Data-driven X-ray image generation for industrial applications

Vladyslav Viacheslavovych Andriiashen

Data-driven X-ray image generation for industrial applications

PROEFSCHRIFT

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

Voorzitter: prof.dr. E.R. van den Heuvel

Promotor: prof.dr.ir. R. van Liere

Copromotor: prof.dr. T. van Leeuwen (Universiteit Utrecht)

Leden: prof.dr. K.J. Batenburg (Universiteit Leiden)
prof.dr. A. Vilanova Bartroli
prof.dr.ir. M. Breeuwer
prof.dr. S. Witte (Technische Universiteit Delft)

Contents

1

INTRODUCTION

1.1   System design for agricultural inspection

Product inspection is an important part of production processes in modern industries. Inspection is performed to evaluate the quality of products — items produced by the industry. The concept of "quality" has many definitions, such as "Quality is fitness for use" [109]. To define quality, it is generally assumed that a product must meet a set of requirements depending on its future use. Requirements may be related to the product's size, color, texture, absence or presence of different features, durability, etc. Some requirements can be verified by visual inspection of the product, while others require specialized equipment.

Different types of product inspection are used in the industrial production process. In some cases, the inspection is performed off-line: the product is transported from the conveyor belt (production line) to a separate facility for quality evaluation. If a single product from the batch meets the quality requirements, the entire batch is considered to be of satisfactory quality. The off-line approach is used when inspection takes a long time or destroys the product. An alternative approach is in-line inspection. In-line systems are built around the conveyor belt to inspect every product on the belt. In-line inspection is valuable for mass production, although not all quality requirements can be verified in-line.

Inspection challenges vary significantly between different branches of industry. The agricultural industry processes raw materials into food products — a broad category that includes grain, canned food, fruits, meat cuts etc. Their quality is characterized from different perspectives, such as food safety, marketing, and storage. A food product must not pose a threat to the health of the consumer. In many modern countries, the risk of foodborne diseases is not negligible [63]. There are governmental regulations to ensure uniform control of products intended for human consumption [37] and to define marketing standards for agricultural

products [38]. Among products fit for consumption, quality is further characterized by observing texture, flavor, and nutritional value [101]. From the perspective of logistics, quality is related to the possible duration of storage time. Various biological processes affect food products over time. This effect may be negligible in the case of immediate consumption, but critical when long transportation is required.

The agricultural industry produces large volumes of food products and benefits from in-line inspection. One important inspection task is foreign object detection. This type of inspection is performed to prevent the production of food contaminated with foreign objects - unwanted entities that affect the quality of the product. There are many different types of foreign objects. Intrinsic objects are a type of objects that are naturally present in raw materials but should be removed during the production process (e.g., bones in meat, stems, and seeds in plants). Extrinsic objects originate from the external environment (e.g., stones, pieces of metal, rubber, plastic). In the thesis, defects in the tissues of a product due to biological processes such as ripening and infestation are also considered foreign objects. In practice, foreign object detection is performed using visual inspection (a human observer inspects the product), imaging methods (an image of the product is acquired and then analyzed), or a combination of both.

The general structure of foreign object detection with imaging is shown in Fig. 1.1. In the diagram, the inspection system is divided into three subsystems. First, there is an image acquisition system that creates an image of the inspected product. There are a variety of image acquisition systems (e.g., visible light, infrared, and X-ray) that produce different images representing different physical properties of the product. The imaging system is chosen based on the foreign objects to be detected. Foreign objects should be visible in the image because they create recognizable image features. The second subsystem is the image analysis, which processes the image to compute a variety of features that may indicate the presence of the foreign object. Finally, the decision-making subsystem uses the analysis results to make a decision — to mark the product as either contaminated or healthy. Contaminated products are later removed from the production process, but this is beyond the scope of the inspection system.

There are many decisions to be made when constructing an inspection system: method of imaging, image analysis algorithm, and decision-making criteria. Different subsystems can be configured independently of each other to better fit the task. Thus, there is a significant value in the system design problem — a problem of constructing an inspection system to meet the needs of the application. This is a challenging problem. An important goal of system design is to achieve the highest possible accuracy of inspection. Accuracy measures how well the system identifies foreign objects when they are present and detects no foreign objects when the product is not contaminated. It is desirable to maintain high accuracy for a wide range of foreign objects and variations in the product.

However, the high accuracy of an inspection system does not necessarily mean that it can be used in practice. Applicable systems face numerous external limitations originating from the industrial environment. For example, in-line systems
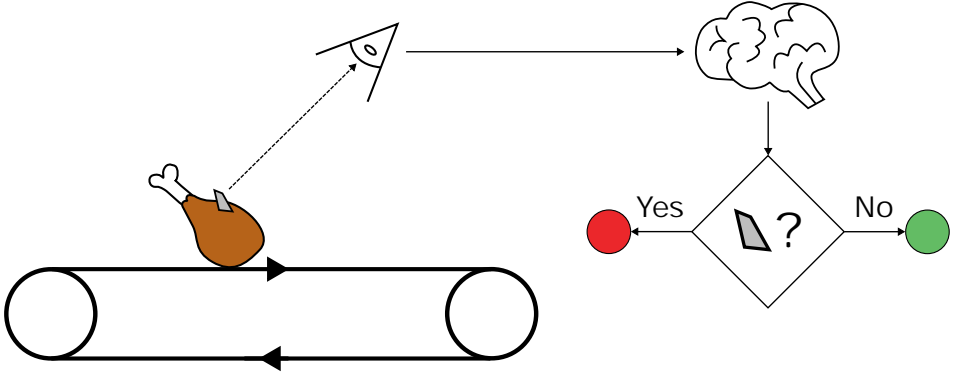
Figure 1.1: Schematic diagram of in-line foreign object detection with imaging. Products (visualized as *chicken drumstick*) are transported by the conveyor belt. The image acquisition system (*eye*) acquires an image of the product. The image analysis system (*brain*) examines the image, and the decision-making system (*conditional block*) concludes whether the product should be removed from the belt (*red circle*) or allowed to move forward (*green circle*).

are constrained by throughput — the number of products moving on a conveyor belt per unit of time. In case of high throughput, an inspection of a single product, from image acquisition to the decision, must not take a long time. If this is not the case, the speed of the conveyor belt will be limited by the inspection system. High-throughput inspection could also have an indirect effect on accuracy because controlling the pose of the product becomes more difficult. This becomes an issue if the foreign object is located on one side of the object and the only available image is taken from the other side.

Another significant limitation is the financial cost. The budget — a limit on the total cost of designing and constructing the inspection system — affects many aspects of the system design. In some applications, expensive inspection equipment is not used because the increase in product quality does not outweigh additional expenses. Research on a better inspection approach may be canceled because it requires too much time and too many experts working on the problem. However, an in-depth financial analysis is beyond the scope of this thesis. On top of expertise in imaging and imaging analysis, it would require a deep knowledge of the market for a particular product and the cost of equipment required to construct a system.

The aforementioned goals and limitations guide the design of the in-line foreign object detection system. The design problem is interdisciplinary. First, the general knowledge of imaging and image processing is used to create a prototype inspection system. Some foreign objects can only be detected with particular image acquisition methods. From a variety of image analysis methods developed in the image processing field, only a subset is fast enough to be used for high-throughput applications. Then, this prototype should be adjusted to a specific

application to reach the necessary accuracy. In practice, this adjustment requires a large amount of application-specific knowledge extracted from large volumes of experimental data.

A distinctive feature of agricultural products is their variety of shapes and internal structures. Both for plants and livestock, small variations during growth lead to similar but not identical structures. While these variations are known by experts in the field, this knowledge is rarely expressed in the form of a mathematical model (with some exceptions, such as shape models of tomato fruits [152]). Such models exist for manufactured products. The geometric shape can be described by a combination of basic shapes (cylinders, spheres, and rectangles) and polygonal meshes. The material composition can be described by a chemical formula. The lack of a model is a fundamental challenge for an automated inspection. Some deviations of the product's structure are foreign objects, and some are natural variations, and there is no rigorous and explicit definition to distinguish them.

The ambiguity of the problem formulation is a well-known problem in computer vision. The simpler algorithms analyze images to find gradients, edges, circles, etc. However, a more advanced approach is necessary to solve a more practical problem, such as detecting a foreign object. Such problems are often solved with machine learning. Similar to product inspection, machine learning methods consist of image processing and decision-making. First, basic computer vision methods are used to extract different image features. Then a decision is made by computing a parameterizable function using these features as input. The parameters of this function are determined by analyzing large amounts of data — images of objects with and without a foreign object. During "learning", the values of parameters are optimized to achieve the highest decision accuracy on the available data. As a result, it is not necessary to have a rigorous definition of a contaminated and an uncontaminated object to distinguish between them. The definition can be extracted from examples and implicitly stored in the decision-making part.

To increase accuracy and robustness, modern machine learning develops towards more complex methods and larger amounts of data. In deep learning — a widely used modern type of machine learning — image processing methods also "learn" from the images. Image processing is implemented as a parameterizable function (e.g., convolution of the image with a kernel), and its parameters are determined in the same way as the parameters of the decision-making. Therefore, features used by deep learning are not constructed manually by experts in the field as in classical machine learning. The deep learning features are extracted from the data and may be more complex and context-dependent. Data-driven features are often discussed in the context of the "explainability" problem [50, 57] — even for an expert in the field, it is often difficult to explain why a deep learning method made a certain decision. Furthermore, low explainability increases the difficulty of another problem of the inspection system design — optimization of the imaging method.

The previous overview assumed that the method of creating an image of a physical product was fixed. Different imaging methods capture different properties of the product and lead to different features that can be used to detect foreign

objects. However, even if the imaging method is already chosen, there are usually a number of flexible settings that influence the process of imaging and the resulting image. The settings are mostly specific to a particular method, with a few more general ones, such as acquisition time and imaging geometry. The accuracy of detection depends on the image features, so changes in imaging settings would ultimately affect the accuracy, and there are optimal settings leading to the highest accuracy. However, deep learning significantly complicates the optimization of imaging. With an expert's knowledge of the imaging method, it is possible to predict the effect of settings on some properties of the image, such as noise level, contrast between different objects, or intensity gradients. However, deep learning relies on complex data-driven features, and it is significantly more challenging to predict the influence of settings on them. A practical solution to this problem is the trial-and-error approach — to change settings and record the resulting accuracy until the desired performance is achieved. As a result, the optimization of imaging becomes data-driven, similar to image processing.

The combination of two data-driven problems leads to a combinatorial explosion of the amount of data necessary to design the optimal system. Yet, the data-driven system design has been highly successful in agricultural applications [81, 137]. There are a number of factors contributing to this success. The most popular method of imaging is visible light photography, which captures only the surface of the product. For many applications, researchers and commercial companies acquire large numbers of images for machine learning and deep learning methods. Furthermore, there are large publicly available datasets, such as ImageNet and CIFAR, that can be used to train a baseline for a deep learning method. Color photos are similar to human vision, so it is possible to optimize the settings of imaging with empirical knowledge (more advanced optimization can be found in hyperspectral imaging [26, 92]). For photography, it has been possible to reach the desired inspection accuracy with extensive experimental work. However, color photography is limited to the surface of the object (near-surface for infrared photos). Another method of imaging is required to detect foreign objects inside the product, and it is not guaranteed that the same brute-force approach will succeed.

1.2   X-ray imaging for inspection

The internal structure of objects can be studied with a variety of imaging methods, including X-ray, ultrasound, and terahertz. X-ray imaging is widely used in medicine and security for non-destructive inspection [107]. Similar to visible light, X-rays are a form of electromagnetic radiation but with a shorter wavelength (0.01 nm - 10 nm). Consequently, X-rays can penetrate thick objects, such as a human body, airport luggage, or a food product. Similar to cameras for visible light, there are X-ray detectors — a type of equipment that can record the intensity of X-rays and produce X-ray images. Classical X-ray images are black-and-white photos in which the value in every pixel corresponds to the intensity of X-rays (Fig. 1.2). When X-rays propagate through the object, their intensity decays

due to X-ray interactions with matter, and the corresponding part of the image becomes darker. The decrease in intensity depends on the physical properties of the object: density, chemical composition, and thickness.

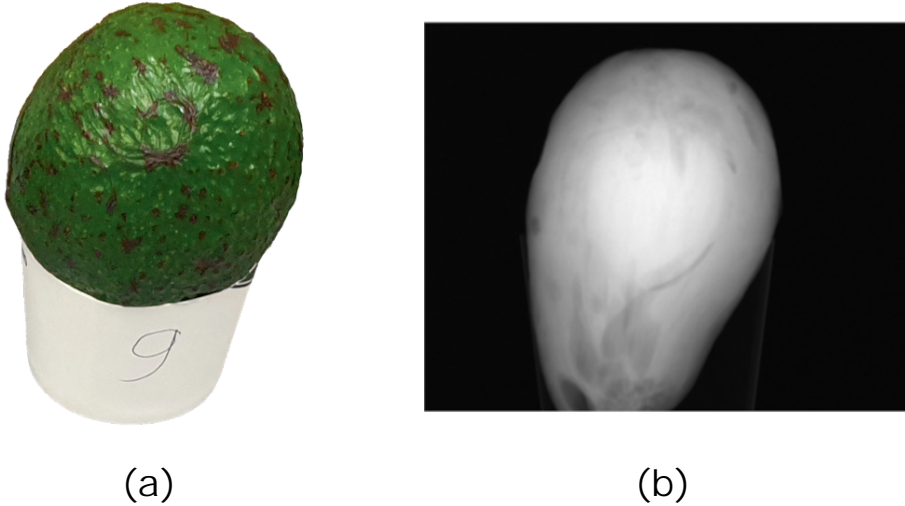(a)   (b)

Figure 1.2: A color photograph (a) and an X-ray image (b) of an avocado fruit. The internal structure of the fruit is not visible in the photograph, but some details (contours of air pockets and a seed) can be seen in the X-ray image.

X-ray imaging has been successfully applied to different problems of agricultural product inspection and foreign object detection in particular. X-rays were used to detect foreign objects (e.g. glass and metal residue) in packaged food, bones in fish and poultry, and infestation in grain and fruits [66]. However, the challenge of detection varies significantly depending on the material composition [89]: metal fragments are easier to detect than plastic fragments or insects. As a result, there have been developed a large variety of X-ray inspection systems tailored to specific industrial tasks.

X-ray imaging does not refer to a single imaging method, it is a general term for a wide range of methods. They differ by the number of images acquired, the settings under which these images are acquired, and the method of X-ray detection. X-ray imaging methods include radiography, computed tomography, spectral imaging, phase-contrast imaging, etc. In the general structure of the product inspection system, different X-ray techniques correspond to the image acquisition method. Image analysis and decision-making have been implemented with a wide range of approaches as well: manual inspection by human experts, conventional computer vision with feature enhancement, machine learning, and deep learning.

The limitations of the in-line high-throughput agricultural inspection narrow

down the range of image acquisition methods that can be used in practice. Classical radiography is often used for high throughputs. In this method, only a single image (radiography) of the object is acquired to maintain the short time of acquisition. To obtain more information within the same period of time, it is possible to acquire multiple X-ray images (e.g., dual-energy imaging) with different properties of X-ray simultaneously. However, detecting foreign objects on radiographies is a challenging task.

The intensity of a pixel in an X-ray image depends on the total decay of X-rays — the difference between the original intensity and the detected intensity. Thus, a value in a pixel is influenced by the full trajectory of X-rays between the source and the pixel (Fig. 1.3). This leads to an effect referred to as an object overlap [106]. When different objects are located along one trajectory at different depths, the corresponding decays of X-rays overlap in the image plane. As a result, it becomes difficult to separate objects using only a single image. Due to the overlap, early applications of the X-ray inspection concentrated on large foreign objects and small foreign objects with high X-ray attenuation, such as metal fragments. In these cases, the presence of the foreign object leads to a large intensity gradient that can be enhanced and analyzed. To detect smaller foreign objects, it is necessary to account for the morphology of the product by, for example, using a deep learning method.
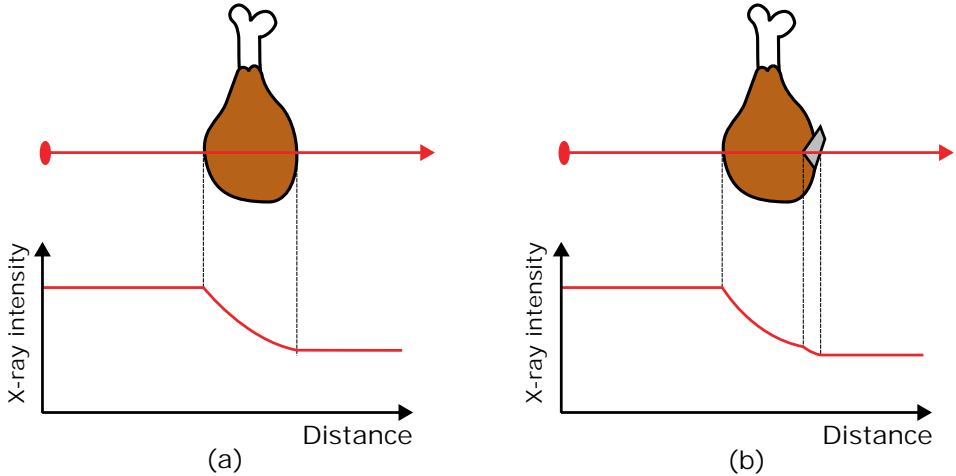
Figure 1.3: Comparison of X-ray intensity decay in an inspected object without a foreign object (a) and in the presence of a foreign object (b). The main contribution to the intensity change is from the product, and the foreign object has a relatively small effect.

The main question determining the applicability of deep learning methods is whether it is feasible to acquire a sufficient amount of images for training. Unlike visible light photography, publicly available datasets containing X-ray images of

agricultural products are rare [35, 104]. The brute-force solution of acquiring a lot of images experimentally also faces new challenges. It is more difficult to predict in advance that the detection problem is solvable. Due to the object overlap in X-ray images, the foreign object can be missed if the product has a complex internal structure with many intensity gradients. With color photos, it is possible to rely on human vision as empirical knowledge to decide in advance whether the foreign object is visible. For X-ray images, a human observer often needs special training (e.g., radiology in medicine) to correctly interpret the data.

Even if the problem can be solved with sufficient accuracy, the necessary number of images may vary. X-ray images are sensitive to changes inside the object. Therefore, the dataset should cover the variety in the internal structure of the product, not only the variety in surface structure. Furthermore, X-ray imaging settings, such as the energy (wavelength) of X-rays, significantly influence the image. Unlike medical applications, there are no standard protocols describing the X-ray imaging settings for the agricultural inspection. The problem of imaging optimization for X-ray is more challenging than for visible-light photos and comparable to hyperspectral imaging.

The increased difficulty of collecting a sufficient amount of images is a barrier to the more widespread use of X-ray inspection. It motivates to investigate whether the data collection can be simplified or accelerated. The scarcity of data is not a unique problem for computer vision tasks. One possible solution is to use data generation — methods that create data for a particular problem computationally using expert knowledge of the imaging method and the problem of interest. In certain cases, the generated data may be indistinguishable from real data. Then there is an argument to replace the experimental data collection with a computational data generation. The problem of simulating X-ray interactions with matter has been studied extensively, and highly accurate results have been achieved. There have been promising studies in which product inspection methods were improved by using generated images [145]. These successes motivate a more general and detailed analysis of image generation as a tool for the system design problem.

1.3   Generation of X-ray images

The concept of "data generation" describes the production of data (referred to as "generated", "synthetic", "artificial") using software and algorithms. This definition is general and encompasses a wide range of approaches. In the context of the system design problem, image generation can be seen as a prediction method. An image generator receives information about the object and the imaging system and predicts what the image would look like. Since there are two data-driven problems, the image generator should be able to address two tasks. For image analysis, the image generator should predict the image of the product (with all possible variations) with different foreign objects of interest in different locations. For optimization of imaging, the generator should predict the image under different

system settings.

The structure of an image generator as a parameterizable prediction method is shown in Fig. 1.4. In contrast to a real experiment, the generator works with digital representations of physical objects and processes. A digitized object is a digital description of the shape and material properties of a physical object. A model of imaging provides a computational method of transforming a digitized object into an X-ray image. The model is parameterizable. Its parameters represent different settings of the imaging system (e.g., X-ray energy, resolution of the image, noise).
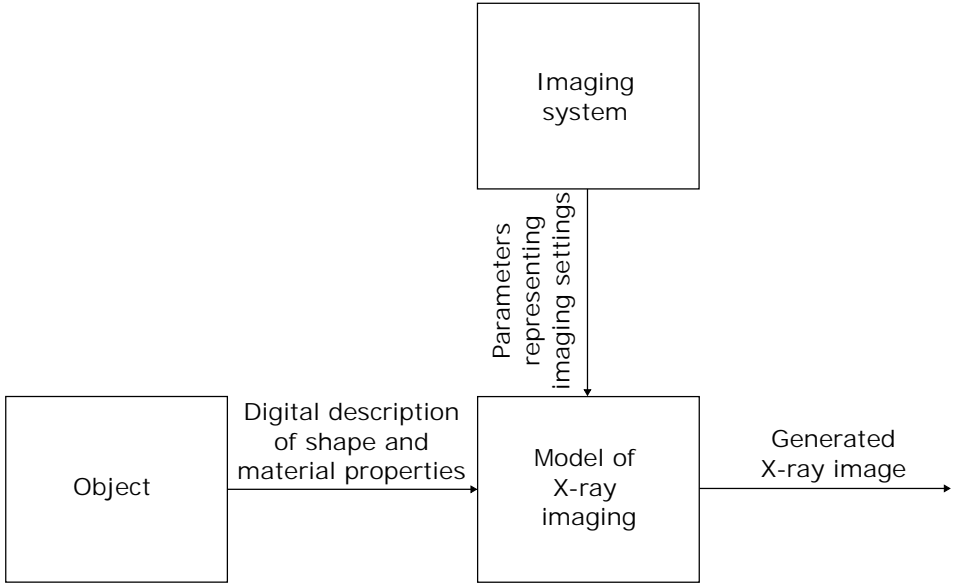
Figure 1.4: Structure of an X-ray image generator. It receives information about two physical entities: the object and the imaging system. This information is used by the model of imaging to create an X-ray image corresponding to a real image of this object acquired with this imaging system.

The image generator has two advantages over an experimental acquisition. To acquire a large number of images, it is necessary to have a large number of real objects. With the generator, it is possible to predict an image without a real object, as long as it is possible to create its digital description. For example, generate many images of the same product with a foreign object at different locations. The second potential advantage lies in the speed of generation compared to the speed of experimental acquisition. It may be faster to generate images under different imaging settings than to change them experimentally. However, these advantages could only be achieved if the generation method is appropriately implemented. The generator needs a method of describing a diverse range of products and possible foreign objects. There should be a way to represent different states of the imaging systems as parameters of the generator. Finally, the model of imaging needs to be

fast and accurate.

1.3.1   Models of imaging

The model of imaging has a defining impact on the image generator. It determines the algorithm of image computation and how the object and the imaging system are digitized. There are two fundamentally different approaches to modeling the imaging process: stochastic and deterministic. They differ in the method of describing random processes comprising the imaging process. An X-ray image is a result of detecting many individual X-ray photons (Fig. 1.5). Each photon moves through different parts of the object and contributes to the image if it reaches the detector. An X-ray trajectory is affected by its interactions with the object: photoelectric effect leading to an absorption of X-ray and scattering causing a change of X-ray momentum. Probabilities of these interactions are theoretically studied and experimentally verified [25]. A combination of all interactions forms a complex distribution of X-ray images that could be acquired. Stochastic models of imaging perform a Monte-Carlo sampling of this complex distribution to compute an image. Deterministic models do not model the distribution and instead approximately compute the average image.
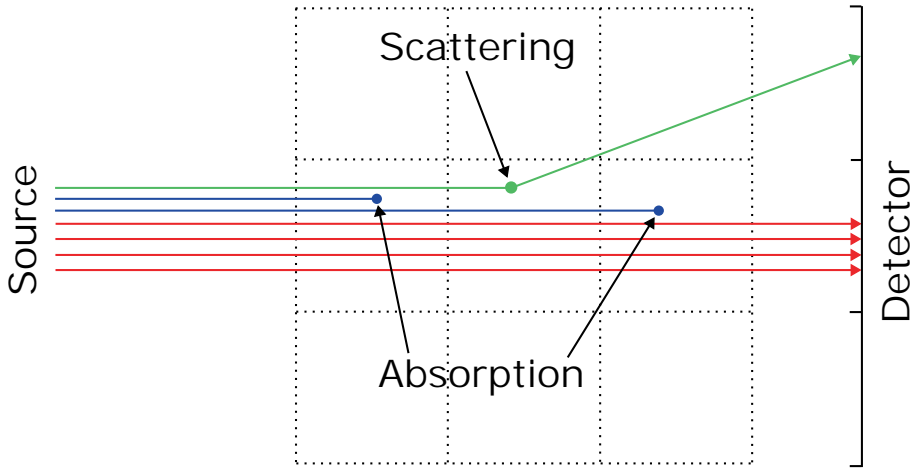
Figure 1.5: Illustration of an X-ray image formation. Dotted cells represent small regions of the object, and colored lines show possible X-ray trajectories. Red lines correspond to X-rays that propagated through the object and were registered by the detector. Blue lines show the trajectories of X-rays that were absorbed after interacting with atoms of the object. The green line indicates a scattered X-ray that was registered but in a different pixel of the detector.

Stochastic models of imaging can be highly accurate. The physics of X-ray interactions is well known, so the accuracy of a Monte-Carlo simulation is only

limited by the ability to accurately describe every element of the imaging process (X-ray source, object, detector). Due to their high precision, these methods have been successfully used in medical applications [62, 125]. However, a similar success is not guaranteed in industrial applications. Monte-Carlo sampling has a high computational cost proportional to the number of X-rays that need to be simulated. As a result, a stochastic generator could be orders of magnitude slower than an experimental acquisition unless there are a lot of available computational resources. Furthermore, digitizing the object and the imaging system is a challenging task. The properties of X-ray interactions are computed based on the chemical composition of the object and the energy properties of X-rays. This information is not trivial to obtain, even if a particular object and system are physically available. Chemical analysis can be used to characterize agricultural products [10, 33], but it usually provides a general assessment, not a distribution of chemical elements over the volume of the product. An accurate description of the X-ray source and the detector requires a lot of specialized equipment, which is rarely available in an industrial environment. Thus, stochastic generation is difficult to use in practice.

Deterministic models provide an alternative way to describe the imaging process. Instead of computing an image based on fundamental processes, image formation is approximated by simpler equations. A widely used example of a deterministic model is Beer's law. It describes the object using the function μ. The value of μ is an X-ray attenuation coefficient — a property of the material describing how much it reduces the flux of X-rays. The X-ray image y is computed following the equation

$$y = I_0 \exp\left(-\int_l \mu \, dx\right), \tag{1.1}$$

where l is a trajectory from the X-ray source to the pixel of the image, and the integral represents the cumulative reduction of X-ray flux along this trajectory. I_0 is the initial intensity of X-rays representing imaging settings.

A significant advantage of the deterministic approach is the possibility to estimate all required properties experimentally, using only an X-ray system. To obtain the value of μ in Beer's law, it is not necessary to know the chemical composition of the object. The simplest method would be to cut a small fragment of the object and measure how much it attenuates X-rays. An alternative, non-destructive approach is to perform a computed tomography scan by acquiring X-ray images from different angles. Then a distribution of μ can be computed following a reconstruction algorithm. While some methods cannot be performed with an in-line X-ray system, they are easier to implement than the complex analysis necessary for stochastic methods. A similar experimental approach can be applied to more complex deterministic models. For example, X-ray scattering can be approximately modeled using kernel convolutions [133], with all necessary properties extracted from suitable calibrations.

1.3.2   Digitization of the object and the imaging system

With a deterministic approach, an X-ray image can be recreated with a generator if a sufficient amount of experimental data is provided. However, for a practical generation of large datasets, it is necessary to predict the properties of objects without extensive study. One way to achieve this is to assume that similar objects are made of the same set of materials - regions with similar X-ray properties. Then it is sufficient to characterize the properties of the material once, and this information can be used to generate images of a variety of objects.

The assumption about common materials leads to a number of problems that need to be solved for every particular inspection task. There is no general definition of materials. With a sufficient imaging resolution, agricultural products can be represented as sets of fibers, cells, and porous structures. It is crucial to represent these as a generalized material with constant X-ray properties. Furthermore, there should be a method of representing the shapes of objects (e.g., sets of voxels or meshes). A set of possible shapes must include different variations of the inspected product, different types of foreign objects, and different locations of the foreign object. This variety of shapes should, in some way, be extracted from a variety of real products.

With the data-driven method of describing objects, it is especially important to account for the effect of imaging settings. If a property is extracted from the image, it implicitly assumes certain imaging settings, such as the energy of X-rays, resolution of the detector, and acquisition time. To predict images under different settings, it is necessary to define how the imaging system affects material properties and represent the state of the system as a set of parameters. This connection could be derived from the data by performing imaging of the same object under different settings. Alternatively, fundamental knowledge of X-ray interactions could be used.

Another significant effect of imaging settings is found in the noise properties of a generated image. As opposed to the stochastic approach, deterministic models do not account for random processes of X-ray image formation and consequently generate a noiseless image. To increase the similarity between generated and real images, noise can be added as a post-processing step [93]. However, noise models are usually used to add an arbitrary amount of noise (e.g., to test a method of computed tomography) and not to imitate a particular imaging system. For accurate generation, imaging settings should be connected to parameters of the noise model, so the magnitude of noise corresponds to a real acquisition.

1.4   Validation of the image generator

Implementing an imaging model, especially a data-driven deterministic model, involves many assumptions and approximations. As a result, the generator diverges from an accurate representation of the imaging process and might predict images with systematic errors. This is a significant issue for the data-driven system design. If the system is optimized fully or partially on generated images containing

systematic errors, its performance in practice on real images from the production process will be compromised. Thus, the accuracy of the generator needs to be studied in detail. There should be a qualitative and quantitative assessment of how similar generated images are to the corresponding real images, and how the differences may affect the system design.

1.4.1 Image-based validation

The accuracy of the generator is a measure of the similarity between generated images and real images. The presence of noise fundamentally complicates the definition of accuracy. Since X-ray images are sampled from a complex random distribution, two real images of the same object acquired under the same settings will never be identical. This effect is insignificant if the flux of X-rays is large and the relative magnitude of noise is low. However, for in-line inspection, the time of acquisition is often short, and acquired images are noisy.

A standard method of evaluating the accuracy of image generation methods is a phantom study [12, 41, 163]. A phantom is an object with a well-known morphological and material structure. While some phantoms represent objects of interest for a particular application, others are combinations of simple geometric shapes made of known materials. For an image generator, a phantom is an object that can be described with high precision, so any inaccuracies would stem from the model of imaging. The model is then validated by comparing different properties of the corresponding real and generated images: intensity profiles, root-mean-square of the difference, point spread function, etc. Due to the noise in individual pixels, these properties can only be measured with statistical uncertainty. However, if the noise level is low, many image properties can be compared to verify if the generator predicts real images accurately (Fig. 1.6a).

In industrial applications with a high expected noise level, a phantom study should be analyzed from a different perspective. When the relative magnitude of the noise is comparable to image features, the accuracy analysis becomes more uncertain (Fig. 1.6b). A generator might use an approximate model that makes systematic mistakes in predicting image intensity. However, if the difference is statistically insignificant, it is not possible to prove that a generator is inaccurate. On the other hand, an inaccurate noise model (e.g., wrong distribution, variance, spatial correlation) might make images visually dissimilar, even if the difference is challenging to estimate quantitatively. Verifying that two images draw from the same random distribution is a challenging task that requires a lot of prior knowledge and data to analyze.

1.4.2 Application-based validation

The similarity analysis is further complicated by the low interpretability of deep learning methods. With sufficient prior knowledge and expertise, it is possible to verify that the generator correctly predicts certain properties of the image. However, it is not explicitly known which of these properties influence deep learning
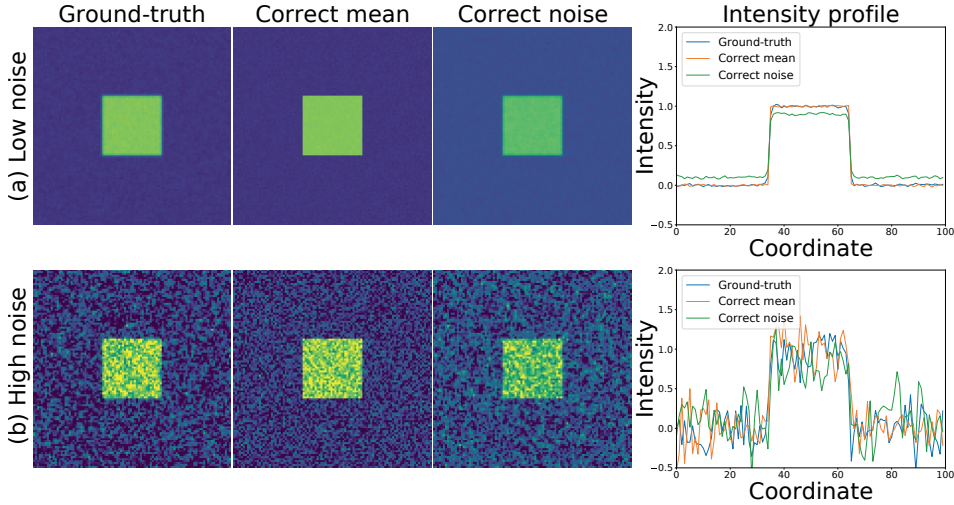
Figure 1.6: Comparison of different image generation approaches. The ground-truth is a square of constant intensity with zero background, the image is corrupted by Gaussian noise and Gaussian blur. Two generation methods are proposed: the first has accurate object and background intensities and a wrong noise model, the second is the opposite. For low noise (a), the difference in average intensities is significant. For high noise (b), inaccurate intensities are less visible than the difference in noise distributions.

methods significantly. While the effect of image generation on data-driven methods is challenging to predict from image similarity alone, it can be measured directly. First, an application is defined by specifying a task, such as foreign object detection in products of a certain type with a number of possible foreign objects. This object variety is then represented by a set of objects. Real images of these objects are acquired experimentally, and an image generator is used to produce a set of generated images of the same objects. Finally, these sets of images are used to train inspection methods solving the specified task, and their performance is directly compared. In this thesis, this approach will be used as an application-driven evaluation of generation accuracy.

This application-driven definition of accuracy has many fundamental challenges that need to be addressed. It does not generalize to all applications: similarity of inspection accuracy is only verified for one task. It does not generalize to all possible objects either: the inspection method is trained and tested on a limited set of objects, and the results may be incorrect if the set lacks a certain variety of objects. It depends on the chosen metric of inspection performance: the behavior of inspection systems may differ if the difference is not described by the metric of performance. This method is computationally expensive: measuring the accuracy of inspection requires not only to generate images, but also to train an inspection method. However, this approach characterizes the most important aspect

of accuracy and produces a quantitative estimate. While specific results are not generalizable, the application-driven accuracy evaluation is a general method that can be applied to a large variety of applications, tasks, and sets of objects.

A metric describing the inspection performance is a significant part of the application-driven accuracy evaluation. This metric should represent all qualities of the inspection system that are important for a particular application. In the simplest form, the inspection system should rather be correct in its predictions than wrong. However, foreign object detection can be characterized in many other ways. What foreign objects can be detected by the system, and what is the frequency of failure for every type? What is the smallest size of the foreign object that can be detected? At what location of the product are they most likely to be missed?

There are many standard metrics to describe the general accuracy in deep learning [153]. For foreign object detection, a popular approach to performance description is to use a set of test objects and ground-truth — labels describing whether objects contain foreign objects. Then an inspection method is used to characterize the objects and produce a set of responses. These responses are compared with the ground-truth to compute different metrics of accuracy: precision, recall, F1-score, etc.

In industrial foreign object detection, a single value of accuracy is not a sufficient metric of performance because the accuracy is significantly affected by the properties of the inspection object. For example, if a test set is composed of objects with large foreign objects producing a lot of intensity contrast in the image (e.g., large metal residue for X-ray imaging), the inspection is likely to be very accurate. However, this accuracy would drop significantly if the foreign objects are small or produce less contrast. If a test set covers a variety of different foreign objects, the resulting accuracy reflects the composition of the set and how accurate the method is under different circumstances.

The problem of correlation between the inspection accuracy and properties of the inspected object is studied in detail in non-destructive testing. One of the popular approaches is Probability of Detection [55] — a regression analysis method that was originally developed for evaluating human observers and then extended to different problems and imaging methods. Probability of Detection is a purely statistical method that views the inspection system as a black box. The method only quantitatively characterizes the correlation between accuracy and any object or image property that has to be determined by a human expert based on the problem knowledge. With a sufficient variety of the test set, the Probability of Detection presents a method to describe the inspection system in detail and addresses the questions relevant to the industrial application.

1.5   Research questions

The thesis addresses the challenges of image generation for the design of foreign object detection systems for industrial applications. While a significant amount

of relevant fundamental knowledge has been developed, adapting it to the real problems of the agricultural industry is a challenging task. In the absence of rigorous definitions of products and foreign objects, it is necessary to develop a method to describe both the shape and material structure digitally. The model of imaging should be chosen to provide the necessary accuracy of generation while not requiring a prohibitively large computational cost. There should be a method to determine which imaging effects could be ignored during generation, depending on a particular application. Requirements of the industrial problem should be connected to the properties of the inspection system, and these properties should be translated into parameters of the image generator. The goal of the thesis is to create a parameterizable data-driven X-ray image generator based on real objects and the imaging system and show how this generator can be used to design more accurate inspection systems.

The challenges of the system design and image generation are formulated as the research questions. They are stated on a separate page each.

Research question 1. *How the theoretical knowledge of X-ray image formation can be used to improve classical and deep learning image analysis methods?*

Foreign object detection requires deep learning methods because the underlying computer vision problem is challenging, products and foreign objects are ill-defined. However, this issue can be addressed from the perspective of imaging. A robust pre-processing method is proposed to estimate the material properties of the product from a dual-energy image. This information is then used to create an image in which the attenuation from the product is effectively "subtracted", and only the foreign object is visible. As a result, a classical image analysis method can be utilized to perform foreign object detection. The performance of this method is compared to a deep learning detection with and without the proposed pre-processing.

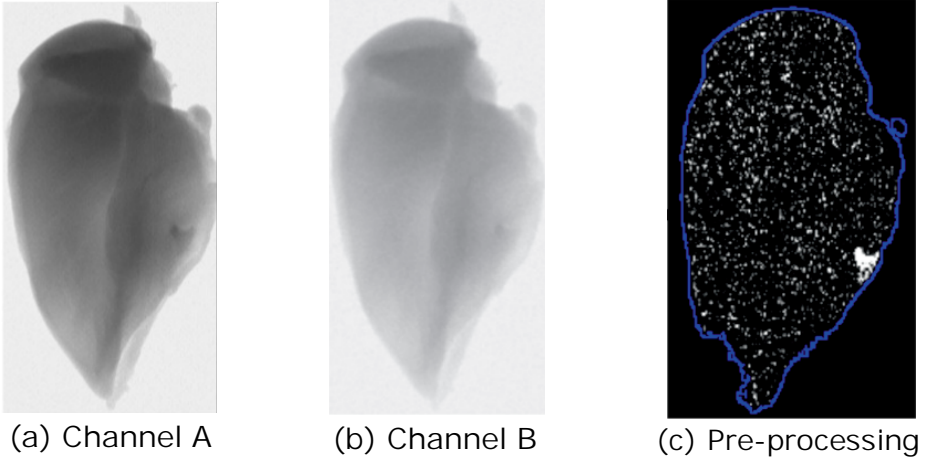(a) Channel A   (b) Channel B   (c) Pre-processing

Figure 1.7: Comparison of raw data (a, b) and the result of the proposed pre-processing (c). The raw data are two X-ray images acquired with different X-ray energy properties. The pre-processed image is based on a quotient between these two images and highlights the foreign object due to the difference in material properties.

Research question 2. *How to generate X-ray images of non-existing objects based on a limited number of real objects and their X-ray images?*

Efficient generation of large volumes of X-ray images relies on the ability to predict images of non-existing objects - variations of the product with different foreign objects at different locations. Computed tomography reconstruction is proposed as a method of creating digital descriptions of real objects. These digitized objects can be deformed to create new objects. The material properties of the products and imaging settings are extracted from the experimental data. It is demonstrated that with a suitable deformation method, highly accurate inspection methods can be trained with a limited amount of experimental work.

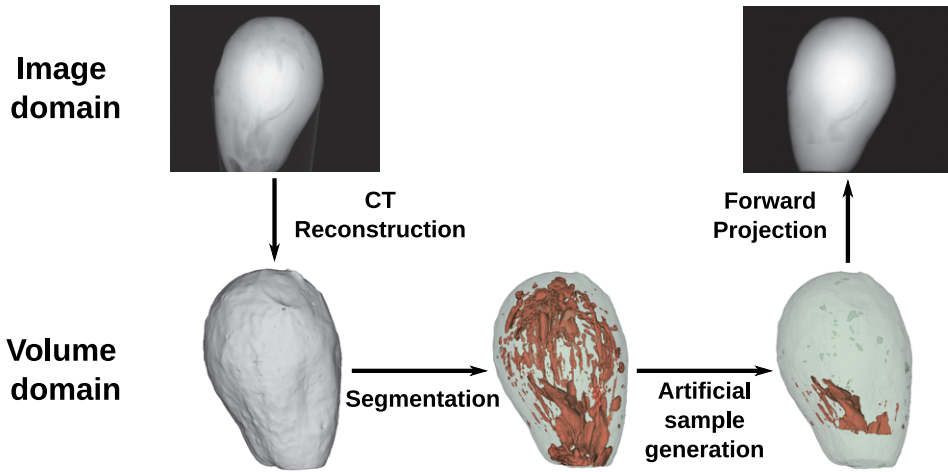Figure 1.8: Diagram describing the CT-based image generation approach. The method takes real X-ray projections as input and generates new projections corresponding to a deformed product volume. The transition between projections and volumes is done using CT reconstruction and forward projection. Volumes are deformed by representing the product on a voxelized grid with discrete voxel values corresponding to different materials.

Research question 3. *How to estimate the effect of X-ray scattering in the generation method on inspection accuracy?*

Application-driven accuracy evaluation is used to determine how much a model of imaging can be approximated before the generated data can no longer be used for the system design. A stochastic and a deterministic model are compared based on the resulting accuracy of the inspection method applied to synthetic data. The deterministic model does not simulate X-ray scattering, and the absence of scattering is shown to be significant only for a subset of objects. The Probability of Detection method is used to demonstrate under which conditions the difference in imaging models is significant for the inspection.
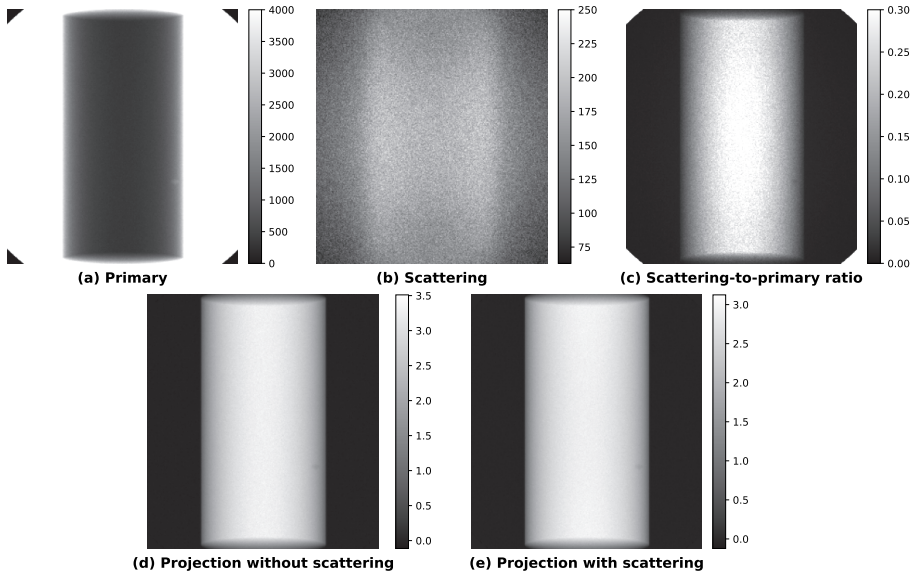
Figure 1.9: X-ray images of a phantom computed with a stochastic generation method: primary image is formed by X-rays propagating through the object without scattering (a); scattering image is formed by X-rays that have scattered (b); the ratio of scattered and primary X-rays in each pixel (c); images after flatfield correction with and without scattering (d and e).

Research question 4. *How to represent the imaging system as parameters of the generator? How to use the generator to design the system?*

The main difficulty of optimizing imaging settings with an image generator is in connecting the properties of the real system with the parameters of the imaging model. A parameterizable image generator with a mixed Poisson-Gaussian model is considered. It is demonstrated how the exposure time of X-ray images can be translated into parameters of the noise model. As a result, the inspection accuracy can be studied as a function of exposure time and, consequently, throughput of the production line.
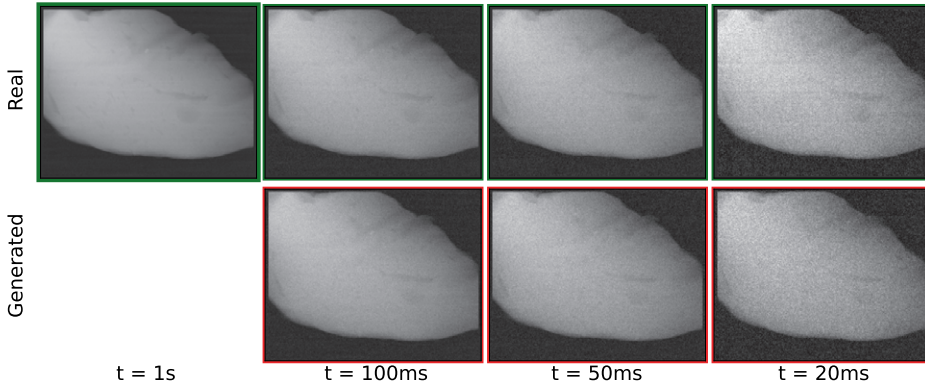
Figure 1.10: Comparison of experimentally acquired X-ray images with different exposure times and corresponding generated images. With a suitable noise model and system calibration, it is possible to achieve both high similarity and fast generation. The similarity between the generator and the real measurement is quantified by measuring the accuracy with Probability of Detection curves.

2

DUAL-ENERGY INSPECTION

Agricultural food products naturally vary in detail regarding their internal structure. Individual samples can be analyzed to assess product quality, to predict maturity state, and to minimize waste. To facilitate early detection of health risks, it is crucial to apply a procedure to inspect the capability to detect foreign object inclusions [32, 84, 85] and contaminations [34, 148]. This task can be performed by a human expert for an individual sample. However, manual inspection cannot provide reasonable speeds for high-throughput cases, such as product processing in a factory.

X-ray imaging is a widely used technique for nondestructive, in-line inspection of agricultural products [105]. It is commonly applied to food samples while they are processed on a conveyor belt on the factory floor. One of the important applications of X-rays is the automatic detection of foreign objects (also referred to as defects in this chapter) that might appear in food products. Examples of such objects are bone fragments in a chicken fillet, plastic debris and bones in fish, and infestation in fruits. One of the well-known approaches in foreign object detection is the acquisition of two dual-energy X-ray absorptiometry (DEXA) [91] projections with different values for X-ray tube voltage. This method is commonly used in medical X-ray imaging for contrast agent detection and body composition analysis.

In-line foreign object detection in food samples on a conveyor belt possesses three major challenges for DEXA analysis. First, high-throughput X-ray acquisi-

---

tion leads to a significant noise level that greatly impacts the detection process. Noise reduction methods have to be applied to reduce this effect. Second, typical foreign objects in the food industry have similar X-ray attenuation properties to that of food samples, resulting in low contrast between foreign object and main object. Contrast-enhancing methods have to be applied to mitigate this effect. Third, a variation in the thickness of the main object causes ambiguities when detecting a foreign inclusion in a sample. For example, a thin sample of a chicken fillet with a bone fragment might have similar dual-energy measurements to a thicker sample of a fillet object without a bone. Hence, a thickness correction method should be taken into account to mitigate this effect.

The main contribution of this chapter is a novel approach to DEXA image pre-processing. For low-contrast foreign objects, it is crucial to analyze how the ratio between two projections acquired with different voltages depends on the thickness of the object. This effect is caused by the polychromatic spectrum of the X-ray tube and the unknown shape of the sample. The correlation between two intensities for different voltages is not the same for the foreign object and the product sample, and this difference can be utilized to distinguish between them. The goal of DEXA pre-processing is to create an image where the main object has an average intensity of zero whereas the foreign object deviates significantly from zero.

This chapter presents a three-step data processing methodology that uses the DEXA pre-processing model in combination with a segmentation method and parameterizable foreign object detection criteria. The methodology is optimized to achieve high detection rates on samples with foreign object inclusions and, perhaps more importantly for industrial applications, achieves low detection rates on samples without inclusions. Although the results are targeted towards bone fragments in chicken fillets, the processing methodology is generic and can be used to analyze the performance of various industrial scenarios of foreign object detection on a conveyor belt. The DEXA pre-processing is based on general concepts of X-ray measurements and can be applied to other materials. Two approaches to image segmentation are considered: an unsupervised active contour model and a supervised deep learning method.

2.1   Related Work

The nondestructive study of products is an important topic for the food industry. It is applied to a variety of food samples [44], and every type of object has different details and typical foreign objects to detect. X-ray imaging can be used for the detection of grain infection, fruit infestation, and bone detection in the fish and poultry industry. X-ray CT is of great interest since it enables volume reconstruction and detection of defects based on the internal structure of the product. However, this approach requires a significant time for data acquisition and reconstruction. Discrete tomography based on limited-angle data [118] can be introduced to balance acquisition time and reconstruction quality. This chapter

concentrates on the radiography approach since it can provide the fastest inspection.

Foreign object detection based on a single projection has been studied for different types of food, such as poultry and apples. Multiple algorithms for fruit inspection rely on the shape knowledge that can be estimated with a certain accuracy [143]. However, knowledge of the shape of the product is not necessary if the foreign object has significantly different X-ray absorption properties. Several algorithms can be applied to enhance foreign object detection by utilizing conventional filtering [103], local contrast enhancement [32], and local adaptive thresholding [98]. These methods do not require DEXA data and can be applied to a single projection. However, they rely on the assumption that an absorption gradient on the border of the foreign object can be distinguished from gradients in the main object.

The addition of the second projection with a different tube voltage for better foreign object detection is a concept that is widely used in medicine. It is applied to body fat measurements [149] that determine the percentages of different tissues in the human body based on their absorption rates corresponding to different tube voltages. This problem may be similar to some types of food inspection (assessment of fat level), but foreign object detection focuses on small inclusions of different objects. Furthermore, in many applications, material identification is performed using dual-energy CT [96]. This approach is more accurate than DEXA because attenuation properties are analyzed for a small region of the internal structure of the object. If the data are limited to a single projection, the measured attenuation distribution depends on both material properties and the unknown shape of the object (thickness along the ray trajectory). In [45], this effect was explained by beam hardening and corrected with a system calibration.

In the poultry industry, the addition of a laser was considered to obtain a thickness profile of the studied object [59, 135]. Knowledge of the exact thickness profile helps to predict an absorption distribution for the main object if it is homogeneous. Thus, the presence of the foreign object can be detected by simple thresholding after subtracting the expected distribution from the measured absorption signal. In this study, only X-ray equipment is used to perform detection, and no additional sources of information is used.

There is a wide range of classical [80] and deep learning [108] methods for image segmentation. Both types of methods have been used for industrial foreign object detection [141, 146]. In this chapter, two approaches to image segmentation are considered. Active contour methods provide an unsupervised and robust way to segment foreign objects. The main downside of many active contour models is that they rely on the edges to perform segmentation. With a high noise level, any edge information becomes unreliable since noise deviations are bigger than a natural absorption gradient that would be observed on a noiseless image. The Chan–Vese energy equation [27] is used to partition an image into two phases without relying on edge detection. On the other hand, deep learning methods are generally considered state-of-the-art for image segmentation [67, 110] even if for some practical problems the difference in accuracy is insignificant [165]. However, many state-

of-the-art methods require large volumes of data to achieve high accuracy. Some deep learning architectures, such as the Mixed-Scale Dense Convolutional Neural Network (MSD) [117], are designed to have fewer parameters and require less data than other architectures. MSD networks have been particularly successful in segmenting X-ray images and slices of CT reconstructions [126, 161].

In this work, the inspection procedure is evaluated based on the detection rate and not segmentation accuracy. Typical studies of the algorithms concentrate on the images with a foreign object and estimate the accuracy of segmentation. However, the detection rate is more important for most industry applications since the majority of the samples is expected to be without a defect. Such a study was performed, for example, in [144] for pear fruit inspection. Methods to distinguish between bones and no bones for different patches of fish images were proposed in [103]. An algorithm with a good detection rate might show suboptimal segmentation accuracy for samples with a foreign object. However, a good inspection procedure requires a balance in performance on normal and defected samples.

2.2   Materials and Methods

2.2.1   General Methodology

The product inspection procedure proposed in this chapter consists of several stages. The corresponding flowchart is shown on Fig. 2.1. First, two X-ray images of the studied sample are obtained using two different voltages of the X-ray tube. These projections should be aligned, and darkfield and flatfield corrected. The images are divided to produce a combined image distribution for the ratio of two absorption rates corresponding to different voltages. This image is further processed using thickness correction procedure. The goal of this step is to create an image where pixels of the main object have close to zero values and a foreign object presence leads to a sufficiently large nonzero intensity. Segmentation is performed on this image to divide it into two phases with different mean intensities. This leads to a set of clusters corresponding to the regions of the foreign object inclusion. Properties of these clusters are used to decide if the sample should be marked as containing a foreign object.

2.2.2   Dual-Energy Projection Pre-Processing

X-ray imaging can be used to create a projection of the studied sample. The value of every pixel in such resulting projections depends on the integral absorption of the object's matter across the corresponding trajectory. The main object and a foreign object absorb radiation differently, and this leads to differences in pixel intensities. However, the shape of the studied sample is not known in advance. Thus, the pixels in the object region do not have constant pixel intensities, as their values depend on the sample thickness. Two images acquired under different voltages provide additional information since material absorption depends on the X-ray photon energy.
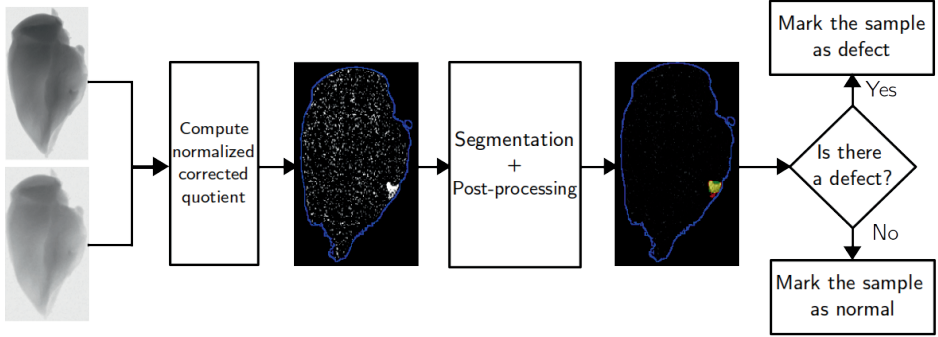
Figure 2.1: Diagram of the foreign object inspection procedure. The input is two projections of the sample acquired with different X-ray spectra. The blue curve on the images approximately shows the sample boundary. The segmentation image uses green color for pixels that were incorrectly classified as foreign objects, red for not detected foreign object pixels, and yellow for detected foreign object pixels.

A dual-energy projection of a sample with a foreign object can be segmented as a two-channel image. The X-ray absorption rate in a pixel depends on both material attenuation properties and the thickness of the object. The absorption rate $M(x)$ computed for the detector pixel x is given by

$$M(x) = -\ln \frac{P(x)}{F(x)} = -\ln \frac{\int_0^\infty I_0(E) e^{-\kappa(E)L(x)} dE}{\int_0^\infty I_0(E) dE}, \qquad (2.1)$$

where P (projection) is an intensity of the detected X-ray radiation when the sample is present, and F (flatfield) is an intensity measured in the absence of the sample. Different voltages are described by the spectrum of the X-ray tube $I_0(E)$ – a distribution of X-ray photon energies E. $\kappa(E)$ is a material absorption curve that also depends on the X-ray energy. $L(x)$ is a profile of thickness along the ray. The argument x refers to a detector pixel, and every pixel has a corresponding X-ray beam trajectory from the source to this pixel. The absorption curve κ does not depend on x since the material is assumed to be homogeneous. If scattering is not considered, attenuation properties of the material are defined by X-ray absorption, and the attenuation rate can be calculated according to Eq. 2.1.

If the tube spectrum is monochromatic, then $I_0(E) = I_0 \delta(E - E_0)$, where $\delta(x)$ is a Dirac delta function. Eq. 2.1 can be simplified:

$$M(x) = -\ln \frac{I_0 e^{-\kappa(E_0)L(x)}}{I_0} = \kappa(E_0)L(x). \qquad (2.2)$$

In this case, the two channels of the dual-energy image are linearly correlated. If a homogeneous material is scanned with two monochromatic beams of energies E_1 and E_2, the corresponding absorption rates are $M_1(x) = \kappa(E_1)L(x)$ and $M_2(x) = \kappa(E_2)L(x)$. The ratio between M_1 and M_2 is constant, does not depend on the

thickness, and is defined by the ratio of attenuation coefficients for two X-ray energies. As a result, two different materials can be easily separated using a dual-energy projection.

In most CT applications, a beam is usually polychromatic since it is produced by conventional X-ray tubes. In this case, the attenuation rate depends on material thickness according to Eq. 2.1. If the thickness $L(x)$ is small, an effective attenuation coefficient κ_{eff} can be computed as a first-order approximation [68]:

$$\kappa_{\text{eff}} = \frac{\int_0^\infty I_0(E)\kappa(E)dE}{\int_0^\infty I_0(E)dE}. \tag{2.3}$$

However, the attenuation rate does not linearly depend on $L(x)$ in general. Thus, a ratio of attenuation rates is no longer a material characteristic and it depends on the thickness $L(x)$.

A nonlinear dependency of attenuation rate on material thickness is visible in the simulated data. For the simulation, the main object is assumed to be a skeletal muscle with an attenuation curve taken from the NIST database (ICRU-44 report [64]). The tungsten tube spectrum for the voltages of 40 and 90 kV is computed according to the TASMIP data [20]. Fig. 2.2a shows how attenuation rates for two different voltages of the tube correspond to each other. On this plot, thickness changes from 0.1 mm to 20 cm, and the attenuation rates are calculated according to Eq. 2.1. The correlation between the two values is almost linear as if both of them depend linearly on thickness. However, the ratio of two attenuation rates changes with thickness, as shown in Fig. 2.2b. This change is not significant; therefore, it is not visible on a correlation plot between two intensities for different voltages. The ratio dependency on thickness can be calculated as follows:

$$\frac{M_1(x)}{M_2(x)} = \frac{-\ln \int_0^\infty I_1(E)e^{-\kappa(E)L(x)}dE + \ln \int_0^\infty I_1(E)dE}{-\ln \int_0^\infty I_2(E)e^{-\kappa(E)L(x)}dE + \ln \int_0^\infty I_2(E)dE}, \tag{2.4}$$

where $M_1(x)$ is the attenuation rate computed for the voltage of 40 keV, $M_2(x)$ corresponds to 90 kV, and $I_1(E)$ and $I_2(E)$ are the tube spectra for voltages of 40 and 90 kV

In the real scan, the nonlinear dependency is further complicated by the noise presence. A mass attenuation function is usually unknown for many food industry materials. Therefore, the thickness dependency of the ratio values cannot be predicted beforehand and should be extracted from the data. Experimental measurement produces distributions of $M_1(x)$ and $M_2(x)$ for two different tube voltages. The quotient distribution $R(x)$ can be computed as follows:

$$R(x) = \frac{M_1(x)}{M_2(x)}. \tag{2.5}$$

The thickness profile $L(x)$ is unknown. As shown in Fig. 2.2a, attenuation rate $M(x)$ is almost proportional to $L(x)$. Thus, in a data-driven approach, the dependency of quotient values on thickness can be studied as a dependency of
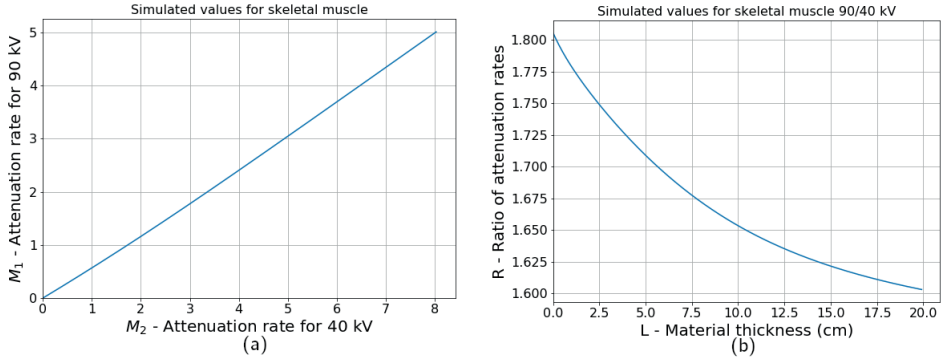
Figure 2.2: Correlation between the absorption of skeletal muscle for X-ray tube voltages of 40 and 90 kV (a). The ratio between the attenuation rate is drawn as a function of thickness. The ratio is not constant due to a polychromatic spectrum (b).

$R(x)$ on either $M_1(x)$ or $M_2(x)$. Values of $M_2(x)$ are lower than $M_1(x)$ for the same x since the voltage of M_2 is higher. Therefore, $M_2(x)$ has a lower absolute value of the error and is used as an argument in the function $R(M_2)$. The function $R(M_1)$ can be studied as well. The dependency $R(M_2)$ is further replaced by a polynomial approximation since a high noise level makes it impossible to recover true function from the data without any additional information.

The order of the polynomial chosen for a function approximation depends on the data quality. In the experimental data used in this work, a linear approximation of $R(M_2)$ is not sufficient and leads to significant discrepancies between the acquired data and the fit. High orders of the polynomial are prone to noise, and the fit does not always converge as a result. The quadratic approximation was chosen as a middle ground since it provides a sufficiently good representation of the data and has a low noise sensitivity. This approximation is given by

$$R(x) = \frac{M_1(x)}{M_2(x)} \approx aM_2^2(x) + bM_2(x) + c, \tag{2.6}$$

where the coefficients a, b, c are computed based on a polynomial regression. The regression is performed for all pixels of the object simultaneously. The values of the fit coefficients implicitly predict $M_1(x)$ for each value of $M_2(x)$ to compute the quotient $R(x)$.

When the dependency of $R(x)$ on $M_2(x)$ is extracted from the data in the form of polynomial approximation, the effect of thickness dependency can be reduced. After a polynomial fit, the distribution of $R'(x)$ can be computed as follows:

$$R'(x) = R(x) - aM_2^2(x) - bM_2(x) - c, \tag{2.7}$$

where $R'(x)$ is a corrected quotient distribution. If the sample consists of a homogeneous material, $R'(x)$ is close to zero regardless of the thickness. However,

inclusion of a foreign object with different absorption properties affects both $R(x)$ and $R'(x)$. $R'(x)$ is easier to use for defect detection since the form variation of the object does not significantly influence this distribution.

2.2.3   Pre-Processing of the Experimental Data

A sample of a chicken fillet containing a fan bone was scanned using a CMOS detector with a CsI(Tl) scintillator (Dexela1512NDT) [36]. The X-ray source was a microfocus X-ray tube with voltages of 40 and 90 kV. The piece of fillet was wrapped in a plastic bag and placed on a holder. This experimental setup imitates a top view similar to the typical data from a conveyor belt.

The same sample was measured with different exposure times to illustrate the impact of the detector noise. Fig. 2.3a shows a combined image computed according to the Eq. 2.5 with two projections acquired with the exposure time of 0.5 s. Fig. 2.3b is a plot of thickness dependency of quotient values based on the experimental data corresponding to Fig. 2.2b for the simulated data. Values of $M_2(x)$ are used instead of $L(x)$ since the thickness profile of the object is unknown. Pixels of the foreign object are marked with a different color to highlight that the noise variance is bigger than a difference between the sample and foreign object in spectral properties. Nevertheless, the bone can be located by a human expert based on the $R(x)$ distribution since the foreign object pixels are located near each other and form a region. If the same product is scanned with a higher exposure time, the level of statistical noise becomes lower and the foreign object is easier to locate. Fig. 2.4 shows the quotient image and quotient plot for the measurement with an exposure time of 5 s. The high noise case is more difficult, and it is the main focus of the next subsections.

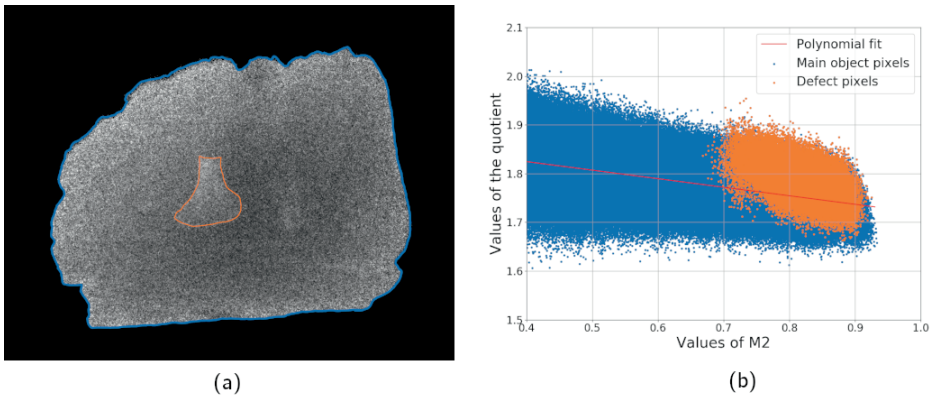(a)   (b)

Figure 2.3: Sample scan with low exposure (0.5s per projection): combined image computed according to Eq. 2.5 (a), and the dependency of its values on the single projection intensity (b).
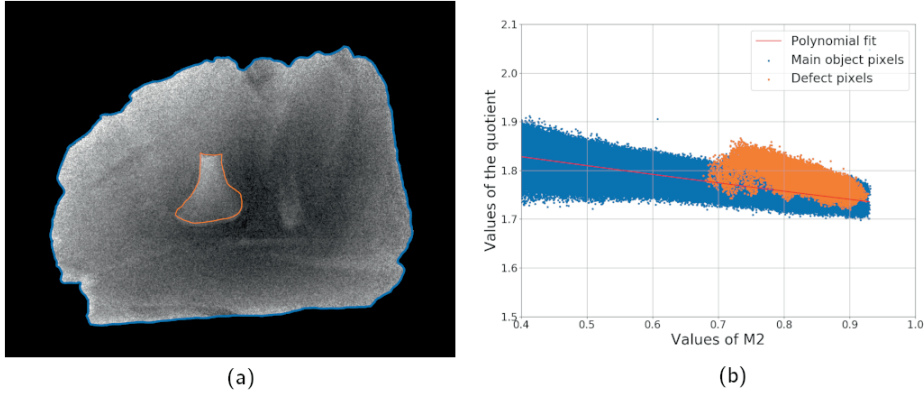
Figure 2.4: Sample scan with high exposure (5s per projection): combined image computed according to Eq. 2.5 (a), and the dependency of its values on the single projection intensity (b).

Even if the foreign object is visible to a human expert, its detection might still be non-trivial for a classical algorithm. Thickness dependency of the ratio values leads to natural gradients in the $R(x)$. The introduction of the corrected quotient $R'(x)$ helps to reduce this effect and makes the image easier to segment. The thickness correction is shown in Fig. 2.5a,b, which represent $R(x)$ and $R'(x)$ distributions. After this procedure, $R(x)$ is close to zero in most pixels corresponding to the main object. A deviation from zero might be caused by detector noise, systematic errors of the experimental setup, and different foreign objects. The presence of a foreign object changes a pixel value, and the difference depends on its thickness. The main task of the segmentation algorithm applied to Fig. 2.5b is to locate big clusters of nonzero pixels excluding noisy outliers. With a significant noise influence, foreign object location is difficult to perform on a pixel level. Thus, it is important to use spatial information.

The largest noise level in the $R'(x)$ distribution is usually found near the edges of the main object. In those regions, a quotient intensity is calculated as a ratio between two small numbers that leads to the high relative error. This effect can be reduced to improve detection accuracy. It can be assumed that the variance of the image values is mainly defined by the statistical noise and depends on the thickness. Therefore, a standard deviation for different thickness values should be computed from the data. The simplest approach to solve this problem is to divide the image into regions with similar M_2 intensity and to compute mean value and standard deviation for every subset. Each region is defined as a set of image points with the intensity values $M_2(x)$ belonging to a certain closed interval of values. If values of $M_2(x)$ start from a threshold value M_0 and size of the interval is Δ, the intensity interval corresponding to ith region can be written as $[M_0 + i\Delta, M_0 + (i+1)\Delta]$. Therefore, the region is a set of points x where $M_2(x) \in [M_0 + i\Delta, M_0 + (i+1)\Delta]$.

Then, the normalized corrected quotient $N(x)$ can be calculated as follows:

$$N(x) = \frac{R'(x) - \overline{R'_i}}{\sigma_i}, \forall x : M_2(x) \in [M_0 + i\Delta, M_0 + (i+1)\Delta], \qquad (2.8)$$

where $\overline{R'_i}$ and σ_i are mean value and standard deviation of $R'(x)$ computed for ith region.

Fig. 2.5c shows a normalized quotient based on the corrected image Fig. 2.5b.
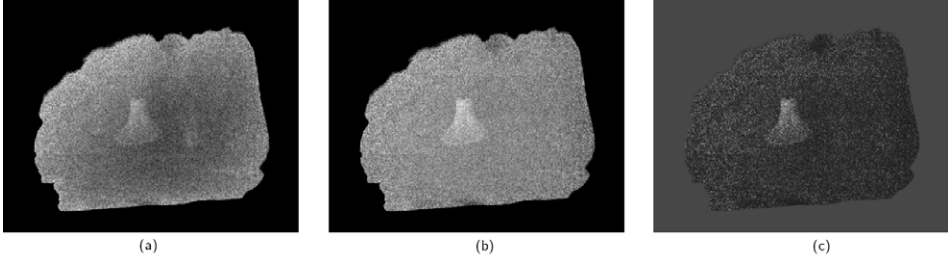
Figure 2.5: Stages of the scan data pre-processing: $R(x)$ computed according to Eq. 2.5 (a), $R'(x)$ defined by Eq. 2.7 (b), combined image after correction and normalization computed according to Eq. 2.8 (c). The sample is a chicken fillet with a fan bone scanned with an exposure time of 0.5s.

2.2.4   Active Contour Segmentation and Foreign Object Detection

Thickness correction pre-processing described in the previous section transforms two-channel dual-energy projections into a single image. High noise level prevents many algorithms from accurately segmenting the image. In this work, an active contour model is used to segment the image based on region properties without relying on intensity gradients. The Chan–Vese method [27] is a variational segmentation algorithm inspired by the Mumford–Shah model [111]. It separates two homogeneous phases in the image by minimizing the energy functional over the phase boundary and their average values:

$$F = \lambda_1 \int_{\Omega_1} (N(x) - c_1)^2 dx + \lambda_2 \int_{\Omega_2} (N(x) - c_2)^2 dx + \mu|\partial\Omega_1| + \nu|\Omega_1|, \qquad (2.9)$$

where Ω_1 and Ω_2 are regions segmented as an object and backgound, c_1 and c_2 are average pixel values in these regions, $|\partial\Omega_1|$ is the boundary length of Ω_1, and $|\Omega_1|$ is the area of Ω_1. In the case of a foreign object location problem, the background refers to the main object and the object refers to the foreign object. The minimization problem is solved by applying the level-set technique: phase boundary is defined as a zero-level of a level-set function. The values of c_1 and c_2 are recalculated on every step depending on the current phase boundary.

The segmentation outcome is implicitly controlled by the ratios between λ_1, λ_2, μ, and ν. The first two terms favor a similarity between pixel value and region average intensity regardless of the spatial properties. The last two terms mitigate the effect of noisy pixels on the segmentation. Low values of μ and ν transform the segmentation into thresholding with minimal removal of outliers (Fig. 2.6a). Different values of penalty weights lead to different boundary detections, noise sensitivities, and overall accuracies of the algorithm. Examples of such effects are shown on Fig. 2.6b,c. The biggest strength of the active contour approach is that parameters have an interpretation and can be related to the image properties, such as object intensities and noise distribution.

(a)   (b)   (c)
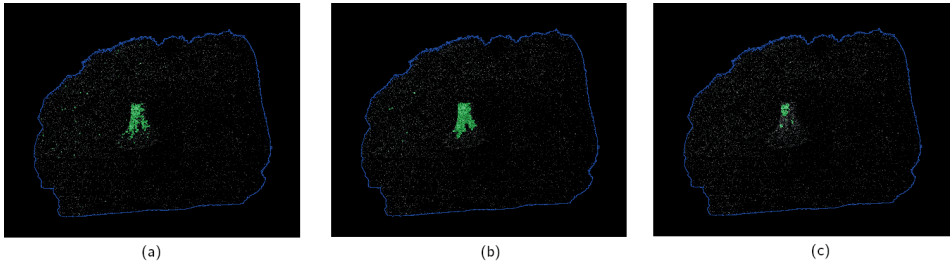
Figure 2.6: Different examples of the segmentation with different values of penalty weights applied to the quotient $N(x)$ image after pre-processing shown on Fig. 2.5c. The panel (a) shows a segmentation with low values of penalty weights $\mu = 1, \nu = 1$, where many noisy outliers are marked as bone fragments. This effect can be reduced by changing the weights as shown on panel (b) corresponding to $\mu = 5, \nu = 1$. High values of penalties, such as $\mu = 20, \nu = 5$, on panel (c) might lead to a safe segmentation excluding a significant fraction of the bone.

As a variational algorithm, the Chan–Vese method has a few implicit parameters that influence the iterative process. The initial state of the level-set function in the described implementation is defined by a simple thresholding

$$\phi(x) = \begin{cases} 1 & \text{if} \quad N(x) \geq T_{init} \\ 0 & \text{if} \quad N(x) < T_{init} \end{cases}, \tag{2.10}$$

where T_{init} corresponds to a significant deviation from zero selected manually based on the data. Thus, pixels with values higher than the threshold are likely to be part of the defect region. On every iteration of the segmentation algorithm, the level-set is recalculated to better minimize the segmentation energy. If the increment norm is smaller than a certain tolerance value, the algorithm converged. The iterative process is also stopped if it takes more iterations than a certain maximal number. These parameters mainly influence the speed and convergence of the method and define the final segmentation in case there are multiple local minima.

Accuracy of the segmentation can be evaluated if a ground truth (correct segmentation of the input) is known for every sample. F1-score is chosen as an

accuracy metrics and is calculated according to the following formula:

$$\text{F1-score} = \frac{\text{TP}}{\text{TP} + 0.5(\text{FP} + \text{FN})}, \tag{2.11}$$

where TP is the number of True Positive pixels (pixels of the foreign object that were correctly identified), FP—False Positive (pixels of the main object that were falsely classified as a foreign object), and FN—False Negative (pixels of the foreign object which were missed). This metric is commonly used in papers about foreign object segmentation. However, it does not evaluate performance on the samples without a foreign object.

The main challenge for the active contour segmentation lies in images without foreign objects. The neighborhood of a noisy pixel with a significant deviation from zero can be considered part of the defect region, even if there is no foreign object in the sample. This problem can be solved by adjusting penalty weighting coefficients μ and ν. If a region Ω_{noisy} with mean intensity N_{noisy} is considered part of the main object, the energy is increased by $(N_{noisy} - c_{main})|\Omega_{noisy}|$ since the mean intensities are different. This energy increment can be avoided if this region is classified as a foreign object. In this case, the energy is modified by penalty terms $\mu|\partial\Omega_{noisy}| + \nu|\Omega_{noisy}|$. The decision about including or excluding the region Ω_{noisy} is based on the ratio between these two terms. It is important to note that the Chan–Vese algorithm is an iterative method. Therefore, the result depends not only on energy terms but on the initial level-set, regularization parameters, and convergence speed, among other things. Nevertheless, penalty weights should be raised to a certain level to exclude noisy clusters in most cases.

If a noise level is sufficiently high, fine-tuning parameters for all types of inputs is a challenging problem. Significant noise fluctuations in the samples without foreign objects require penalty weights to be high. On the other hand, the accuracy of foreign object segmentation becomes worse since many pixels on the foreign object boundary are included in the main object. A post-processing procedure is introduced as an additional way to exclude noisy pixels from the foreign object region and to make the algorithm more robust. A segmented defect region can be divided into clusters of neighboring pixels. For each cluster, the mean intensity and size can be computed. Segmentation quality can be enhanced if a certain threshold on the cluster size is set and small clusters are ignored. The main reason for employing this strategy is the existence of noisy pixels with an intensity that is several times higher than the average foreign object intensity. If penalty weights are adjusted to exclude such pixels, small foreign objects might be excluded as well.

In the proposed methodology, post-processing is the removal of segmented clusters with a size lower than a certain number of pixels. If there are no clusters after pre-processing, the sample is marked as normal. Otherwise, it is considered that the sample contains a foreign object. For every sample in the experimental dataset, it is known whether it contains a foreign object or not. Thus, it is possible to compute a confusion matrix and F1-score for the inspection procedure. In this case,

accuracy is measured on a sample level, unlike segmentation precision. These metrics are more important for algorithm performance evaluation since they include all possible cases. If the segmentation is fine-tuned to achieve the best segmentation accuracy, it might become too sensitive to noise. Therefore, due to noise fluctuations, it classifies normal samples as containing a foreign object.

2.2.5 Deep Learning Segmentation and Foreign Object Detection

Deep learning provides a more general and less interpretable approach to image segmentation. A deep learning segmentation method is a function $f_\phi : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X} \subset \mathbb{R}^{m \times n}$ is a set of input images and $\mathcal{Y} \subset \{0, ..., L\}^{m \times n}$ is a set of segmented images. ϕ are parameters of the algorithm, and the main practical difference between deep learning and classical methods is the number of parameters. While the Chan-Vese method is characterized by a small number of parameters, the Mixed-Scale Dense Convolutional neural network (MSD) [117] used as a deep learning method in this chapter has thousands of parameters.

The MSD architecture shares many concepts with more standard convolutional neural networks [122, 129]. However, a number of modifications are made to improve the performance and reduce the number of parameters. An MSD network consists of layers (Fig. 2.7). Each layer takes an image as input and applies multiple operations to it to compute a feature map - an intermediate image. The first layer is applied to the input image of the network, and other layers receive previously computed feature maps as input. Finally, the last layer computes the resulting segmentation.
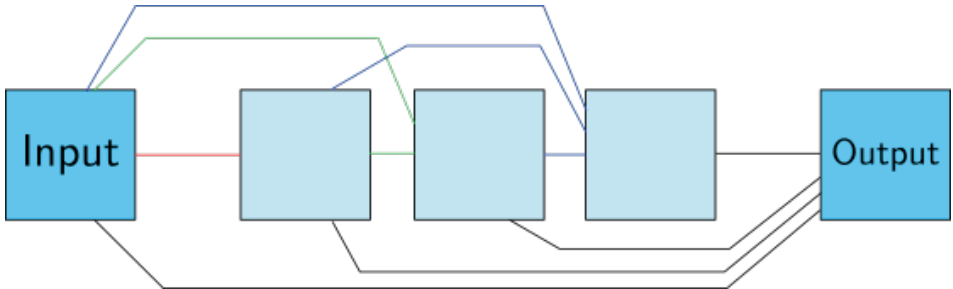
Figure 2.7: An example of an MSD network. Light blue squares represent feature maps. Lines of different colors represent convolutions with different dilation values. Each feature map is computed based on all previous feature maps, and all of them are used to produce the final image.

Standard convolutional neural networks use a convolutional filter to process the image. Each pixel of the output image is computed as a sum of pixels of the input image weighted by the convolutional kernel. The kernel is usually significantly smaller than the image, and all kernels are parts of the parameter set ϕ. MSD

networks use dilated convolutions based on sparse and large kernels. With dilated convolution, networks are able to quickly capture large-scale features of the image. Furthermore, all feature maps in an MSD network are of the same size as the original image. As a result, each subsequent layer can use all previously computed feature maps.

While the structure of an MSD network is defined by a number of meta-parameters (width, depth, dilations, etc.), the values of its parameters (convolution kernels, weights, and constant biases) need to be determined. The parameters are computed by training a network — solving an optimization problem. The network is applied to a set of images with known ground-truth segmentations. Network parameters are optimized to minimize the difference between segmentations produced by the network and the ground-truth.

Due to the large number of parameters, it is possible to overfit deep learning methods, so that their performance on the training data is significantly better than on different data. Therefore, it is important to separate training data and test data, which are used to measure accuracy of the network [48]. While the Chan-Vese method is applied to the full available dataset, for MSD applications the dataset is split between training, validation and testing. Training and validation subsets are used to find optimal network parameters, and the test subset is used to characterize the accuracy of the network.

Accuracy of the MSD network is characterized using the same metrics as for the Chan-Vese method: F1-score for segmentation and F1-score for the foreign object detection. The same post-processing is used for MSD segmentations to determine whether there is a foreign object. An additional advantage of MSD is that it can be applied to other inputs, not only normalized corrected quotients $N(x)$. MSD segmentation is additionally tested on the 2-channel image (M_1, M_2) directly and on both channels separately.

2.3   Results

2.3.1   Dataset Description

The thickness correction procedure was tested on the scans of chicken fillets on a conveyor belt. The images were acquired with a line detector since these are commonly used in industrial setups. The majority of fillet samples contained a bone that should be detected as a foreign object. Every sample was scanned four times with different positions on a belt. There are 100 images with a fan bone, 100 images with a large rib bone, and 96 images with a small rib bone. In 192 scans, the chicken fillets did not contain a foreign object. These types of bone differ by average size, form, and typical position in a fillet. The dataset was semi-manually segmented to create an approximate ground truth for accuracy estimation.

The biggest challenge of the dataset is the small difference in attenuation curves between the chicken fillet and bone. Both are organic materials and do not contain a signicant concentration of heavy metals with a visible K-edge. Therefore, the

attenuation curves do not have spectral features that can be easily detected, such as K-edges in contrast agent detection. However, bones contain a significant amount of calcium that is not present in meat. This leads to a difference in spectral properties that might be of the same order of magnitude as noise if the foreign object is small enough.

The ground truth for the dataset was used to calculate the average properties of different bone types. Table 2.1 shows the comparison between pixel values after thickness correction and normalization. Bone size refers to the average area of a bone of the corresponding class measured in pixels.

Table 2.1: Comparison of mean pixel intensity and foreign object sizes for different types of bones

Bone class	Pixel value	Bone area, px
Fan bone	3.7 ± 2.6	370 ± 120
Large rib bone	3.1 ± 2.5	290 ± 70
Small rib bone	3.0 ± 2.6	160 ± 35

2.3.2 Thickness Correction

The described pre-processing was applied to the dataset to compute $N(x)$ distributions. A thickness dependency for every sample was interpolated by a polynomial of the second degree. If an absorption rate was lower than a certain threshold equal to 0.2, a pixel was ignored both in interpolation and division. The size of the intensity bin Δ was set to 0.1. Fig. 2.8 shows projections of different samples from the experimental dataset, the corresponding $N(x)$, the ground truth, and the thickness dependency plot.

The dataset contains a variety of examples with different types and sizes of bone fragments and normal samples without foreign objects. Cases 2.8(a)–(d) show the effect of thickness correction on samples with a foreign object. The foreign object might be visible on a single projection, even if it has a small size, such as a shattered rib bone in sample 2.8(d). However, the contrast on a single projection significantly depends on the exact location of the foreign object. If the main object form causes intensity gradients near the foreign object, it might be missed without additional information. The main benefit of thickness correction lies in removing intensity changes corresponding to the main object and highlighting the defect location. The thickness dependency plot shows that quotient values corresponding to the foreign object often have a similar deviation from zero as noisy outliers. This corresponds to the low exposure scanning procedure discussed previously.

Example 2.8(e) illustrates the main advantage of using thickness correction for DEXA data. Both projections contain a region with a well-visible border that is not a foreign object. However, the correlation between images does not correspond to a material with attenuation properties different from the main object. In the corrected quotient, this region is the same as other parts of the sample.

The thickness dependency plot does not contain any significant outliers as well. Such intensity changes might appear on samples with and without foreign objects. Therefore, it is crucial to remove them in order to prevent a high false positive rate.

Some projections in the dataset contain systematic experimental effects that can look simialr to a foreign object after thickness correction. In sample 2.8(f), a set of pixels near the border has a high $N(x)$. However, no bone is present in the object in this case. Automatic data acquisition might lead to misalignment artifacts, small movements of the object between scans, and a change of shape. These artifacts have $N(x)$ values similar to the real foreign objects and might be recognized as such.

2.3.3   Accuracy of the Chan-Vese method

$N(x)$ images were used as an input for the Chan–Vese segmentation algorithm. The method implementation was based on the C++ code by Pascal Getreuer [56]. Furthermore, a Python wrapper was written and used as an interface. In the results, the default algorithm parameters are fit weights $\lambda_1 = \lambda_2 = 1$, time step $dt = 1$, convergence tolerance $tol = 10^{-4}$, maximal number of iterations $N_{max} = 200$, heavyside regularization $\epsilon = 1$, curvature regularization $\eta = 10^{-8}$, and initial level-set threshold $T_{init} = 5$. The accuracy of the algorithm is studied for different values of μ and ν since they significantly influence the outcome and have a geometrical interpretation. The post-processing pixel count threshold is set to 30 pixels. This means that defect clusters containing fewer than 30 neighboring pixels were removed from the final segmentation.

In this subsection, the segmentation accuracy is estimated on a pixel level for the samples containing a foreign object. For every object, the resulting segmentation is compared with ground truth to count the number of true positive, false positive, and false negative pixels. F1-score is calculated according to Eq. 2.11. The values of the F1-score are shown on Fig. 2.9a for different combinations of penalty weights. The value of the metric is averaged over all images with foreign objects. The best segmentation accuracy is achieved with $\mu = 14$ and $\nu = 2$.

As explained in Results, the dataset contains different types of bone as a foreign object. The values in Fig. 2.9a are averaged over all foreign object types. Thus, it does not show how defect class affects segmentation accuracy. The corresponding figures for every type of bone are shown on Fig. 2.10. The dependencies of the F1-score on penalty weights are similar for all foreign objects. Every bone class has a combination of Chan–Vese parameters that achieves the best segmentation accuracy for that defect type, and these parameters might be different from each other. However, the best instances for a single defect class also perform well for the whole dataset, as shown in Table 2.2. Thus, all types of foreign objects can be segmented with the same algorithm settings. Different ratios of bone types in the dataset would affect the algorithm performance but not significantly.

As highlighted previously, the detection procedure should be evaluated on the samples without a foreign object. The decision-making based on the segmentation
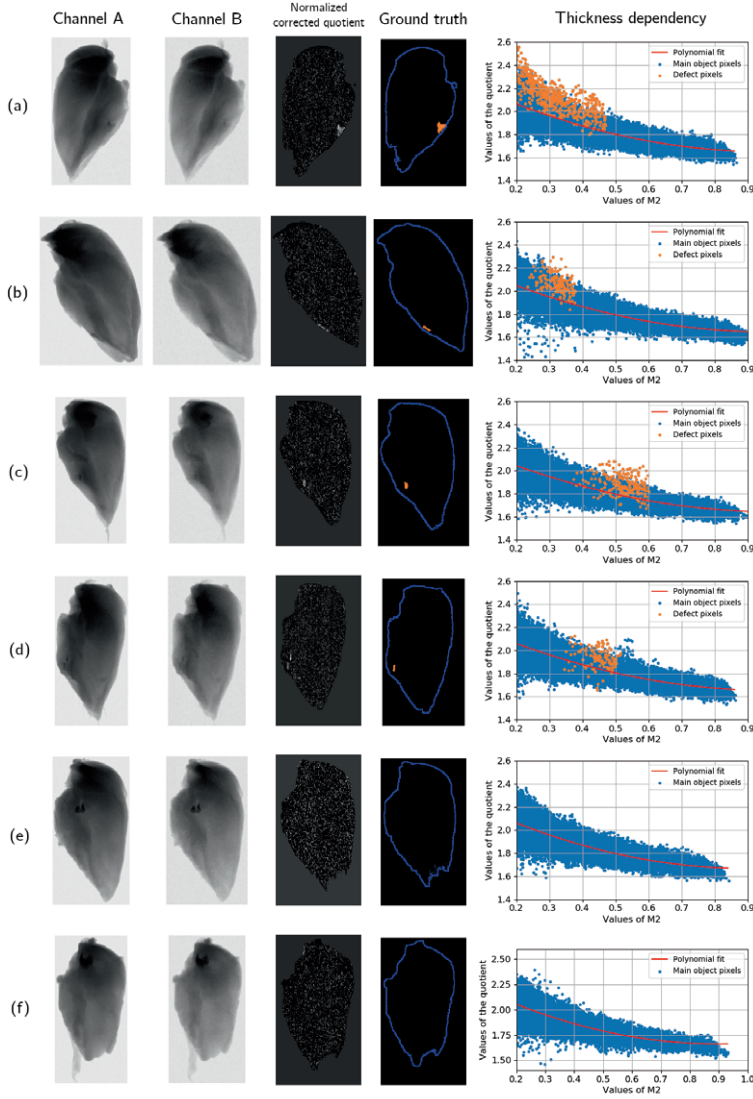
Figure 2.8: Different samples from the experimental dataset. For every object, two projections acquired with different spectra, the $N(x)$ distribution, the segmented image, and the thickness dependency plot are shown. Sample (a) contains a fan bone, sample (b) has a large rib bone, cases (c,d) show small rib bones, and samples (e,f) do not contain foreign object. The boundaries of the samples are approximately drawn as blue curves, but they are not used during the inspection procedure. The defect location is marked in orange and corresponds to the ground truth images from the dataset. Ground truth in sample (d) is partially wrong and does not include the second part of the shattered bone.
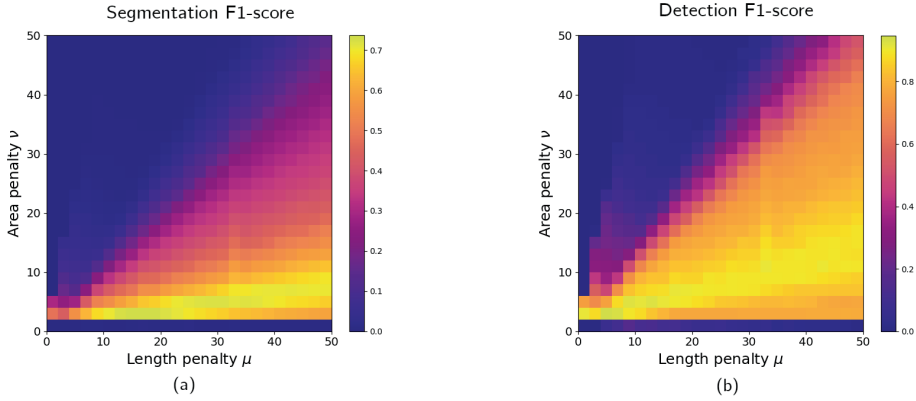
Figure 2.9: Dependency of F1-score on length penalty μ and area penalty ν for different tasks: image segmentation (a) and foreign object detection (b). For segmentation, the F1-score is computed using a ground truth segmentation known for every sample and averaged over all images from the dataset containing a foreign object. For detection, the metric is calculated on a sample level for the entire dataset consisting of the objects with and without a bone.

Table 2.2: Comparison of the best Chan-Vese parameters for different classes of foreign objects. F1-score is separately calculated for all images with the same bone class and for all samples with a foreign object.

Defect class	μ	ν	Single class F1	F1 for all defect classes
Fan bone	20	2	77%	72%
Large rib bone	14	2	75%	74%
Small rib bone	14	2	70%	74%

and post-processing was tested on the whole dataset: 296 images with different types of bone and 192 images without a foreign object. The test results contain the number of images with a bone where a presence of foreign objects was detected and the number of boneless images that were correctly identified as empty. These values correspond to the true positive and true negative cases. The algorithm accuracy was evaluated using the F1-score.

Fig. 2.9b shows the dependency of F1-score on Chan–Vese energy equation penalties μ and ν. High accuracy (more than 90%) can be obtained with multiple combinations of parameters, and the best value of F1-score was achieved with $\mu = 4$ and $\nu = 2$. The confusion matrix for this instance of the inspection procedure is shown in Table 2.3. The algorithm correctly marks 97% of the normal samples as not containing a foreign object. Chicken fillets with a bone were successfully identified in 92% of cases.
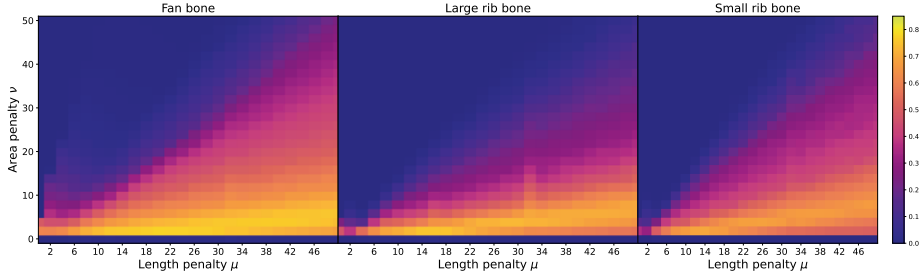
Figure 2.10: Foreign object segmentation F1-score for every class of defect in the dataset: fan bone, large rib bone, and small rib bone. The F1 value is averaged over all samples with a corresponding defect type.

Table 2.3: Confusion matrix for the inspection procedure with Chan-Vese parameters $\mu = 4$ and $\nu = 2$

		Predicted	
		Defected	Normal
Actual	Defected	271	25
	Normal	5	187

2.3.4 Accuracy of the MSD method

For MSD segmentation tests, the PyTorch implementation from [69] was used. All MSD networks had the same structure: a depth of 15 layers, a width of 1, dilations in the range from 1 to 10. Dice's loss was used during training. The number of input channels was different depending on the input: 1 – for channels A, B, and the quotient, 2 – for a combined input of A and B. Each network was trained until convergence for 500 epochs. To mitigate random effects affecting training (e.g., random initialization of weights), 10 instances of the same network with the same training data were trained independently, and their average F1-score was used as a metric of accuracy. The dataset was split: 40% for training, 10% for validation, 50% for testing. Images without foreign objects and with different classes of foreign objects are evenly spread between these subsets.

Table 2.4 compares F1-scores of segmentation for different foreign objects and types of input. The networks are optimized for all classes of foreign objects at the same time, and F1-scores for different classes are achieved with the same network parameters. The MSD network applied to the normalized corrected quotient $N(x)$ outperformed Chan-Vese segmentation. For each foreign object class, it achieved a higher value of F1-score. Furthermore, similar accuracy was achieved without performing the pre-processing. The network applied to a 2-channel image was as accurate for fan bones and large rib bones, and there was an improvement in the segmentation accuracy for small rib bones. While the proposed Chan-Vese

method requires dual-energy images, the MSD network was able to segment bone fragments even on separate images. Channel A leads to a better accuracy than Channel B, and both are outperformed by a combined 2-channel image.

Input	M_1	M_2	(M_1, M_2)	Quotient $N(x)$
Fan bone	62%	33%	87%	89%
Large rib bone	54%	23%	86%	84%
Small rib bone	64%	37%	85%	77%

Table 2.4: Comparison of the segmentation accuracy of the MSD network with different types of input images.

The MSD network performed even better when evaluated based on the accuracy of foreign object detection. It was possible to achieve almost 100% F1-score for detecting foreign objects with both the 2-channel image and the quotient. The corresponding confusion matrices are shown in Table 2.5 and 2.6. Furthermore, the detection method based solely on Channel A was able to perform the detection with 96% accuracy (Table 2.7). Thus, with a deep learning method, even a single image could lead to a higher detection rate than what could be achieved with an active contour method and more data. The results with Channel B were worse than the Chan-Vese baseline, only 76% F1-score was achieved.

Table 2.5: Confusion matrix for the inspection procedure based on the MSD segmentation of $N(x)$

	Predicted	
	Defected	Normal
Defected	151	1
Normal	0	100

Table 2.6: Confusion matrix for the inspection procedure based on the MSD segmentation of (M_1, M_2)

	Predicted	
	Defected	Normal
Defected	152	0
Normal	0	100

Table 2.7: Confusion matrix for the inspection procedure based on the MSD segmentation of M_1

		Predicted	
		Defected	Normal
Actual	Defected	145	7
	Normal	4	96

2.4 Discussion

2.4.1 Thickness Correction

The thickness correction pre-processing was performed on the experimental data and does not rely on prior knowledge about the samples. The dependency of the quotient value on the thickness of the main object material is estimated as an average function for pixels of the projection. The theoretical foundation of the pre-processing implies that this function can be predicted with sufficient knowledge of the inspection system. As a result, it can be used to construct more precise measurement systems and to achieve better contrast between the main and foreign objects. However, in the scope of this chapter, a heuristic approach was chosen to show the applicability of this method to a wide range of data.

One of the major downsides of $R(x)$ images is the resulting significant level of noise. If noise in single-energy projections follows a Gaussian distribution, the $R(x)$ noise has a ratio distribution. Thus, significant outliers from the mean value are more likely than with the Gaussian. The exact properties of the resulting distribution depend on the mean value and variance for both original distributions. In practice, this means that high noise appears frequently, especially in the boundary regions of the image.

Several approaches were considered to mitigate a high noise level. First, normalization is applied to the corrected quotient $R'(x)$. On the boundary, large deviations from the zero level are divided by the significant variance value. A downside of this procedure is that variances are computed under the assumption that the distribution is Gaussian. Second, not all boundary regions are used for the $N(x)$ computations. If a pixel value (absorption rate corresponding to the pixel that can be computed after darkfield and flatfield corrections) is less than a certain threshold, a pixel is considered part of the background and ignored. This boundary cut might remove some parts of the bone from the image if it is located near the main object border. Nevertheless, a smaller number of noise outliers leads to a better segmentation in general.

In this work, the $N(x)$ images were used as an input for a segmentation algorithm. However, it does not mean that other derivatives of two projection channels should not be considered. In the data-driven approach, the quotient is used as a simple combination of two channels that has an additional meaning for a monochromatic beam. With more knowledge about the inspection system and

product materials, a better combination of the two channels might be constructed. The main focus of the pre-processing procedure is to remove thickness dependency, additional steps can be considered to improve foreign object contrast. Normalization of the $R'(x)$ can be viewed as an implicit introduction of the Gaussian noise to the active contour model.

2.4.2   Active Contour Segmentation

The Chan–Vese method operates well even with noisy data, and a high noise level is common for the conveyor belt product inspection. The energy that is minimized over two phases in the image is a formal way of defining a connected cluster of points in the presence of high noise. Thus, the segmentation algorithm determines bone borders consistently based on the objective criteria. At the same time, the ground truth made by a human operator might be more subjective. When a discrepancy between the segmentation and ground truth occurs, it can be caused by many factors. On one hand, the Chan–Vese method might not converge or reach a local minimum, and thickness correction might produce a very noisy input. On the other hand, the ground truth in a single sample can be inconsistent with other data.

The active contour models contain several parameters that do not have a physical meaning. The maximal number of iterations, convergence tolerance, and time step influence the speed of the method and resulting segmentation. The optimal choices of these parameters balance computational speed and algorithm accuracy. For the detection procedure, it is crucial to obtain the result as fast as possible. Therefore, a large time step and low tolerance can be considered.

Fig. 2.9 shows that good accuracy can be achieved with a range of algorithm parameters. The best penalty weight pair does not lead to a significantly better F1-score than its neighborhood in the parameter space. Thus, a search for the optimal inspection settings converges quickly. In these plots, the grid step for μ and ν was set to 2. A smaller step was not chosen to prevent overfitting to the experimental data. The form of the F1-score dependency on μ and ν implies that a similar result can be achieved with a broader dataset.

2.4.3   Foreign Object Detection

The active contour model in the described methodology is defined for two homogeneous phases according to Eq. 2.9. This means that the method is well-suited for the cases when a single foreign object is present in the main object. Two foreign objects of significantly different classes (i.e., bone and a plastic piece) might be segmented incorrectly since they should be marked as a defect phase, but their average intensities vary significantly. This problem can be solved with a change of the active contour energy and the introduction of several level-set functions. However, in practice, it is unlikely that multiple foreign objects appear on the sample since a single foreign object is expected to appear rarely. At the same time, the introduction of more defect types in the energy equation might decrease

the detection accuracy even more since the majority of samples contains no foreign objects. In the experimental dataset, some images contain a shattered bone, and both pieces are segmented properly since they correspond to the same defect class.

The described methodology is not limited to the food industry. The main novelty of this inspection procedure is the thickness correction procedure. The effect of thickness on quotient values is relevant for any dual-energy single projection measurement. It is not necessary if the foreign object has significantly different attenuation properties (e.g., detection of metal pieces in the luggage). Nevertheless, the thickness is important to account for if the data contain high noise level and low-contrast foreign objects.

The detection task with optimal parameters achieves 95% accuracy on the experimental dataset. Out of 192 samples, 5 samples were misclassified as containing a bone. Some of them can be attributed to systematic experimental errors, such as misalignment or sample deformation. In other cases, a noisy cluster is not segmented as a foreign object if the convergence tolerance and the maximal number of iterations are changed. In 25 samples out of 296, a bone was not detected. The main factor leading to false negative cases is detector noise that requires strict length and area penalties for the segmentation. Furthermore, with a high exposure time, some materials present in the factory environment, such as poultry fat, can be recognized as a foreign object.

It is important to note that the best Chan–Vese parameters for defect detection are different from those that provide the best segmentation accuracy. For a binary outcome, it is not important how precisely the bone is located on the image. The segmentation method often marks only a central part of the bone and ignores its boundary. At the same time, the best detection parameters lead to better performance in difficult cases: the presence of small bones that are indistinguishable from noise and significant noise fluctuations that look similar to small bones. Furthermore, the execution time is lower for the detection procedure, which is important in the industrial environment.

2.4.4 Deep learning segmentation

The Chan-Vese algorithm was chosen as a segmentation method because it performed better than other well-known techniques, such as thresholding and watershed algorithms. However, the MSD network was able to achieve significantly higher accuracy despite not being a state-of-the-art deep learning method. The comparison between Chan-Vese and MSD should be analyzed qualitatively rather than quantitatively because the methods were not compared on exactly the same test data (a subset of data had to be used for training). Nevertheless, supervised deep learning is generally considered more accurate than classical methods. Furthermore, applying neural networks (not training) is usually less computationally expensive than executing an iterative active contour segmentation. This computational efficiency is especially important for in-line foreign object detection.

Several advantages of Chan-Vese segmentation compared to deep learning methods need to be highlighted. Unsupervised methods can be used to prepare ground-

truth for supervised deep learning methods. While a Chen-Vese segmentation is not perfectly accurate, it could be used before the manual segmentation by a human expert to accelerate the process of data collection. The Chan-Vese segmentation is interpretable. With the thickness correction pre-processing, every pixel value can be computed using the physical properties of materials, tube parameters, and detector model. The penalty weights can be interpreted as a balance between the defect signal and the noise level of the image. This information can be used to improve the scanning protocol, to evaluate the cost efficiency of different detectors for a certain task, and to estimate the size limits of the detectable foreign objects.

It is important to note that the MSD network did not require the proposed pre-processing to achieve accurate segmentation and foreign object detection. Similar and in some cases better results were also achieved with a 2-channel image as input. Due to the low interpretability of the MSD network, it is challenging to provide a fundamental explanation of this effect. However, it is important to note that the pre-processing procedure could have been imitated in the layers of the neural network. The pre-processing is data-driven, no prior information about the chicken fillet and bone fragments was used. Similar quotient corrections could be, at least approximately, performed with convolutional layers. Furthermore, it is possible that the quotient alone is a worse input for a deep learning methods than a raw 2-channel image. The pre-processing of the quotient introduces assumptions, such as a polynomial fit, and raw data contain the full acquired information about the studied object.

2.5 Conclusion

The thickness correction pre-processing proposed in this work enhances the detection of foreign objects in the dual-energy projections of conveyor belt samples. The described methodology does not rely on a good contrast between a foreign object and the main object on a single projection. Instead, it utilizes the difference in attenuation properties that can be detected with a dual-energy acquisition. The active contour segmentation algorithm allows for analyzing data with a significant noise level if a proper energy model is chosen. The performance of the inspection was evaluated based on the ability to distinguish samples with and without a foreign object. It was shown that 97% of samples without a foreign object can be correctly identified while maintaining a 95% accuracy of the foreign object detection on the experimental dataset. The proposed approach does not require prior knowledge about the samples, and necessary material properties were extracted directly from the projections. The methodology was tested on bone detection in chicken fillets. However, the thickness correction procedure does not rely on any specific properties of this problem and can be extended for other foreign object detection tasks. Tests with a deep learning image segmentation method showed no advantage of using a thickness-corrected input over raw data.

3

CT-BASED IMAGE GENERATION

X-ray imaging is widely used for non-destructive and high-throughput product quality control of agricultural [66, 84, 99] and manufacturing [102] products. Because of the unique penetration properties of X-rays, they are well suited for detecting unwanted "foreign objects" that have strong density contrast in a single X-ray view. In this chapter, we use the term *foreign object* (FO) in a general manner, referring to the presence of undesirable structures within a certain base product. Examples of such objects include insect infestation in grain [42], bones in fish [103] and chicken [31] fillets, but also structural damages such as fractures in light-alloy casting [52] and porosity in welds [46]. On a single X-ray view, different object features overlap with each other and make image analysis complicated even for a human expert. While clearly delineated, dense foreign objects could be detected by an observer or a simple algorithm, defects with a complex shape or low attenuation contrast with surrounding materials require more sophisticated approaches, as demonstrated in Chapter 2.

Deep convolutional neural networks (DCNNs) are the state-of-the-art methods for complex image analysis tasks. The previous Chapter has shown that a supervised [119] deep learning method could reach an almost perfect foreign object detection rate depending on the task. However, high accuracy is achieved by learning from a sizeable set of high-quality, annotated training images. While training data for standard computer vision tasks are readily available, high-quality

annotated X-ray image data is cumbersome to obtain for industrial quality control applications. Moreover, the combination of changes in the structure of products and foreign objects from one batch to another can render the use of deep learning approaches infeasible, as the amount of data required for retraining the network is prohibitively large.

In this chapter, we present a simple yet powerful approach that overcomes these limitations, providing a fast, data-efficient deep-learning strategy for real-time detection of foreign objects in industrial products with high-throughput. Our approach relies on CT-scanning of a few – or even just one – product to create a basic digital product model that can be used for image simulation. By then modifying the CT volume – also including a variety of digital foreign object configurations – and computing artificial X-ray projections of these generated structures, a large training dataset is artificially created that provides an extensive sampling of the combinations of base and foreign objects.

Our experimental results show that although our training data is generated using a simple physics model for X-ray simulation, the model trained on such artificial images can be used effectively on real-world X-ray data, providing accurate and real-time foreign object detection from single-image X-ray views. As a source of real X-ray images, we use datasets containing X-ray projections of pieces of modeling clay with pebble stones inside [162], and also X-ray CT scans of avocado fruits. We formulate a foreign object detection problem (later referred as FOD) that should be solved with a single X-ray image, and the correct prediction (ground-truth) is defined based on the properties of studied samples. This formulation corresponds to the problem of detecting (rather than segmenting) a bone fragment in a chicken fillet described in Chapter 2. For both modeling clay and avocado, we present two generation strategies that require a different amount of expert knowledge. The basic modification strategy uses a limited number of transformations and reduces the amount of real data acquisition by roughly 50% while preserving the model's accuracy. We also present a more complex generative model based on a CT scan of just one object that achieves the same accuracy level that would otherwise require tens of samples. The generation of artificial data gives the user full control over sample features in the training set, improving detection accuracy when working with small real-world datasets.

3.1   Related work

X-ray imaging is successfully applied to foreign object detection in a variety of applications. Fish bones could be detected with an accuracy of 99% using conventional image processing and feature extraction [103], and CNNs can segment bones with 75% per-pixel accuracy [140]. Different stages of pest infestation can be distinguished in soybeans, reaching 86% accuracy of uninfested sample classification [28]. In poultry, X-ray projections are often combined with laser imaging systems to detect bones in fillet samples with more than 95% accuracy [31]. Manually defined features of X-ray images can be used to achieve 85% accuracy of welding

defect classification with 5 types of defects [46]. CNNs can provide mean average precision of 92% for localization of casting defects[52].

Depending on the specific application, the foreign object detection problem can be formulated in different ways. The direct approach is to *detect* whether a foreign object is present in the X-ray image. The outcome could be either binary (yes/no) or from a discrete set of options (multiple FOs that might be present). Additionally, FO detection can be framed as a binary segmentation, where every pixel of the image should be classified as a defect or not. Alternatively, a *semantic segmentation* can be performed by computing the full clusters of pixels containing a single FO.

Several authors have proposed to apply deep learning approaches to FO detection in X-ray images [52, 86, 146]. Such approaches could achieve high accuracy if a significant amount of annotated training data is provided. The data requirement can be a key obstacle in practical use since they must represent a broad set of possible base objects and foreign objects in different locations. For a segmentation task, the annotation of a large amount of data becomes an additional problem requiring a human expert. The approach in [146] required 180 pear scans, and 51 volumes were segmented manually. The GDXray dataset [104] used in [52] contains 2727 X-ray images of automotive parts and includes annotations of bounding boxes of defects. For the case study in [86], a dataset of 4725 X-ray images with ground-truth for binary segmentation was used.

In [161], it was proposed to perform CT scans of objects to generate data for foreign object detection (FOD) in X-ray images. Every scan contains a large number of individual projections that can be used as inputs for FOD. A possibility to perform volume reconstruction and create a 3D model of the sample can be used to generate ground-truth automatically. With this methodology, imaging of thousands of objects can be replaced with tens of CT scans for certain FOD problems, such as fruit inspection. However, reaching a sufficient number of scans might be infeasible if the problem involves multiple FOs. The base objects and all FOs need to be represented by a variety of data, and every CT scan captures a single combination of the base object with a particular FO. Thus, the training data would require all possible pairs that cover a variety of base objects combined with every variation of each FO, leading to a combinatorial explosion.

In many fields, the possibility to create artificial data is considered to overcome the challenge of acquiring large datasets. Monte-Carlo simulations using 3D models of objects of interest were proposed for a variety of manufacturing defect detection problems [102], luggage inspection [61], breast X-ray imaging [41]. While this approach could generate realistic data, it requires detailed knowledge of the X-ray imaging setup (source and detector properties) and objects that are supposed to be studied (3D model and material properties). Several studies [140, 145] indicate that realistic synthetic data can be created with simpler models, such as Beer's law, and use these data to train accurate deep learning models.

3.2 Methods

Our approach for creating artificial X-ray images requires at least one CT scan of a product and generates new images by changing the volume of the sample. This methodology is illustrated in Fig. 3.1. A set of X-ray projections can be converted into a sample volume using CT reconstruction algorithms. The volume is then segmented and deformed to automatically generate a variety of artificial volumes. Finally, X-ray images are computed based on artificial volumes using a forward projector.
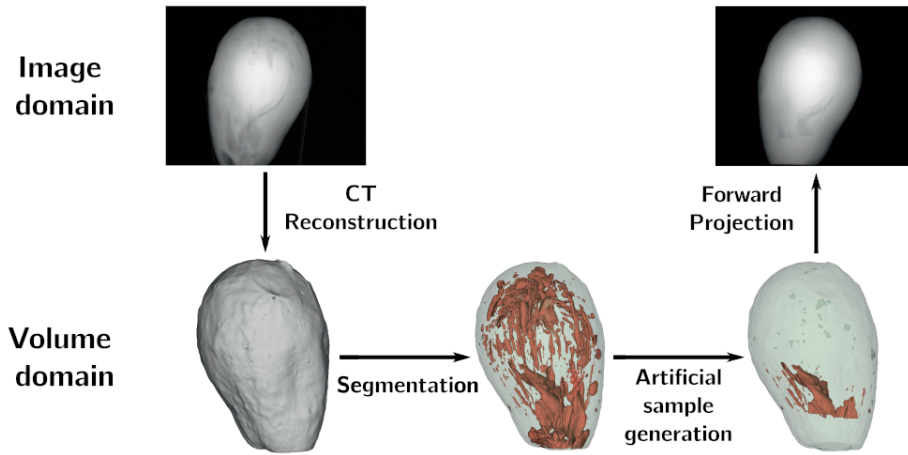
Figure 3.1: Proposed approach to data generation. All augmentations are performed on volume level and then forward projected to get an X-ray image.

In X-ray CT, the object **reconstruction** produces a 3D distribution of the attenuation coefficient inside the inspected volume. Every voxel of the reconstruction corresponds to a small region of the object, and the attenuation coefficient is mainly defined by the properties of the matter there. Furthermore, the reconstruction contains artifacts (e.g. beam hardening, noise, rings) induced by the difference between the model implied by the reconstruction algorithm and the actual physical imaging process.

The next step is a **segmentation** of the reconstructed volume. We consider the sample as a combination of a limited number of materials. In a segmented volume, every voxel corresponds to a single material. Both the base object and FO could consist of multiple materials, and on segmentation they can be separated from each other. Additionally, the attenuation coefficient of every material is extracted from the data, by computing an average intensity of reconstructed volume using segmentation as a mask. Thus, any segmented volume can be converted back to

the 3D distribution of attenuation using average intensities of materials.

Artificial sample generation is performed to create new object volumes similar to the real-world samples. Deep learning models construct a feature space based on the training data, and a successful generation strategy should provide a diverse range of objects to train an accurate model. The main challenge lies in identifying which differences between samples are important and how to make a representative set that covers a variety of possible products that might be encountered by a deep learning model.

The construction of a representative dataset requires the introduction of prior knowledge about the objects - *coverage*. It provides constraints in the data space and helps to represent every sample as a set of parameters in the configuration space. As a simple example (Fig. 3.2a), consider an object structure where a unit solid sphere (main object) contains a small ball (foreign object). Due to the symmetries of the objects, every possible sample in this set can be represented as a point in the configuration space with four dimensions, namely the radius of the inner ball and three coordinates of the vector connecting the centers of the objects (subspace of parameters is shown on Fig. 3.2b). A straightforward generation algorithm for this dataset would evenly sample all possible angles, distances and sizes of the foreign object. Not all parameters are important to produce a variety of images. For example, different polar and azimuthal angles are equivalent to different views of the same object.

(a)   (b)
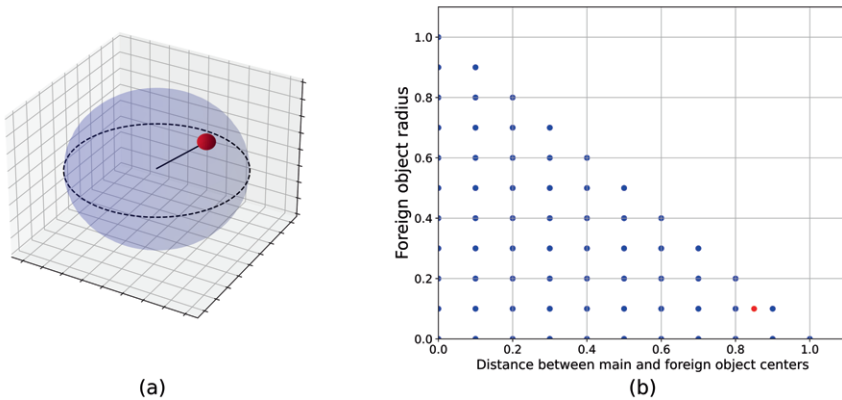
Figure 3.2: Example of FOD problem where a small ball is located inside a bigger one. Each sample can be described using four parameters, namely radius of the inner ball and three coordinates of the vector connecting centers of the objects. Subfigures: one possible sample of this dataset (a); subspace of configuration space corresponding to the fixed polar and azimuthal angles (b). The configuration shown on subfigure (a) is marked as a red dot in this subspace.

In real-world applications, the base object and FOs are usually more complex and cannot be described with a small number of parameters. Some problems,

such as detection of cracks during casting or inspection of canned food, have a
well-defined main object structure that can be used as a constraint and simplify
parameterization. We aim to formulate a more general generation strategy that
does not use such constraints. Instead of attempting to map a low-dimensional
configuration space and sample objects from it, we propose to apply deformations
to the known real volumes. For a particular deformation model with different
possible parameters, the output volumes cover a region around the original volume
in the data space. We assume that with a variety of deformations we can cover
enough data space to train an accurate deep learning model and validate this
assumption in the experiments.

Finally, the **forward projection** operator can be used to create artificial pro-
jections of the modified volume. The proposed methodology uses a simple ap-
proach to forward projection, in which X-ray attenuation follows Beer's law, and
the image is calculated by computing line integrals. This method does not take into
account many experimental effects, such as a polychromatic spectrum of the tube
and detector noise. We additionally apply post-processing to the noiseless projec-
tions in order to make them more similar to real data. Mixed Poisson-Gaussian
noise is applied to every pixel according to

$$I_{noisy} \sim \frac{1}{\lambda}\mathcal{P}(\lambda I_0) + \mathcal{N}(0, \sigma^2), \tag{3.1}$$

where I_0 is a noiseless photon count, λ and σ are noise parameters. The values
of these parameters can be found by performing a calibration of the detector [82].
As a result, the forward projector can be computed significantly faster than a
Monte-Carlo simulation, and the data have a noise distribution roughly imitating
the real one.

3.3   Results

3.3.1   Deep learning model: training, validation and accuracy evaluation

As an example of deep learning model, we train DCNNs with EfficientNet B4
[134] architecture and cross-entropy loss to perform classification of X-ray images.
The corresponding task in Chapter 2 was X-ray image segmentation with an MSD
network combined with a post-processing of the segmented image. With a fully
DL-based classification, a separate post-processing method is unnecessary, and a
higher accuracy is expected.

The network is trained with different numbers of training images N_{train} to
evaluate how the size of the training set affects the network accuracy A. The
accuracy is determined by applying the model to the same set of N_{test} images
and comparing true classes y_i and predicted classes $\hat{y}_i$. The score A is computed

according to the formula

$$A = \frac{1}{N_{test}} \sum_{i=0}^{N_{test}-1} \mathbb{1}(\hat{y}_i = y_i), \tag{3.2}$$

where $\mathbb{1}(x)$ is the indicator function. This is an alternative accuracy metric to the F1-score introduced in Chapter 2, Eq. 2.11. The test set contains equal number of images for every class to ensure that the accuracy score is not biased towards certain classes.

In both case studies, there is a a collection of experimentally acquired CT scans split between training (S_{train} samples) and test (S_{test} samples). The test set always consists of real data, p projections per sample, resulting in $N_{test} = pS_{test}$. When the network is trained on real data, the training set is constructed similarly: $N_{train} = pS_{train}$. To evaluate the data generation performance, the network is trained only on artificial data, and the same number of projections is created with fewer volumes: $N_{train} > pS_{train}$. The goal of the data generation method is to achieve similar A with fewer S_{train}.

Training of the network stops based on the validation score. The training data are split in advance between training and validation subsets, approximately 15% of images is used for validation. The performance of the trained neural network might significantly depend on random factors, such as initialization and nondeterministic algorithms. As a result, the value of A changes if the same network is trained multiple times. To characterize the uncertainty ΔA, every model is trained four times with different random seeds and the same training data.

3.3.2 Data

The CT-based approach for creating artificial X-ray images is illustrated on two FOD problems. The first case study is based on the dataset [161, 162] that was acquired at the Computational Imaging group of Centrum Wiskunde & Informatica (CWI) in Amsterdam, the Netherlands. These data were collected as an example of X-Ray FOD and contain images of modeling clay with pebble stones. We use these data as a model problem for FOD with a piece of clay being a homogeneous main object and stone as a foreign object. Even though this problem does not have an inherent industrial value, it poses challenges similar to industrial inspection tasks, such as detection of bones in poultry or glass in peat. Firstly, a piece of clay was remolded for every scan, so the dataset contains a variety of main objects with unpredictable shape. Secondly, despite a significant contrast, a pebble stone can still be not detected on a single projection due to unknown shape of the sample and small size of the stone.

The training and test set contain $S_{train} = 30$ and $S_{test} = 30$ samples from the original collection with zero, one, and two stones. For a classification task, the number of foreign objects in the sample is chosen as a class (examples of all classes are shown on Fig. 3.3). We have additionally acquired CT scans of 10 samples without pebble stones since the original dataset only had 11 such scans.

The experimental settings of these additional scans are similar to the settings in the original collection. For every sample, $p = 72$ projections are used as X-ray images for FOD. Furthermore, the original data are downsampled with a factor of 2, for an effective pixel size of 300 μm and voxel size of 190 μm.

To acquire the second dataset, we performed X-ray scans of twenty five avocado fruits at the FleX-ray laboratory [36] of CWI in Amsterdam, the Netherlands. The measurements were performed with the voltage of 90 kV, power of 45 W, exposure time of 300 ms per projection, and magnification factor of 1.3. The original X-ray image size was $1912px \times 1520px$ with a pixel size of 75 μm, 1440 images were acquired for every sample. The fruits were stored without refrigeration for two weeks and scanned throughout this period. We observed that over time the amount of air inside avocado was increasing. Thus, a freshly harvested fruit can be distinguished from a ripe one based on the presence of air. Speed of this process significantly varies among different samples due to inherent biological variability. To formulate a FOD problem, we consider presence of air pockets inside avocado as a foreign object.

The experimental dataset contains $S_{train} = 36$ and $S_{test} = 20$ objects (most fruits have multiple scans corresponding to different days), $p = 72$ projections per object. For every sample, a class is assigned based on the fraction of air volume with respect to the whole fruit. We have chosen an arbitrary threshold of 1% air to separate two classes of objects (examples of both classes are shown on Fig. 3.3). For faster deep learning model training, images and reconstructions were downsampled with a factor of 4, leading to the effective pixel size of 300 μm and voxel size of 230 μm.

3.3.3 Sample generation

Data generation reduces the amount of manual data acquisition by constructing a representative training set consisting of artificial X-ray images. This approach could only work if the artificial and real data are similar enough, so the trained model can be applied to real images. To evaluate similarity, we train the same DCNN on real images and artificial images that are supposed to be identical to real. More precisely, we perform the segmentation and forward projection without any changes of the volume. First, this procedure checks if the segmentation is accurate enough to preserve contrast and morphology of object features. Second, we evaluate how much missing physical effects influence the training process of DCNN. Furthermore, this comparison verifies if the addition of noise could mitigate shortcomings of the simple forward projection operator.

In the presented case studies, formulating a deformation model is a challenging task since neither base nor foreign object have a fixed shape or blueprint. Thus, it is not trivial to check algorithmically if the deformed object is still an object of the same type that can be encountered by the inspection system (e.g. an avocado fruit after deformation is still avocado). The first deformation model we propose to use is an affine transformation that includes resizing, rotation, shear and translations. By applying it to the foreign object, it is possible to change
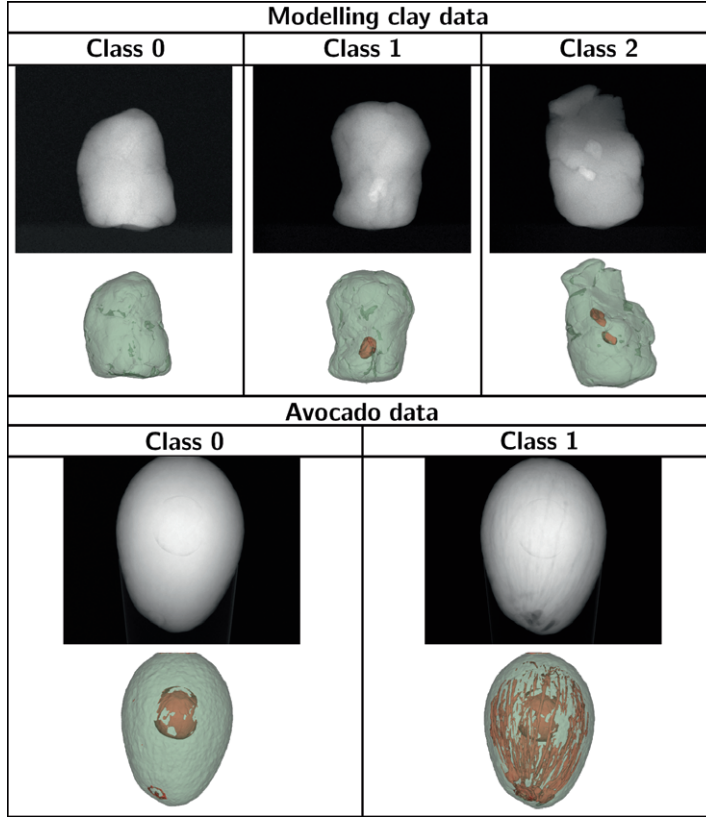
Figure 3.3: Examples of objects from different classes for two case studies. The first row corresponds to input X-ray images, and the second row shows visualization of the volume based on the segmentation (base object is colored green, foreign object - red).

its placement inside the base object and generate a variety of volumes using a single scan as a blueprint. Furthermore, resizing and shear could slightly change the shape of the object while preserving morphological features since the size of agricultural products often varies in a certain range. However, such reasoning might break if the transformation matrix is sufficiently different from identity matrix (e.g. a deformed object is orders of magnitude smaller than the original one), so a manual inspection of deformed objects is necessary to determine a range of parameters. It is possible that this range of parameters is too small to produce visibly different images rendering affine transformation unsuitable for some FOD problems. For example, in avocado case study, air pockets might form around specific regions of the fruit, and translations are not always applicable. To reflect this case, a region removal method is used as a non-linear deformation model. This algorithm splits any object into regions by randomly assigning a number of point

in the object to different clusters and then performing a nearest neighbor search for all other points. Then a certain number of these clusters is removed and the rest is kept. The region removal method is used to change the shape of the object considerably while preserving its position.

For both case studies, we present two possible generation strategies that require different amount of real data and expert knowledge to define deformation models. The first one (later referred as basic modification) takes multiple samples of the same class and generates images for other classes by applying deformation models to the foreign object. In the modeling clay case, the algorithm starts with a segmentation of the sample containing two stones (Class 2) and generates two artificial samples: one for Class 1 by removing one of the stones and one for Class 0 by removing both. Pieces of stone are removed by replacing stone material on segmentation with the modelling clay material. Thus, the number of training images N_{train} is $3pS_{train}$ since the samples of only one class are used by the generation method. For avocado FOD problem, the algorithm takes a volume of Class 1 avocado and applies region removal to reduce the amount of air. As a result, every real-world sample is used to create two avocado volumes of both classes, and $N_{train} = 2pS_{train}$. It is crucial to remove only a fraction of air. Otherwise, the network wouldn't see examples of Class 0 objects with air and misclassify images in the test set.

The second, more data-efficient strategy applies deformation models to a single real-world object to generate as many different artificial volumes as possible. The first step is to apply affine transformation (a mix of resizing and shear) to the real volume. Parameter range is chosen by a human expert to ensure that a deformed volume is not identical to the original one, but still looks like a real sample. Thus, scaling coefficients are in a neighborhood of 1, and shear coefficients are around 0. The exact limits used for our implementation can be found in the code repository (referenced at the beginning of the chapter). In the modelling clay case, the original volume is chosen from Class 1. Furhermore, in addition to the affine transformation of the whole object, the pebble stone is duplicated using a second affine transformation. This FO tranformation includes not only resize and shear, but also translation and rotation to create a Class 2 sample with two pebble stones as shown on Fig. 3.4a. This is an arbitrary choice that highlights that a real sample with two FOs is not required to generate artificial samples with two FOs. Then, as in basic modification, the stone objects are removed from this artificial volume to create Class 0 and 1 samples. For avocado case (Fig. 3.4b), an affine transformation is applied to the Class 1 sample, and Class 0 objects are generated by performing region removal. In both case studies, the number of generated images N_{train} is pGS_{train}, where G is the number of artificial volumes generated for every real scan, and $S_{train} = 1$.

3.3.4   Software implementation

ASTRA Toolbox [142] is used to perform reconstruction of the sample volumes based on the X-ray CT scans using FDK algorithm [51]. The segmentation uti-
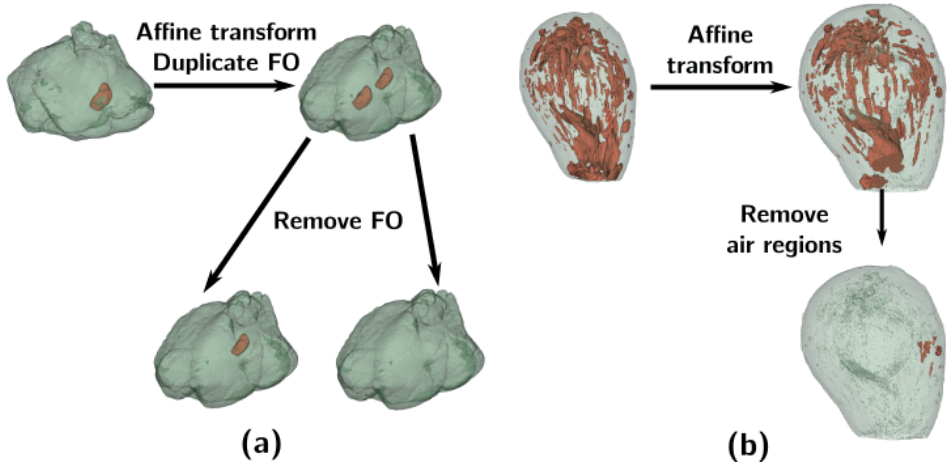
Figure 3.4: Artificial sample generation based on one real-world object: for modelling clay (a); for avocado (b).

lizes image processing algorithms (Otsu, convex hull, morphological operations) implemented in the scikit-image package [147]. Deformation models apply affine transformation implemented in the SciPy package [151], and the region removal algorithm was implemented by us. Forward-projection uses the operator from ASTRA Toolbox.

3.3.5 Accuracy of modeling clay classification

For modeling clay FOD, the DCNN performance for different amounts of real data used for training is drawn as a blue region on Fig. 3.5a, green dots correspond to the accuracy scores A of the individual instances of the same network. Training images based on $S_{train} > 20$ CT scans are enough to achieve $A = 90\%$. Quality of the artificial data was tested on a subset of $S_{train} = 21$ real-world samples. When the model is trained on generated images corresponding to the real volumes, the average A stays on the same level of $\approx 90\%$ as with real data. The values of A for noiseless and noisy artificial projections are shown in Table 3.1. Mixed noise is applied according to Eq. 3.1 with the values of noise parameters $\lambda = 1.17$ and $\sigma = 20$. Simulation of noise reduces ΔA, but does not improve the average performance.

The accuracy score of networks trained on generated data is shown as orange region (basic modification) and a red point (generation from one) on Fig. 3.5a. For basic modification, the values of A approximately correspond to accuracy score achieved with twice as many real samples. Thus, this generation strategy reduces the amount of data acquisition by half and achieves the same predictive strength.

For the generation from one sample ($S_{train} = 1$), the value A mainly depends on the number of artificial volumes G generated from this sample (Fig. 3.5b).

For $G < 20$ the resulting performance rises and then saturates. The best value of $A = 87\% \pm 3\%$ is achieved with $G = 45$ and corresponds to the performance that could otherwise be achieved with $S_{train} = 15$ real objects. The important difference is that ΔA does not decrease when the number of generated objects G rises as opposed to real data. Furthermore, with $G > 60$ the mean value of A is getting worse due to a higher ΔA.

Training data, $S_{train} = 21$	Accuracy score A
Real data	$91.1\% \pm 0.7\%$
Noiseless artificial	$91.8\% \pm 1.3\%$
Noisy artificial	$91.8\% \pm 0.4\%$

Table 3.1: Comparison of model accuracy for different types of training data: real images, artificial noiseless images, and artificial images with noise added as a post-processing filter.
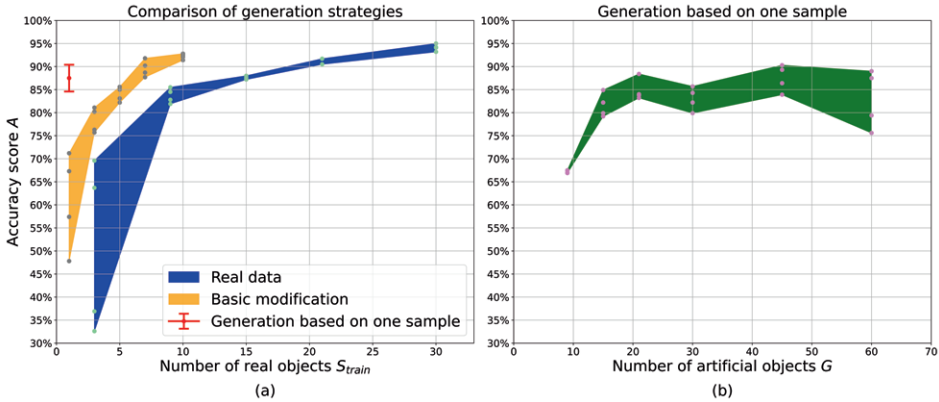
Figure 3.5: Dependency of modeling clay classification accuracy A on the amount of data provided for training. Comparison between training on real data and generated images with a different number of real samples S_{train} (a). Training on the artificial images generated based on one experimentally scanned sample (b).

3.3.6 Accuracy of avocado classification

According to the experiments with real data (Fig. 3.6a), scans of $S_{train} = 24$ fruits are enough to achieve binary classification accuracy $A = 95\%$. The standard deviation ΔA decreases to 2% and stays at that level when more data are used for training. The subset of $S_{train} = 24$ objects was also used to check the quality of the forward projection operator. As shown in Table 3.2, the network accuracy significantly drops if the model is trained on noiseless data. However, the addition of mixed Poisson-Gaussian noise improves the performance of the model. The

difference in accuracy between models trained on real and noisy artificial images lies in one standard deviation range.

The accuracy score of networks trained on generated data is shown as an orange region (basic modification) and a red point (generation from one) on Fig. 3.6a. Networks trained on images generated with the basic modification strategy reach the best real-data performance $A \approx 95\%$ with $S_{train} = 12$ instead of 24 real objects. Furthermore, this method outperforms training on real data achieving $A = 97\% \pm 2\%$ with all available Class 1 samples. Thus, a 50% reduction in the amount of data acquisition S_{train} is also possible in the avocado case without loss of performance.

The generation from one sample strategy leads to a model accuracy $A = 95\% \pm 3\$$ with $S_{train} = 1$. The value of A slightly depends on the number of generated volumes G (Fig. 3.6b), but the standard deviation ΔA stays at 3% even for a large number of $G > 50$. In the avocado case, the exact values of A significantly depend on the parameters used for air region removal. A variety of Class 0 samples can be generated from a Class 1 sample since air clusters to be removed can be chosen in different ways. For this reason, the experiments with data generation were repeated multiple times with different random seeds and distributions for the amount of air to be removed from the sample. In case of generation from one, the experiment was also performed with a different original object. With a small number of S_{train} for basic modification and small G, the values of A could drop significantly with respect to the reported ΔA. However, by generating $G = 100$ artificial volumes it was possible to reach the performance of $A > 92\%$ with different random distributions and original real-world volumes.

Training data, $S_{train} = 21$	Accuracy score A
Real data	$95.3\% \pm 1.7\%$
Noiseless artificial	$90.5\% \pm 4.4\%$
Noisy artificial	$94.5\% \pm 1.5\%$

Table 3.2: Comparison of model accuracy for different types of training data: real images, artificial noiseless images, and artificial images with noise added as a post-processing filter.

3.4 Discussion

The results show that for two case studies with real data, accurate classification models can be trained on artificial X-ray images based on a small amount of data acquisition. For modelling clay case, one experimentally scanned object provides enough information for the generation method to achieve accuracy corresponding to the training on 15 real samples. Despite using only one real-world avocado, the model trained on generated fruits reaches the performance that would otherwise require more than 20 objects. Such a significant reduction in the amount
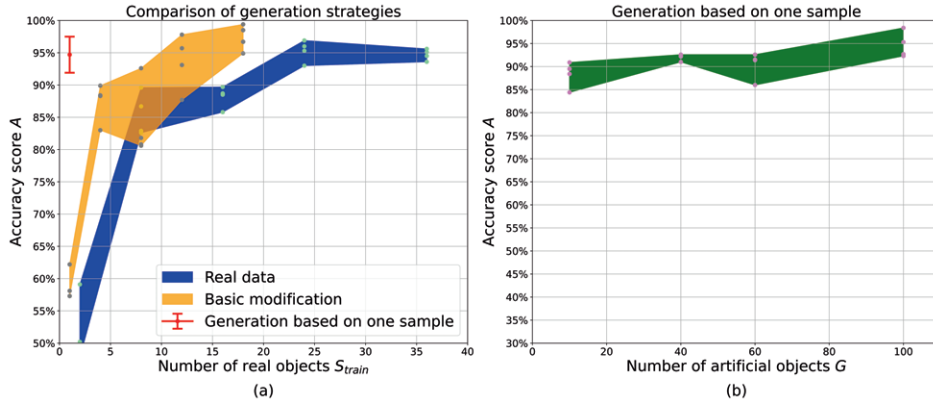
Figure 3.6: Dependency of avocado classification accuracy A on the amount of data provided for training. Comparison between training on real data and generated images with a different number of real samples S_{train} (a). Training on the artificial images generated based on one experimentally scanned sample (b).

of data acquisition requires a change in data acquisition approach and poses new application-specific challenges related to data generation. To effectively apply this methodology, the user has to guide the acquisition of CT data, formulate deformation models, and adjust the forward projection operator.

Despite FOD problem being formulated for a single X-ray projection, this methodology specifically requires CT scans of samples that provide enough data to reconstruct volumes. Although in the case studies a laboratory setup was used for both tasks, in the industrial case they should be performed with two separate systems. A detector on a factory floor is used to take low-resolution images with low exposure for a fast FOD, and a CT scanner acquires a large number of projections from different angles with high exposure and resolution. Even though a CT scan takes more time than a single projection, the data generation approach requires fewer real-world samples with a proper choice of deformation models.

Access to 3D models of base and foreign objects gives user a direct control over all possible combinations that can be encountered by a quality control system. Deformation models applied to the segmented volume can be used to put foreign objects in different locations of the base object, and transformations of the base object can change features visible on a single X-ray view. In the manual data acquisitions, this variability would require the user to provide many different samples representing different combinations of the base object and FO. For FOs of biological nature (air gaps, infestation, chemical changes in tissues), it is not possible to construct an arbitrary sample in real life, but it can be generated algorithmically by providing necessary models. The main limitation of the sample generation approach lies in the application-specific expert knowledge to make sure that a variety of possible samples is represented. Additionally, artificial volumes

might be impossible to exist in reality, due to biological or physical constraints, and such samples could mislead the model and decrease its predictive ability.

Artificial sample generation can be utilized to reuse the existing data and extend problem formulation. Possibility to deform sample volumes can be used to adapt the existing deep learning model to another size of FOs or a different shape of the base objects. New types of FOs can be scanned separately and injected in the existing models of the base objects. Data generation strategies with many real-world samples, such as basic modification, can be used to fix class imbalance in the manually acquired datasets.

The data generation approach relies on the assumption that artificial X-ray projections can be similar to real projections. Depending on the application, the user has to adjust a forward projection operator. In the presented case studies, the operator did not include many physical effects that can be simulated with Monte-Carlo methods. Nevertheless, the results show that a more accurate operator would not significantly improve the accuracy of the deep learning model. Even though we did not compare our approach with other simulation methods, we evaluated the performance based on real data, and a perfect operator would replicate real images. Good enough performance of the simple operator can be explained by a low resolution of the inspected images and structure of the deep learning models that might ignore or blur small details.

A significant challenge of this methodology is its sensitivity to the application-specific decisions made by the user. Some implementation details might significantly change the generated data, and this effect can only be evaluated by training a deep learning model. Such behavior was observed for artificial avocados where the amount of air in the generated volumes could significantly change the resulting accuracy of the model. However, a similar problem can be highlighted for conventional training on real data where the performance is implicitly controlled by the acquired images. Picking a representative set of real-world samples is also a challenge, especially if the different batches of products have different FO patterns. The data generation approach provides a more direct way to control coverage of the training data and a faster way to expand the existing dataset.

3.5 Conclusion

Efficient application of deep learning methods to industrial quality control requires large annotated datasets that are rarely available. We present a CT-based method for creating artificial X-ray images that can be used to train DCNNs instead of real data. This methodology requires a small number of CT scans, significantly reducing the amount of necessary data acquisition. Deformation models applied to the segmented sample volumes generate a variety of possible combinations of the base and foreign objects and create a representative dataset. The sample generation strategy can be adjusted based on the amount of available data and a-priori knowledge about the problem and makes the approach flexible to new features in the data and extensions of the problem formulation. The results show that even

a single CT scan can train a model capable of accurate foreign object detection in real-time. By changing the approach to data acquisition, this methodology simplifies the use of deep learning methods for industrial quality control tasks.

SCATTERING EFFECT

QUANTIFICATION

X-ray imaging is used as a non-destructive testing technique suitable for objects of different shapes made of different materials [30, 74]. We focus on the task of foreign object detection: a product may contain an undesirable structure (an object made of different material or a void) that should be detected by an inspection system. Depending on the task, detection may only require indicating that a foreign is present, or locating it and specifying its size and other properties. Examples of foreign objects in this context include spallation in metal alloys, blowholes in castings, razors in airplane baggage, bones in fish fillets, etc. Due to differences in density and chemical structure, defects affect an X-ray image (later referred to as a projection) of the product. Their presence can be detected by analyzing the projection with an algorithm or a human expert (Fig. 4.1). A projection can be acquired in tens of milliseconds, making X-ray imaging an in-line inspection technique. While individual projections can be inspected by a human expert, high throughput of data requires a fast and accurate foreign object detection algorithm. The main challenge in analyzing X-ray projections is the overlap between features of the object located at different depths. This complicated task is conventionally solved by human experts relying on application-specific knowledge.
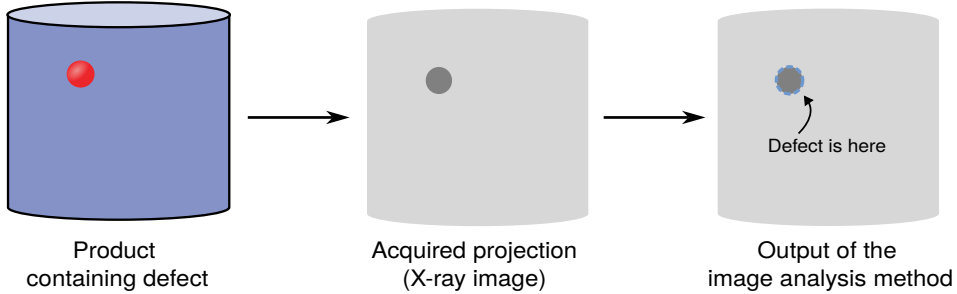

Figure 4.1: General scheme of foreign object detection via X-ray imaging. An X-ray acquisition system is used to make a projection of the product of interest. Defects affect projection intensity even if they are inside the product, and can be detected by analyzing the projection.

In recent years, deep learning methods, such as Deep Convolutional Neural Networks (DCNNs), have been successfully applied in many areas of industrial product inspection [87, 102, 157]. Deep learning algorithms assume that the image processing task can be solved by a function with a large number of parameters. The values of these parameters are found by optimizing the result of the function for a large set of projections (referred to as training data). Compared to conventional image processing algorithms, DCNNs provide higher accuracy at the expense of interpretability. The high predictive power of DCNNs stems from their ability to generalize image patterns present in the training dataset. Such datasets, with a few exceptions [104], are not available for X-ray foreign object detection. The previous Chapter demonstrated that computational methods can be used to generate large datasets of X-ray images to train accurate deep learning methods. However, the generation method proposed in Chapter 3 was not a perfect representation of the X-ray imaging process and contained many simplifications. Therefore, a more in-depth study is needed to determine whether this approach can be applied with similar success under different circumstances.

Experts from various fields have tried to simulate X-ray data using computational methods [15, 61]. Depending on the desired degree of similarity to real data, the methods range from ray-tracing to Monte-Carlo simulation of each individual X-ray photon. The computational cost of the simulation algorithm for X-ray imaging comes from the models of X-ray interaction with objects and the detector. Most methods do not estimate the effect of X-ray photon scattering, since it significantly increases the computational complexity of the problem [60]. Instead, only primary radiation - photons that pass through the object without scattering - is simulated. The effect of X-ray scattering is also absent from the generated data in Chapter 3. If the training data are generated using only the primary X-ray signal, the DCNN will miss the influence of scattered photons that are present in the real test data. Hence, a decrease in accuracy is possible. Due to the low interpretability of DCNNs, it is impossible to predict in advance how

a change in the training data would affect performance on a wide range of test cases. Balancing the computational cost of the data generation approach and the resulting DCNN performance requires an accurate and robust metric of detection accuracy.

The Probability of Detection (POD) method [55] was introduced in non-destructive testing and evaluation to analyze the reliability of inspection techniques. Recently it has been applied to X-ray radiography and CT [139, 158]. POD analysis originally addressed the problem of determining the smallest defect size that can be found with an inspection system under specified measurement conditions. The introduction of *Probability* stems from the existence of many possible products with the same foreign object size, and only in a fraction of them the inspection is successful (e.g. due to the location of the defect). The POD curves aim to estimate these fractions by statistical analysis of the inspection results on the test dataset. The POD method can be applied to any inspection technique including deep learning methods. While the POD curves do not analyze the influence of the training set on the DCNN, they connect the performance of the networks to the application requirements.

We present a methodology that quantitatively evaluates the effect of X-ray scattering in the training data on the accuracy of DCNNs for defect detection in X-ray projections. To demonstrate its application, we perform computational experiments on a model problem consisting of the detection of a cavity inside a cylindrical object. Using a Monte-Carlo simulation algorithm, we generate two versions of the same training data that include and exclude scattered X-ray photons. These data are used to train two DCCNs. We use POD curves to evaluate the correlation between network accuracy and test projection properties. This allows us to separate the influence of training and test data properties and to highlight conditions under which simulation of scattering is crucial. We discuss how this methodology can be applied to other industrial problems and which task properties significantly influence the results.

4.1 Related work

4.1.1 Deep learning in real-time foreign object detection

The problem of foreign object detection can take different forms depending on the desired output of the algorithm. *Classification* returns a single label corresponding to the type of defect present in the projection (e.g., steel defects [97]). *Object detection* returns bounding boxes for each foreign object and labels corresponding to their types (commonly used in baggage inspection [11] and quality control of metal details, such as aluminum castings [116]). *Segmentation* provides a set of pixel masks: each pixel of the projection is labeled if a foreign object is present there (e.g. spallation in aircraft engines [18]). There is a difference between *semantic* (one mask for all foreign objects of the same type) and *instance* (different foreign objects of the same type are separated) segmentation. This is not relevant

for our model problem, where there is only a single foreign object.

In contrast to color photographs, there is a lack of publicly available X-ray projections of industrial products for training and benchmarking DCNNs. A notable exception is GDX-ray [104] - a dataset containing projections of welds, castings, baggage, and natural objects. Data for a specific foreign object detection problem have to either be acquired manually or generated computationally. There are two main approaches to image generation: transforming real-world projections with image-to-image methods and simulating X-ray imaging. Image-to-image algorithms can be used to add defects to existing images of objects without defects [145]. Generative Adversarial Networks (GANs) can also be used to create new data similar to the real data [136] or to perform style transfer from one imaging modality to another [9].

4.1.2   Simulation of X-ray imaging and the effect of scattering

Simulation of X-ray imaging requires knowledge of the experimental setup and the 3D structure of the studied object. There are two categories of simulation algorithms: probabilistic and deterministic. The *probabilistic* approach uses Monte-Carlo (MC) methods to imitate the stochastic nature of real X-ray interactions with matter, and can produce highly accurate results. The particle physics toolkit GEANT4 [1] is used as a gold standard to verify other algorithms. The GEANT4-based software GATE is often used to simulate different modalities of medical imaging (PET, SPECT, CT) and dosimetry [76]. The main drawback of these methods is the computational cost. The generated projections contain stochastic noise that can only be reduced by simulating a large number of X-ray photons. The *deterministic* approach uses ray-tracing algorithms to compute projections faster by using a simplified model of X-ray interactions. aRTist [15] (Analytical RT Inspection Simulation Tool) was developed as a fast simulation software to generate realistic X-ray projections based on the mesh model of the object. A similar approach was proposed for baggage inspection [61] using GPU-based ray-tracing. It was later shown [60] that X-ray scattering can also be simulated with additional computational cost.

Simulation of X-ray photon scattering is missing from many probabilistic algorithms. When it is implemented, it increases the computational cost by orders of magnitude [60]. A faster alternative is to approximate the distribution of scattered photons; in particular using convolutions [17, 133]. Convolution kernels can be extracted from Monte-Carlo simulations or from experimental measurements. However, these methods are limited by the difficulty of finding a small number of kernel parameters that work for various objects. Deep learning scattering estimation [94] has been proposed as an alternative to kernel-based methods, but it requires a larger amount of data with known scattering patterns.

If the scattering effect is not reduced experimentally or compensated algorithmically, the projection quality may be compromised [123]. This problem is thoroughly studied in radiology, where scattering can reduce the contrast in projections, making them unsuitable for diagnostic purposes [14]. The amount of

scattered radiation is usually measured by calculating the scattering-to-primary ratio (SPR). The ratio depends on the field of view, the air gap between the patient and the detector, and X-ray energy [23]. SPR can be reduced by hardware techniques, such as anti-scatter grids. There are several metrics that quantify the change in SPR: Contrast Improvement Factor (CIF), Bucky Factor (increase in absorbed dose), and change in signal difference to noise ratio SdNR [39].

Many important insights can be found in radiology articles on the connection between scattering, image quality, and quality for diagnostic purposes. The effect of scattering on some image properties, such as the point spread function, can be calculated [21]. Using clinical trials, it is possible to infer the effect of image properties such as noise power spectrum, resolution, and point spread function on image quality for diagnostic purposes [77, 95]. To our knowledge, there have been no studies correlating the presence of scattering with a decrease in diagnostic accuracy that would allow a quantitative evaluation of the impact of anti-scatter techniques.

4.2 Methods

We formulate the foreign object detection task as a segmentation problem. The goal is to compute a labeled image $y \in \mathcal{Y}$, where each pixel has a label according to its classification, from the measured X-ray projection $x \in \mathcal{X} \subset \mathbb{R}^m$. The set of possible labels $\{0, ..., k\}$ corresponds to the defect types and depends on the particular detection task. An algorithm performing segmentation is described as

$$f_\phi : \mathcal{X} \to \mathcal{Y}, \tag{4.1}$$

where ϕ are the parameters of the algorithm. Supervised deep learning considers hypothesis functions $f_\phi \in \mathcal{H}$ with a high-dimensional parameter space and determines ϕ by solving the optimization problem

$$\phi = \arg\min_\phi \sum_{i=1}^{N_{train}} l(f_\phi(x_i), y_i), (x_i, y_i) \in S_{train}, \tag{4.2}$$

where $S_{train} = \{(x_i, y_i)\}_{i=1}^{N_{train}}$ is a training dataset consisting of projections and correct predictions (ground-truth) and l is a loss function. By including and excluding the scattering signal, we create two projections x_i^{MC} and x_i^R that correspond to the same object and have the same ground-truth y_i. Consequently, these two types of projections form two versions of the training dataset S_{train}^{MC} and S_{train}^R. Since these sets are not equal, they lead to the different solutions ϕ^{MC} and ϕ^R of the optimization problem.

Due to the low interpretability of deep learning methods, the accuracy of f_ϕ is usually characterized by applying the function to the test set $S_{test} = \{(x_i, y_i)\}_{i=1}^{N_{test}}$. With an accuracy metric $h : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$, the accuracy of the function with parameters ϕ is characterized by the sequence

$$A^\phi = \{h(f_\phi(x_i), y_i)\}_{i=1}^{N_{test}}.$$ (4.3)

We propose to analyze A^ϕ using POD curves, which can be determined by computing the regression between the accuracy A_i^ϕ for each object and the vector of corresponding properties of the object and projection $\mathbf{t}_i$ (vector of independent variables). A simple example of the object property correlated with the detection accuracy is the foreign object size (later referred to as s_i), and the corresponding vector $\mathbf{t}_i = (\,1\ s_i\,)$ We assume that with a sufficiently good choice of properties, $\mathbf{t}_i$ is correlated with the probability of detection - the probability P that the value of A_i^ϕ exceeds the threshold A_{thr}. The regression is computed using a generalized linear model according to the equation

$$g(P(A_i^\phi > A_{thr})) = \mathbf{t}_i\boldsymbol{\beta},$$ (4.4)

where $g : \mathbb{R} \to \mathbb{R}$ is a link function (defining the shape of the POD curve) and $\boldsymbol{\beta}$ is a vector of regression coefficients. Then $\boldsymbol{\beta}$ can be used to characterize the performance of the algorithm with parameters ϕ in a data-driven way. For a desired accuracy A_{thr} achievable with probability P_{good}, a set of constraints for $\mathbf{t}$ can be computed with a confidence interval given by the regression uncertainty. By comparing the thresholds of $\mathbf{t}$ we determine whether the difference between ϕ^{MC} and ϕ^R is significant with respect to the object properties. Consequently, this answers the question whether S_{train}^{MC} can be replaced by S_{train}^R without losing detection accuracy.

We test this methodology on generated data following the approach formulated in our previous work [5] (Fig. 4.2). The data are generated based on a collection of 3D volumes that define the physical properties of the objects independent of the imaging method. Two algorithms (including and excluding scattering) are used to transform the volumes into projections x_i^{MC} and x_i^R by simulating X-ray imaging. These projections are combined into datasets S_{train}^{MC} and S_{train}^R to train DCNNs $f_{\phi^{MC}}$ and f_{ϕ^R}. Both networks are tested on the same dataset S_{test}^{MC}. The POD analysis is performed to find a correlation between A^ϕ and foreign object size. Thus, we can compare the smallest defect size detectable with ϕ^{MC} and ϕ^R and conclude if the difference is statistically significant.

4.2.1   Object modeling

Studying the influence of scattering imposes several constraints on the image generation approach. In our methodology, we want to avoid cases where a DCNN fails to detect a foreign object because it was not exposed to a certain morphology during training. This problem is well-known in deep learning as Out-of-Distribution performance [70, 71]. For example, if a network was trained to detect defects in large objects, it might fail on a test projection with a small object, regardless of scattering. The objects used for the training and test sets should have the same distribution over $\mathcal{X}$, and as many objects as possible should be in both sets. In real-world applications, test objects are measured experimentally, while training

Figure 4.2: Application-driven approach to evaluate the difference between simulation with and without scattering. First, a large number of 3D volumes is generated by combining different variations of object and defect geometry. Two forward projection methods are used to transform the 3D volumes into X-ray projections. Each dataset containing a variety of projections is used to train a DCNN. They are applied to the same collection of test data with scattering. The performance of the DCNNs is evaluated using POD curves.

objects are created digitally. Thus, it is difficult to guarantee that the distributions of training and test objects match.

We address these limitations by choosing to work with a model problem. Each 3D volume is created by combining a small number of objects according to a parameterizable algorithm. Training and test sets of volumes are created with different random sequences of parameters, each parameter having the same distribution for both sets. We assume that a large enough number of volumes is generated, so that the training and test datasets have similar distributions over $\mathcal{X}$. We check this assumption qualitatively by visual inspection of the detection performance, but a detailed evaluation is beyond the scope of this work. In addition to the geometry, the volume is characterized by the material properties. To consider a variety of scattering patterns, we repeat the experiments with the same geometry and different material compositions.

The model problem we use in this study is the detection of cavities in cylinders. The object is a homogeneous cylinder parameterized by radius and height. The foreign object is an ellipsoidal void inside the cylinder. It is defined by ellipsoidal axes, height, and distance from the rotational axis of the cylinder. Materials of the cylinder are chosen from the common materials of industrial products: PMMA (type of plastic, synthetic polymer $(C_5O_2H_8)_n$), aluminum, and iron.

4.2.2   Image generation

To ensure that the scattering distribution in the generated data is accurate and similar to real measurements, we use Monte-Carlo simulation as implemented in the GATE framework [76]. An accurate Monte-Carlo simulation requires a detailed knowledge of the experimental setup: X-ray source, studied object, and detector. However, not all of these are crucial for an accurate scattering simulation. The most important properties are the energy spectrum of the emitted X-ray photons, the attenuation curves of the materials present in the object, and an accurate description of the object geometry (defined as a voxelized volume, a mesh, or a collection of simple shapes). GATE simulates each emitted photon individually and records its coordinates if it is registered by the detector.

If the simulation considers only X-ray absorption, the number of photons detected in a pixel p is given by

$$I_p^{abs}(E) = I_0(E) \exp\left(- \int_{l_p} \mu(E, x) dx \right), \tag{4.5}$$

where $\mu(E, x)$ is a 3D distribution of the object's attenuation coefficient for the X-ray energy E, l_p is the trajectory from the X-ray source to the pixel p. The acquired X-ray projection x^R is computed by summing the photons of all energies and correcting for the flatfield (the projection acquired without the object)

$$x_p^R = -\log \frac{\int_0^\infty I_p^{abs}(E) dE}{\int_0^\infty I_0(E) dE}. \tag{4.6}$$

The same equation was used in Chapter 2 (Eq. 2.1) to define image intensity. This formulation does not include the effect of the detector gain [155]. When X-ray scattering is simulated, the number of photons detected I_p is $I_p^{abs} + I_p^{scat}$. The number of scattered photons I_p^{scat} is not defined by the trajectory l_p. Instead, each voxel of the object v contributes to the amount of scattering according to the equation

$$dI_{pv}^{scat}(E') = \frac{1}{\sigma_v(E)} \frac{d\sigma(\theta_{pv}, E)}{d\Omega} I_v(E) d\Omega_{pv}, \tag{4.7}$$

where $\frac{d\sigma}{d\Omega}$ is the differential scattering cross-section, $\sigma_v(E)$ is the total cross-section for all scattering from the voxel v, θ_{pv} is the scattering angle, $I_v(E)$ is the energy spectrum of the photons in the voxel, and $d\Omega_{pv}$ is the solid angle corresponding to p seen from v [54]. Differential cross-sections of Rayleigh and Compton scattering as a function of scattering angle and energy are known. Simulation of scattering makes the signal in a pixel dependent not only on a single trajectory but on the entire object. Furthermore, multiple scattering is possible, and $I_v(E)$ includes photons scattered from other parts of the object. With a sufficiently large number of simulated photons, Monte-Carlo methods consider many possible scattering trajectories and compute the total number of scattered photons I^{scat}. Similar to Eq. 4.6, raw measurements should be flatfield corrected to compute the projection with scattering

$$x_p^{MC} = -\log \frac{\int_0^\infty \left[I_p^{abs}(E) + I_p^{scat}(E)\right] dE}{\int_0^\infty I_0(E) dE}. \tag{4.8}$$

The model problem simplifies the definition of object geometry and material composition. The object geometry is defined as a mixture of a cylinder and an ellipsoid which are the basic object shapes for GATE. While any object can be simulated based on its polygonal mesh, the computational cost of photon tracking increases with object complexity. Limiting object geometry to basic shapes accelerates the simulation. The attenuation properties are automatically calculated for a given chemical formula. We compute the X-ray source spectrum with an empirical model of an X-ray tube (implemented in xpecgen [72]), which provides an energy spectrum depending on the voltage. Several voltage values are chosen for each material, since the scattering distribution depends strongly on both properties.

An image produced by a Monte-Carlo simulation is inherently noisy due to the limited number of generated photons. It can be shown that the number of registered photons follows Poisson distribution [155] (Compound Poisson for energy-integrating detectors). Noise in real X-ray projections is additionally influenced by other factors, such as noise in the detector electronics [156], focal spot size, and scintillator blur [130]. We choose to perform a simulation with a perfect detector that registers every photon reaching it (technically implemented as a sufficiently thick layer of a heavy material). Simulation of a realistic detector is possible with Monte-Carlo methods but is beyond the scope of this study. Such modeling would

require detailed knowledge of a real detector, which is often lacking in industrial tasks.

The output of GATE includes the coordinates where each X-ray photon was detected and how many times it scattered before detection. After matching the coordinates with different pixels of the detector, a total distribution of primary photons of all energies I^{abs} (Fig. 4.3a) and a distribution of scattered photons I^{scat} (Fig. 4.3b) integrated over energy and all possible scattering centers can be calculated. The distribution of SPR is defined as $\frac{I^{scat}}{I^{abs}}$ (Fig. 4.3c). Following Eq. 4.6 and 4.8, distributions of detected photons are converted into pre-processed projections with and without the scattering signal (Figs. 4.3e and 4.3d).

(a) Primary   (b) Scattering   (c) Scattering-to-primary ratio

(d) Projection without scattering   (e) Projection with scattering

Figure 4.3: Different distributions of the X-ray signal that are computed with a Monte-Carlo algorithm: distribution of primary photons registered by the detector (a), distribution of scattered photons (b), scattering-to-primary ratio (c), projection image without scattering after pre-processing (d), projection image with scattering (e).

4.2.3   DCNN Training

We train segmentation DCNNs to perform foreign object detection. This approach not only detects the presence of the foreign object but also provides its location on the projection. The training process requires input data (previously described x^R and x^{MC}) and the ground-truth with defect locations. To construct the ground-truth, the Radon transform is applied to 3D volumes. The result of this Radon transform is a 2D distribution of material length along the ray from the source to

the detector. This image is used to outline the location of the foreign object and compute its geometric properties. The DCNN trained on projections that include scattered photons is referred to as the network trained on MC data. The one trained on data without the scattering signal is referred to as the network trained on Radon data.

DCNN training as an optimization problem for network weights ϕ is affected by random factors (initialization, GPU computation). The performance of a single network is not sufficient to draw conclusions about the influence of training data. Therefore, each training is repeated multiple times to have multiple instances ϕ of the same DCNN. The performance of the instances is then averaged to determine the performance of the DCNN.

The DCNN architecture determines the structure of the weights ϕ and has a significant effect on the performance. Before selecting an architecture for a majority of the experiments, we performed one analysis with multiple state-of-the-art segmentation architectures (the results are given in the Appendix). It was observed that different architectures with commonly used metaparameters converge to a similar level of accuracy. Thus, the influence of scattering could be studied for one architecture, and similar results are expected for others.

We choose the Mixed-Scale Dense Convolutional Neural Network (MSD) [117] for a majority of experiments, as in Chapter 2. This architecture is an alternative to classical convolutional neural networks with scaling, and is aimed towards easy training and reusability. The MSD network consists of multiple layers with the same set of operations (convolution, summation and activation). All layers produce feature maps of the same dimensions and are connected to each other (dense connections). Dilated convolutions are used to capture image features at different scales (mixed scale). Due to dense connections and a small number of parameters (compared to other popular architectures such as UNet), an MSD network can be trained quickly and provide accurate segmentations. All networks share the same architecture parameters: a depth of 30, a width of 1, dilations in the range from 1 to 10, 1 input channel, and 2 output channels. The total number of weights is close to 5000. We have tested the performance with different values of width and dilations and have not observed a noticeable improvement compared to the standard parameters. The optimization of weights is done by ADAM optimizer with a learning rate of 10^{-3}. Network training and testing is done in PyTorch framework with MSD implementation from [69].

4.2.4 Performance evaluation

Many metrics h can be applied to evaluate the similarity between the output of the segmentation DCNN and the ground-truth. We use the F_1 score (as in Eq. 2.11), which can be calculated according to the equation

$$F_1 = \frac{2TP}{2TP + FP + FN},$$

(4.9)

where TP (True Positive) is the number of defect pixels that were correctly segmented, FP (False Positive) is the number of pixels that were incorrectly marked as corresponding to the defect, and FN (False Negative) is the number of defect pixels that were missed by a segmentation algorithm.

Fig. 4.4a illustrates that F_1 varies significantly between different test cases with different defect sizes. By size, we refer to the largest intersection between the foreign object and the primary X-ray trajectories (the largest thickness). This is not the only way to characterize the foreign object. For example, its area or perimeter can also be used to represent size. These properties are often correlated because expected defects have similar shapes. Following the POD methodology, we convert the segmentation accuracy into a binary variable by the defining segmentation as successful if $F_1 > 50\%$. In Fig. 4.4b, the projections are divided into groups by defect size binning and a fraction of successful segmentations is computed for each bin. Segmentation accuracy improves as the foreign object size increases.

The properties of the POD curve are determined by computing the regression according to Eq. 4.4, where $\mathbf{t}$ is a constant and the foreign object size. The link function g defines the shape of the POD curve and reflects the dependence between the probability of successful segmentation and the defect size. The probability is close to zero for small foreign objects, one for large foreign objects, and there is a transition in between. The generalized linear model in this case is represented by the equation

$$g(P(F_1 > 50\%)) = \ln \frac{P(F_1 > 50\%)}{1 - P(F_1 > 50\%)} = a + bs, \qquad (4.10)$$

where a and b are the fit parameters. Their values are computed by solving an optimization problem of maximizing likelihood of getting the observed detectability outcomes for the test set with known values of defect size. The logit function $g(x) = \ln \frac{x}{1-x}$ can be replaced by other functions such as log-log, complementary log-log, probit, etc. An example of a POD curve is shown in Fig. 4.4c (computed using the statsmodels package [127]).

The parameters a and b describe the performance of the algorithm f_ϕ. To conclude whether two algorithms have the same performance in this context, the values of a and b should be compared, taking into account the uncertainty interval. In POD analysis, it is common to approach this problem in a task-driven way and compute the value of s leading to a fixed level of good detection probability (usually 90%) according to the equation

$$s_{90} = \frac{g(90\%) - a}{b}. \qquad (4.11)$$

This value s_{90} describes the size of the smallest detectable foreign object. Due to regression uncertainty, the true value of s_{90} lies in a region around the computed value. Thus, if two sets of parameters ϕ^R and ϕ^{MC} have different values of s_{90}, there is a probability that the true value is the same in both cases, and the difference is only caused by the fit uncertainty. We address this problem by
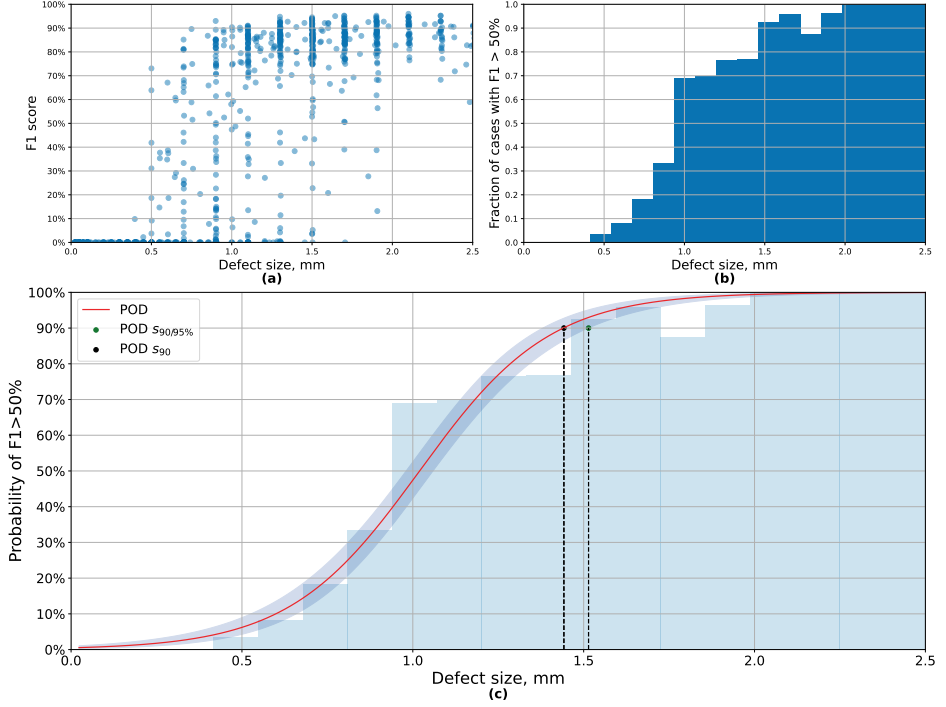
Figure 4.4: Example of a POD curve and its relation to accuracy for a single projection: the correlation between F_1 score and defect size for all projections in the test set (a), a histogram computed after binning the defect size indicating that the fraction of projections with $F_1 > 50\%$ increases with defect size (b),the POD curve representing a probability of $F_1 > 50\%$ with the smallest detectable foreign object size s_{90} and its higher bound highlighted (c).

computing an upper bound for s_{90} which is referred to as $s_{90/95}$. The true value of s_{90} is less than $s_{90/95}$ with 95% probability. Therefore, if the computed value of s_{90} for ϕ^R is less than $s_{90/95}$ for ϕ^{MC}, the difference in performance is statistically significant. Otherwise, it can be explained by the fit uncertainty with less than 95% probability.

The accuracy of segmentation is expected to depend on many projection properties. In particular, the defect location and surrounding object features are also used in POD analysis [29]. When the POD curve is computed, the performance for a particular value of defect size is averaged over different objects, different defect locations and their shapes. More projection properties could be included in a multivariate POD curve to produce a more detailed description of detection performance. We use a multivariate fit to study the influence of SPR on detection. The corresponding generalized linear model is given by the equation

$$g(P(F_1 > 50\%)) = a + bs + cSPR. \tag{4.12}$$

4.2.5   Implementation details

The geometric properties of the objects for the training and test sets are generated in advance and kept the same for all materials and voltages. We generate 1250 objects for the training set (1000 for training, 250 for validation) and 1000 objects for the test set. The radius of the cylinder ranges from 1 to 25 mm and the height varies from 20 to 55 mm. Ellipsoidal cavities are generated as deformed spheres: first, a radius is chosen to be in the range from 0.1 mm to 1 mm, and then the ratio between each axis and this radius is in the range from 0.7 to 1.3. A cavity can be placed at any distance from the axis of rotation. The acquisition geometry remains the same for all datasets. The source-object distance (SOD) is 200 mm and the source-detector distance (SDD) is 300 mm, resulting in a magnification factor of 1.5. The detector plane is a 75×82.5 mm^2 rectangle with a pixel size of 0.3 mm. Thus, the generated projections have a size of 250×275 px^2. One projection contains MC simulations of 10^9 emitted photons. Examples of these projections are shown in Fig. 4.5.
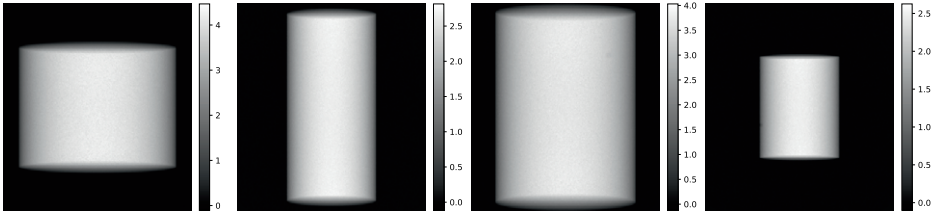

Figure 4.5: Examples of projections corresponding to different generated volumes.

Different material compositions and voltages are used to explore a variety of scattering distributions. In practice, heavier materials are inspected at higher voltages. Otherwise, only a small fraction of the X-ray radiation penetrates the object, resulting in a high noise level. The combination of material and voltage can be characterized by the value of HVL (Half-Value Layer) - the thickness of the object at which the intensity of X-ray entering it is reduced by half. PMMA is simulated at 90 kV (HVL = 27.8 mm) and 150 kV (HVL = 31.7 mm). Aluminum is simulated at 90 kV (HVL = 4.28 mm), 150 kV (HVL = 6.25 mm) and 300 kV (HVL = 17.9 mm). Iron is simulated at 300 kV (HVL = 3.85 mm) and 450 kV (HVL = 4.6 mm). Fig. 4.6 shows different scattering distributions for the same object depending on material and voltage. The number of scattered photons (not necessarily SPR) increases with higher voltage and lower atomic number. Furthermore, there is a qualitative change in the spatial properties: for materials with low atomic number scattering is more uniform, and for heavier materials it follows

the shape of the object. This can be explained by different ratios of Rayleigh and Compton scattering and probabilities of further absorption in the object for scattered photons.
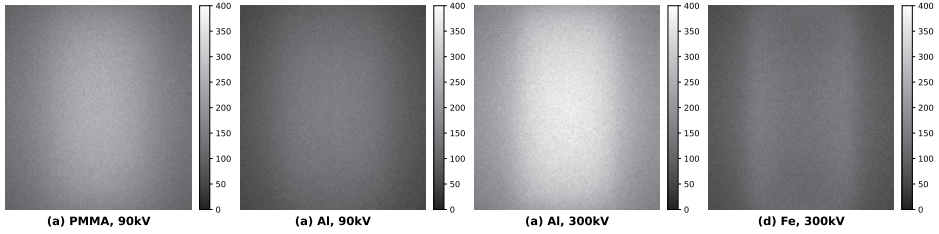

Figure 4.6: Comparison of scattering distributions for the same object made of different materials and inspected with different tube voltages: PMMA at 90kV (a), aluminum at 90kV (b), aluminum at 300kV (c), iron at 300kV (d). For the same voltage, the number of scattered photons is higher for materials with lower atomic numbers. For the same material, the amount of scattering is higher for higher voltage. Iron objects have a different scattering pattern with more photons near the edges.

4.3 Results

4.3.1 Comparison of DCNN performance

For each combination of material and voltage, two DCNNs with the same architecture are trained on data without scattering (Radon data) and with scattering (MC data), 10 instances each. Averaging the segmentation accuracy of the DCNN on the test dataset does not lead to a good performance estimate due to the high variance. For example, in the case of iron at 450 kV, the network trained on Radon data has a segmentation accuracy of $43\% \pm 40\%$ on the test dataset, while the network trained on MC data achieves $44\% \pm 40\%$. A high variance is observed for each dataset, making a direct comparison of the DCNNs impossible.

For a detailed analysis of the performance, POD curves are computed for both DCNNs for each voltage and material combination. Fig. 4.7a shows a pair of POD curves for PMMA at 90 kV with no difference in performance between the networks trained on Radon and MC data. The POD curve for the Radon data network lies within the 95% confidence interval of the MC POD curve, making the difference between them statistically insignificant. The opposite case is shown for iron at 450 kV (Fig. 4.7c) where higher POD values for the network trained on Radon data lie below the confidence interval for the network trained on MC data (in particular, s_{90}). Hence, there is a statistically significant difference between the training sets, and MC data lead to better performance. Nevertheless, the difference between s_{90} is only 5% for iron at 450 kV and smaller for other datasets (e.g. for

aluminum at 150 kV as shown in Fig. 4.7b). Furthermore, for defects larger than s_{90}, the average segmentation accuracy is 82%±16% for the network trained on MC data and 82%±18% for the network trained on Radon data. Although the variance is smaller than for the whole dataset, the performance of the two networks is still effectively the same.
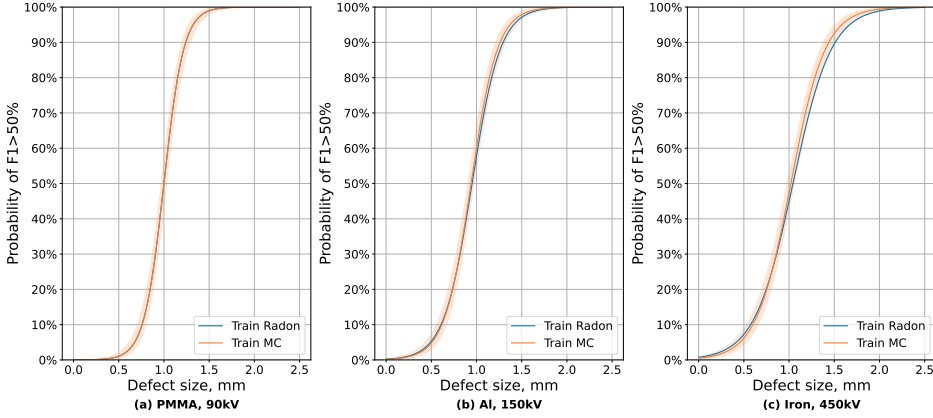

Figure 4.7: Comparison of POD curves for different datasets: PMMA at 90 kV (a), aluminum at 150 kV (b), iron at 450 kV (c). The difference between training on projections with and without scattering is negligible for PMMA. It becomes larger for aluminum but within a 95% confidence interval. For iron at 450 kV, the network trained on MC data performs better than the network trained on Radon data considering the confidence interval.

The values of s_{90} for both networks in all cases are shown in Table 4.1. In most cases (except for iron at 450 kV) the difference in the smallest segmentable foreign objects between the networks is less than the confidence interval. This serves as an upper estimate for the possible performance difference. Depending on the material and voltage, the relative difference in s_{90} is between 0% and 3% and is unlikely to exceed 5% even if the fit coefficients are not precisely determined.

4.3.2   Influence of scattering-to-primary ratio

The negligible difference in the performance of the networks trained on Radon and MC data raises the question of biases in the test data and projection properties that influence segmentation. For each dataset, we perform a multivariate POD fit according to Eq. 4.12 to evaluate the influence of SPR at the defect location (averaged over a region containing the defect). After determining the POD coefficients, the value of SPR can be fixed at a certain level to obtain POD for that amount of scattering signal (Fig. 4.8 for PMMA, aluminum, and iron).

Several observations can be made after splitting the dataset based on the SPR value. First, as the SPR increases, the difference between the networks trained on

Material, kV	MC, s_{90}	MC, $s_{90/95}$	Radon, s_{90}	s_{90} difference
PMMA, 90 kV	1.24 mm	1.29 mm	1.24 mm	0%
PMMA, 150 kV	1.31 mm	1.36 mm	1.31 mm	0.2%
Al, 90 kV	1.39 mm	1.46 mm	1.42 mm	1.9%
Al, 150 kV	1.26 mm	1.32 mm	1.29 mm	2.4%
Al, 300 kV	1.17 mm	1.22 mm	1.18 mm	0.8%
Iron, 300 kV	1.59 mm	1.68 mm	1.62 mm	1.7%
Iron, 450 kV	1.44 mm	1.51 mm	1.51 mm	5.0%

Table 4.1: Smallest detectable defect size for different materials and training datasets.

Radon and MC data also increases. This can be determined either by performing a multivariate fit or by making a univariate fit for a subset of the test data (with a sufficiently high SPR). Second, the values of s_{90} increase as the SPR rises. Third, different datasets have different distributions of SPR. For example, aluminum at 150 kV has 500 objects with SPR < 0.1 while iron at 450 kV has 400 such objects. Hence, even if there is no difference between the networks trained on Radon and MC data for low SPR in both cases, the overall POD curve is more affected for aluminum because such cases are more common.

It is important to note that the POD curves only show a correlation between the DCNN accuracy and the different properties of the projection. Furthermore, the value of SPR at the foreign object location is correlated with other properties such as attenuation rate as shown in Fig. 4.9. The attenuation rate μ in a pixel is calculated as $-\log \frac{I}{I_0}$ where I_0 is the total number of photons emitted in the direction of the pixel and I is the number of photons registered (including scattered radiation). The noise in the MC simulation follows a Poisson distribution and depends on I. Regions with high SPR often have high μ, leading to high relative noise levels. This correlation may explain the increase in s_{90} for a higher SPR.

In Table 4.2 we compare the s_{90} values in each dataset at the minimum and maximum SPR levels (the minimum is close to 0 in each case). With the sole exception of PMMA at 90 kV (where it can be explained by fit uncertainty), higher SPR leads to a greater difference in performance between the networks trained on Radon and MC data. The geometric structure of the test dataset leads to a large number (approximately 40-50%) of cases with low SPR, mostly due to the small size of the object. In these cases, there is little difference between training on data with and without scattering. Consequently, the performance on the whole test dataset becomes more similar, even if in other cases the MC network has a smaller s_{90}.
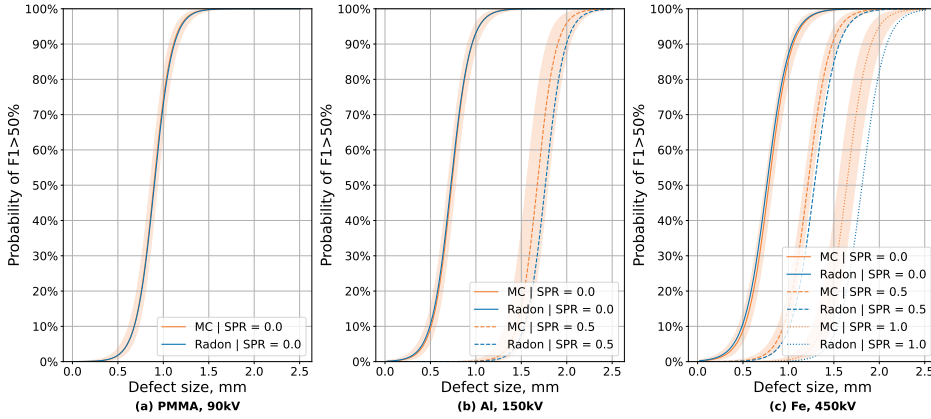
Figure 4.8: Comparison of multivariate POD curves computed at different values of SPR for different datasets: PMMA at 90 kV (a), aluminum at 150 kV (b), and iron at 450 kV (c). Increasing the SPR results in a larger difference between the POD curves of the networks trained on Radon and MC data. Pairs of POD curves are plotted for SPR values of 0., 0.5, and 1.0 if such projections are present in the dataset (the largest SPR for PMMA is 0.16, for aluminum it is 0.55).

4.4   Discussion

We presented a methodology that evaluates the difference between training data for DCNNs using a model problem as an example. This methodology can be applied to various foreign object detection tasks, as individual parts of the methodology can be replaced to better fit a particular problem. The main requirement is the ability to completely remove the X-ray scattering signal for a set of data. It is then possible to compare the dataset containing the scattering signal with the dataset without scattering. For a particular detection task, one should specify an algorithm to perform the task and assemble the test dataset for performance evaluation. POD analysis is one of many possible ways to evaluate the accuracy of the algorithm. As a more general form of linear regression, POO curves provide a robust technique that can be applied to a wide range of problems without extensive expert knowledge. Depending on the task, a more accurate model relating the DCNN performance to object properties may be suggested.

In the presented model problem, both training and test data are generated using a Monte-Carlo algorithm. This approach provides an easy way to obtain two versions of the same data that differ only in the presence of scattering. Alternatively, the same result can be achieved using the real data and either software or hardware scattering reduction techniques to obtain projections without the scattering signal. The main disadvantage of using experimental data is that the scattering would only be approximately removed, whereas Monte-Carlo simulation can ensure that no scattering is present in the data. The analysis made on sim-
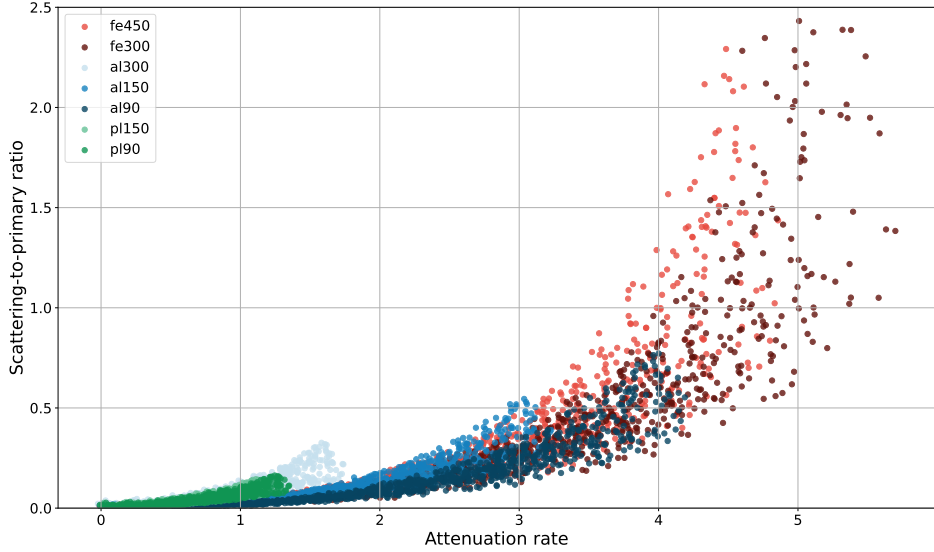
Figure 4.9: Correlation between projection properties at the defect location (SPR and attenuation rate) for different datasets (markers are colored with red for iron, blue for aluminum, and green for PMMA). SPR increases with attenuation rate, and the dependency may be similar for datasets with different materials and voltages.

ulated data is not necessarily applicable to real measurements, but Monte-Carlo algorithms have been validated to be sufficiently accurate for medical purposes [41, 124]. While MC methods can be applied to any problem, objects with complicated morphology and high variance present a computational challenge to the proposed methodology. The computational cost of a single photon simulation depends on the geometric representation and might increase by orders of magnitude for a detailed polygonal mesh. Thus, optimizing the level of detail in a model is crucial for real-world applications. Furthermore, high variability leads to a large number of objects that have to be simulated to produce a representative training set.

Accurate MC simulation requires extensive knowledge of the modeled X-ray system to account for all possible experimental effects. Our model problem does not represent a specific setup, and a number of simplifications has been made. We have not included detector noise in the MC data, although it could be implemented as an image post-processing. According to the calibration of our X-ray system at 90 kV [36], 10^9 simulated photons correspond approximately to a high exposure of 1 s at 45 W of tube power. In this case, the Poisson noise component simulated by MC methods is the main source of noise, and the electronic noise of our detector could be ignored. Furthermore, we have not observed significant changes in the analysis due to noise correlations present in real projections. We

Material, kV	Max SPR	At SPR $= 0$		At Max SPR	
		MC, s_{90}	Radon, s_{90}	MC, s_{90}	Radon, s_{90}
PMMA, 90 kV	0.16	1.11 mm	1.12 mm	1.59 mm	1.57 mm
PMMA, 150 kV	0.14	1.22 mm	1.19 mm	1.57 mm	1.63 mm
Al, 90 kV	0.77	1.00 mm	0.97 mm	2.38 mm	2.51 mm
Al, 150 kV	0.55	0.96 mm	0.96 mm	1.99 mm	2.09 mm
Al, 300 kV	0.33	1.07 mm	1.07 mm	1.47 mm	1.50 mm
Fe, 300 kV	5.26	1.02 mm	1.03 mm	5.84 mm	6.06 mm
Fe, 450 kV	2.29	1.05 mm	1.04 mm	3.02 mm	3.44 mm

Table 4.2: Difference in s_{90} for minimal and maximal SPR values

expect that detector noise could be neglected in similar high exposure cases and become significant in low dose studies. Our MC simulation includes only an X-ray source, object, and detector, and does not contain additional objects that may be in the field of view in industrial applications, such as object holders, conveyor belts, and radiation shielding. These objects could contribute to the scattering signal, especially backscatter, which is not present in our simulation.

We performed foreign object detection using the MSD architecture for a segmentation DCNN. In the appendix, we show that similar results can be achieved with other DCNN architectures, including DeepLabv3, UNet++, and FPN. Changing metaparameters, such as the number of layers and dilation sequences, may slightly improve the performance for a particular training dataset. Choosing the architecture and fine-tuning the metaparameters are necessary to achieve the best possible performance. Since our goal was to study the effect of scattering, we stopped changing the metaparameters after reaching an accuracy level that could not be easily improved. We have tried to train a classification network instead of segmentation, but the results were inferior. Such a difference in performance can be explained by the absence of the defect mask in the training data, which complicates the learning process. We have chosen the MSD architecture due to the relatively small number of weights (compared to other popular architectures such as UNet), the possibility to train an accurate model without a large amount of training data, and a history of successful application to X-ray tasks. Comparison with other DCNN architectures shows that a small number of weights does not significantly reduce accuracy. It is important to note that vision transformers [43] currently outperform DCNNs in standard computer vision tasks (e.g. ImageNet classification), although the novel CNN architectures might compete with transformers [90]. We used DCNNs due to their successful application to industrial problems and the large number of available implementations. Furthrermore, it is not clear whether vision transformers could outperform DCNNs without large datasets.

When using the difference between POD curves as a metric of performance difference, it is important to note that many POD curves with different feature vectors could be computed for the same network. In preliminary experiments

we included object thickness at the defect location and defect distance to the center as parameters in the POD fit. Defect size was chosen as the main POD argument due to a major influence on segmentation accuracy and interpretability. Other variables could be added to Eq. 4.12, but the conclusion about the impact of SPR remains the same. Another important assumption is that the chosen training dataset covers all possible combinations of object and defect. Thus, an incorrect segmentation is caused by the input properties and not by encountering a previously unknown foreign object type or unusual location. Providing sufficient coverage is easier for manufactured objects with known shapes (e.g. castings), and more complicated when both the object and the defect have a large variety in morphology. The threshold of 50% F_1 score to convert network accuracy into a binary variable is chosen arbitrarily. If increased, it will shift the POD curves to the right with respect to the defect size.

While the presented results are only valid for a selected problem, they illustrate the benefits of using a POD-based methodology. A significant difference in accuracy was only present for a small subset of test projections, and this effect was not visible for other metrics due to variance. Furthermore, POD curves address the concept of the smallest detectable foreign object, which helps to evaluate whether a system is useful for a particular real-world application. We expect a similar effect of scattering on inspection performance for other problems. If a defect is too small, the contrast in X-ray attenuation is less than the noise level, and the inclusion of scattering does not significantly affect the detection process. If a defect is too large, the contrast would be larger than a variance in the signal due to scattering. Consequently, the biggest influence of the scattering signal should be seen for barely detectable defects when the contrast is similar to the noise level.

It is important to note that any foreign object detection problem has a number of system properties that affect the noise level and SPR and, consequently, the effect of excluding the scattering signal from the training data. The distance between the object and the detector (known in radiology as the air gap) has a significant effect on SPR values because scattered photons may not reach the detector. Furthermore, a smaller field of view and small objects can reduce the effect of the scattering signal. A longer exposure time decreases the noise level and s_{90}, so the effect of scattering as additional noise might become more significant.

4.5 Conclusion

The practical application of image generation techniques for defect detection in industrial products requires a balance between the computational cost of data generation and the resulting accuracy of the algorithm. We have proposed a methodology to quantitatively evaluate the effect of simulating X-ray scattering on data generation for training DCNNs. The POD curves have been used to study the network performance in detail by correlating it with the properties of the test projections. Using a Monte-Carlo simulation algorithm, we have ensured that the difference between the DCNNs was caused only by the presence of the scattering

signal. Our performance analysis was successfully applied to the various test cases. The POD curves have shown under which test conditions the difference between DCNNs is significant and when it is negligible. In particular, we have shown how the scattering-to-primary ratio affects the network accuracy and the influence of the image generation approach. The presented methodology can be used to decide if the defect detection performance is sufficient for a particular task and what level of accuracy in image generation is necessary to achieve that performance.

Appendix: Different DCNN architectures

For the dataset of iron at 450 kV we have trained DCNNs with different state-of-the-art architectures using the Segmentation Models package [73]. We have tried UNet++, FPN (Feature Pyramid Network), and DeepLabv3+ as semantic segmentation architectures, and EfficientNet, MobileNet, and ResNet (not included in the results because the performance was significantly worse) as encoders. The accuracy of each network on the test dataset was evaluated with POD curves in the same way as for MSD (Fig. 4.10a). The networks have similar performance, the difference in s_{90} between the best and the worst model is around 10%. The difference between training on Radon and MC data can be evaluated for other network architectures, and the effect is similar to MSD. For example, Fig. 4.10b is made with the same data as Fig. 4.7c, but the network architecture is DeepLabv3+ instead of MSD.

The similarity in network performance indicates that the difficulty of segmentation in the model problem stems from the image and noise properties and not from deep learning. This supports the assumption that the generated projections have enough variety to train a DCNN.
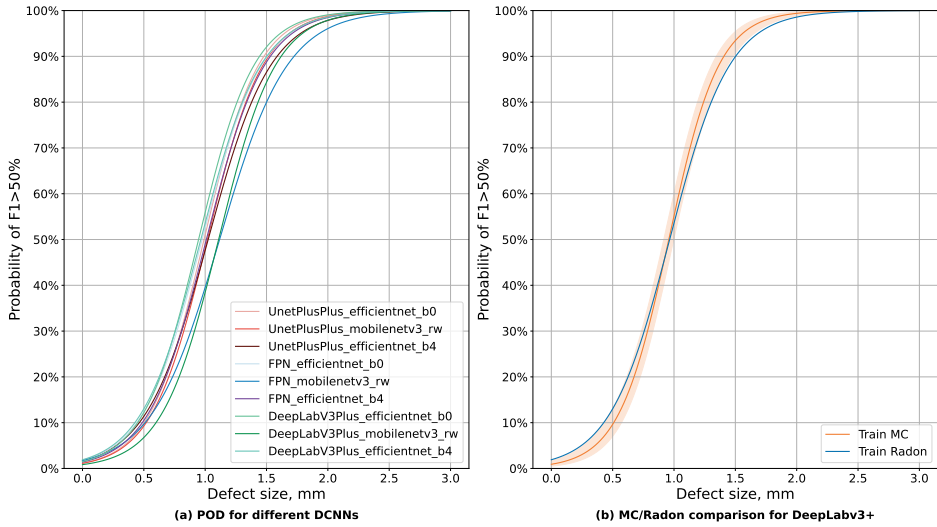
Figure 4.10: Comparison of POD curves for the DCNNs with different architectures tested on the iron dataset at 450 kV. DeepLabv3+ outperforms other architectures tested, MSD achieves one of the highest accuracy levels despite having the least number of parameters. The difference in s_{90} is around 10% when comparing the best and the worst model (a). Comparison of the networks trained on MC and Radon data when the network uses the DeepLabv3+ architecture with the EfficientNetB0 encoder. The decrease in s_{90} is similar to MSD (b).

SYSTEM DESIGN OPTIMIZATION

X-ray imaging systems are widely used for a high-throughput in-line inspection of industrial products on a conveyor belt [99, 100]. Unlike visible light cameras, X-ray radiation is able to penetrate the product and provide information about its internal structure. This is particularly useful in a foreign object detection (FOD) problem, where the goal is to inspect a *main* object and determine whether it contains an undesirable region or a separate body - a *foreign* object. Examples of foreign objects (FOs) are browning in apples, infestation in grain, and bones in poultry [44]. While X-ray inspection could achieve high accuracy in laboratory studies [112, 115], an industrial environment introduces many limitations, such as high-throughput, acceptable dose level, and physical space constraints. Designing a practically viable X-ray system - configuring the experimental equipment and the image analysis algorithm - is a challenging problem of balancing inspection accuracy and application constraints. Furthermore, this problem is highly application-specific; methods developed for one product are usually not effective for other products [99].

A standard approach to analyzing the inspection accuracy is extensive experimental testing. To achieve automated inspection without frequent manual checks, the inspection system has to be reliable. In the context of FOD, reliable systems should detect FOs with a low failure rate for a wide variety of products. The shape of the foreign object is not known in advance, and it could be located in different parts of the main object. The direction from which the product would

---

be viewed is also not known in advance. All these variations need to be tested to confirm that the system is reliable. However, this time-consuming test would only be valid for a specific configuration of experimental equipment and image analysis. Any change in the hardware settings would have an unpredictable effect on the acquired images and the performance of the system. Designing an inspection system from scratch involves a large number of independent settings. Performing an experimental test for all possible permutations of settings is infeasible. A different approach is therefore required.

There is a rising interest in creating virtual representations of physical products and systems - digital twins [78]. Digital twins of X-ray systems [2, 13, 19] are seen as a tool for predicting the X-ray image under the desired experimental settings without an experiment. Such a model could solve the system design problem. If the data could be generated significantly faster than the experimental acquisition time, then it would be practically possible to test a large number of different system settings. The problem of generating X-ray images with computational methods has been studied in detail with vastly different approaches [15, 16, 61]. While it is possible to implement a highly accurate virtual model of an X-ray setup [76], the computational cost is significant. As a result, sufficiently fast data generation would be too expensive and infeasible. This computational challenge leads to the question of how accurate the X-ray image generator needs to be to be used in practice.

In some cases, approximately accurate generated images could even be used to substitute the real ones [145]. This possibility has been demonstrated in this thesis, in Chapters 3 and 4. It is possible that perfect correspondence is not necessary, as long as the features used by the image analysis algorithm are generated accurately. For example, it has been shown in Chapter 4 that X-ray scattering can be removed from the data without a significant impact on the detection accuracy. However, to conclude that the performance of the inspection system is the same (or similar) for two different datasets, a quantitative way of characterizing the performance is required. This problem is well-known in nondestructive testing, and Probability of Detection (POD) curves [55] are often used as a solution. POD curves provide a quantitative estimate of how the detection rate depends on the value of the product properties, and for which values reliable detection is expected. Product properties associated with reliable detection can be measured and compared to conclude if the performance on two datasets is the same. In Chapter 4, the size of the defect was used as one such property, and these relevant properties have to be determined in each case.

We propose a novel data-driven method for efficient performance estimation of X-ray imaging systems under different experimental settings. Our contributions are as follows: We combine the theoretical foundations of X-ray image formation and X-ray data generation techniques to create images corresponding to specific experimental settings of the given inspection system. We show which properties of the system should be measured during calibration to make such generation possible. As an example of the FOD problem, we focus on the detection of small rib bones in chicken fillets with different X-ray exposure times. For the same

experimental settings, a real dataset is acquired with the experimental setup; and a generated dataset is created with a calibrated generator and the dataset of high-quality real images. We show that the use of generated data yields quantitatively comparable performance estimates to the real data, and explore the performance for settings that were not tested. We discuss the advantages of the proposed approach over the standard experimental testing and how our method could be applied to other industrial problems.

5.1 Related work

The problem of X-ray data generation is well-studied and can be solved with a variety of methods that are commonly categorized into probabilistic and deterministic. Probabilistic (Monte-Carlo) methods [16, 76] require a significant amount of computational resources and extensive knowledge of the X-ray system and the studied object. Deterministic (ray-tracing) methods [15, 61] are faster by orders of magnitude but do not include many experimental effects present in a real X-ray acquisition. Furthermore, deterministic algorithms produce a noiseless image, and the noise pattern has to be generated separately. Properties of noise in X-ray images were studied in detail to estimate measurement uncertainty [121] and improve analysis of low-dose X-ray data [93]. Equations connecting mean value and variance in X-ray data are derived in literature and experimentally verified [93].

Image analysis for the FOD is a sophisticated problem that can be addressed by many methods based on contrast [131], edge detection [166], and machine learning trained on manually extracted features. A large variety of main and foreign objects has to be incorporated into the algorithm to achieve the necessary robustness. Deep learning methods have been successfully applied to many computer vision tasks: first, Convolutional Neural Networks (CNN)[138], and more recently, Vision Transformers (ViT) [43]. Deep learning algorithms require a large amount of training data to determine the image features that are crucial to solving the task in question. With sufficient data coverage, these methods provide state-of-the-art results and achieve high accuracy, execution speed, and robustness.

A significant disadvantage of deep learning methods is a lack of interpretability. Some techniques, such as class activation maps, might correlate the algorithm's decision with some properties of the input image. In contrast, in conventional algorithms, such as edge localization with the Canny filter, the accuracy can be computed analytically based on the noise model [79]. However, this is impossible for real problems with complex image morphology. In practice, a deep learning method has to be analyzed as a black box, and the performance analysis may benefit from statistical analysis used to assess the accuracy of human experts (e.g. Receiver Operating Characteristic (ROC) curves and Probability of Detection (POD) curves [55]). Such methods imply that a sufficient variety of test cases can be provided to accurately estimate the performance of the algorithm.

Optimization of X-ray imaging has been studied in medical imaging to determine the geometry of experimental setup [128] and dose of radiation [58]. The

existing work proposes different theoretical methods that explore the connection between the properties of the experimental setup and image quality. Image quality is usually defined by performing phantom measurements and computing different image features, such as the signal-to-noise ratio. However, the correlation between such metrics and the performance on real problems is rarely studied. A possible solution to this problem was presented in [65] where a binary classifier was trained to predict if an object would be correctly identified by a human expert based on a set of image quality metrics.

5.2   General concepts

Our approach to analyzing industrial FOD problems is illustrated on Fig. 5.1. An applicable system has to satisfy industrial requirements. In the case of X-ray, we highlight a list of frequently relevant constraints:

- Geometric - the inspection system should fit into the existing conveyor belt setup.

- Deposited dose - inspected objects should not absorb too much X-ray radiation.

- Radiation safety - properties of the X-ray source, such as voltage and power, are limited by the permits and radiational shielding to prevent interference with other parts of the factory.

- Resolution - the system is supposed to detect features of a certain size.

- Throughput - the inspection system should operate at a speed comparable to the conveyor belt.

These requirements directly affect the settings of the X-ray system: possible magnification, X-ray source power, exposure time etc. Furthermore, they determine the type of X-ray inspection technique that can be applied. For high-throughput, the acquisition may be limited to a single projection, while in other cases in-line CT may be possible. The same requirements can be satisfied in many ways leading to many possible systems. From the industrial point of view, the FOD problem is formulated by defining the physical properties of the objects on the conveyor belt. For X-ray, the important ones are material composition (what is the object made of) and morphology (what is the shape of the object).

System settings and object properties influence the resulting X-ray image acquired during the inspection procedure. We argue that this image could be computed with a properly configured image generator without performing an experiment. The decision about the FO presence is made by the image analysis algorithm that takes the X-ray image as input. Image analysis is only indirectly influenced by the system settings and object properties. Instead, the accuracy of the analysis
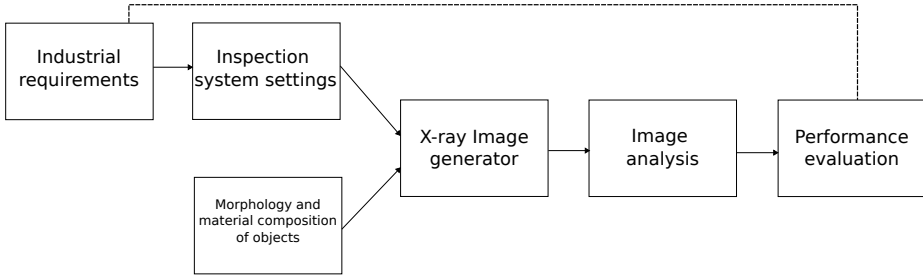
Figure 5.1: Overview of our approach for the FOD. The problem is defined by the physical properties of the objects to be analyzed and the inspection system settings. This leads to a variety of input images that have to be analyzed by the same image analysis algorithm. The performance of the algorithm is then evaluated as a function of the system design and the properties of the objects

is mainly determined by image features. Finally, there is a performance evaluation step that involves testing the inspection system on a large variety of products with and without FOs. The performance evaluation concludes whether the proposed system design is applicable to the problem taking into account the industrial requirements and the desired accuracy level.

In the context of Fig. 5.1, we consider the image generator to be sufficiently accurate when it yields the same (or quantitatively comparable) performance estimate as the real inspection system with the same settings and objects. To make such a generator, we start with a model describing an X-ray image formation. This model contains system-specific parameters that we propose to extract from the calibration of the real system. The goal of the calibration is to connect an abstract model of X-ray interactions to the particular X-ray imaging device under different system settings. When a sufficiently accurate generator is available, the industrial requirements should be translated into different sets of feasible system settings. For each set of settings, a performance estimate is computed using the image generator. If the requirements constrained the system too much, all estimates would not be sufficiently accurate, meaning that the problem has no practical solution under the current limitations. Otherwise, we could use the settings that lead to the best performance and confirm that they work in the real system.

5.3 Methods

To illustrate the proposed approach of predicting inspection performance using image generation, we concentrate on a specific case study: detecting bone fragments in chicken fillets under different conveyor throughput levels. The application-specific version of Fig. 5.1 is shown on Fig. 5.2. We consider a single industrial requirement - throughput of the system - that influences the exposure time possi-

ble for an X-ray image acquisition. A large number of high-quality X-ray images is used to represent a variety of fillets and bones that could be found on a conveyor belt. We consider a dual-energy X-ray setup as the inspection system. Dual-energy systems acquire two projections of the same object under different properties of X-ray radiation. Dual-energy images are often used in industrial [120] and medical [53] applications because they improve contrast and allow more accurate material identification. An image generator based on Beer's law with the mixed Poisson-Gaussian noise model is proposed to create images for different values of exposure time. The image analysis is performed by a segmentation deep convolutional neural network (DCNN), and the detection performance is analyzed with POD curves.
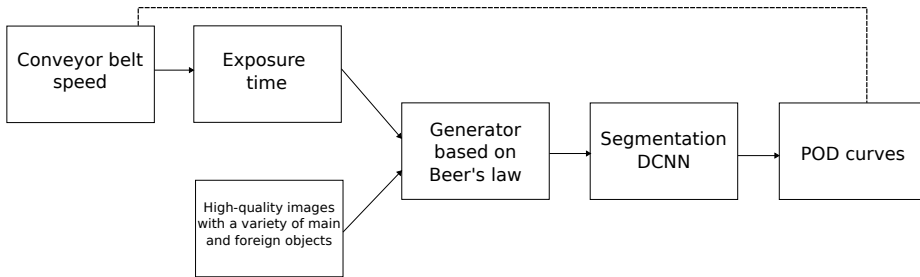

Figure 5.2: System design problem for the detection of bone fragments in chicken fillets under different levels of throughput. This figure illustrates the technical choices made to solve this problem: high-quality X-ray images as a representation of object morphology, an image generator based on Beer's law, DCNN as the image analysis method, and a POD curve as a performance metric

5.3.1   Industrial requirements and inspected objects

For in-line product inspection, the X-ray imaging system is built around the high-throughput conveyor belt. The imaging system consists of an X-ray tube which acts as a source of X-ray radiation and a detector which acquires an image by interacting with X-ray passing through the object. The conveyor belt speed is one of the most important application properties. The processing line is expected to run with a certain number of products per unit of time, and the inspection system must operate under this constraint. While this limitation affects image analysis, this chapter will focus on the influence of the limited measurement time on the acquired data. On a conveyor belt moving at the speed v, an object moves a distance vt, where t is the exposure time. Thus, the acquired projection is a mixture of many images of the object at different positions during the measurement. In practice, long acquisitions of fast moving objects lead to a motion blur effect. The magnitude of blur depends on the imaging resolution. If the distance vt is equal to multiple projection pixels of size a, then the same region of the object

will affect several pixels, producing a visible blur artifact. While some motion blur is inevitable when inspecting a moving object, we define an acceptable exposure time as

$$t = \frac{a}{v}, \tag{5.1}$$

so $vt = a$. This equation is also used for line detectors which acquire an image of a moving object line by line and are often used for high-throughput inspection [154]. Thus, to achieve higher throughput, the exposure time should be reduced. However, many other technical factors should be considered. For example, the exposure time is also limited by the detector electronics and cannot be reduced below a certain threshold.

Similarly to the FOD in other agricultural products, poultry inspection does not have a rigorous definition of main and foreign objects. Chicken fillets are similar in shape to a human expert, but the exact shape is unknown prior to the inspection. The chemical composition of chicken meat is known only approximately. Bone fragments could be found in a variety of locations inside the fillet depending on the cutting process. We choose a data-driven way to describe the variety of main and foreign objects, and their relative positions. A reference dataset is used as a problem definition. This dataset contains high-quality X-ray images of different chicken fillets with different bone fragments in various positions. The definition of high quality will be given later, and it refers to the low noise level to capture the X-ray attenuation properties of fillets and bones. The goal of the image generator is to transform a reference high-quality image into the image corresponding to a particular value of the exposure time.

5.3.2 Image formation

To estimate the influence of the imaging system and object properties, we use a model of X-ray imaging based on Beer's law. It assumes that the X-ray intensity I decays after passing through the object according to Beer's law

$$I = I_0 \exp\left(-\int_l \mu(x)\,\mathrm{d}x\right), \tag{5.2}$$

where I_0 is the intensity of the incident radiation, l is the X-ray trajectory, and $\mu(x)$ is the distribution of the attenuation coefficient. The goal of imaging is to estimate the properties of the object based on $\mu(x)$, which is connected to the measurable value of I.

For further analysis, we will make a number of assumptions about the studied object. $\mu(x)$ is a combination of attenuation in the main and the foreign object. We assume that both are homogeneous. A homogeneous object consists of a single material with a constant attenuation rate μ and occupies a volume V. Thus, if the studied object has no FO and consists of only one material,

$$\mu(x) = \mu \mathbb{1}_V(x), \text{ where } \mathbb{1}_V(x) = \begin{cases} 1 & \text{if } x \in V \\ 0 & \text{otherwise} \end{cases}. \tag{5.3}$$

While V is the physical shape of the object, which remains the same regardless of the imaging method, μ depends on both the object and image acquisition. X-ray tubes produce X-ray radiation with a wide range of energies characterized by the energy spectrum $I_0\Phi(E)$. For any material with a known chemical formula, an attenuation curve $\mu(E)$ can be computed to estimate the attenuation probability for every X-ray energy. It can be shown [3] that μ in Beer's law is approximately defined by the equation

$$\mu = \int_0^\infty g(E)D(E)\Phi(E)\mu(E)\,\mathrm{d}E \tag{5.4}$$

that also includes the detector properties: gain $g(E)$ and sensitivity $D(E)$. The same equation without an explicit mention of the gain and the sensitivity was used in Chapter 2 (Eq. 2.3). While it was demonstrated there that μ is also influenced by the shape of the object, this effect is not taken into account in the further equations.

The intensity I in Eq. 5.2 represents the true value, but it is not necessarily equal to the value measured by the imaging system. The generation of X-ray photons I_0 follows Poisson distribution. It can be shown [155, 156] that the number of registered photons I also follows Poisson distribution under certain conditions. Additionally, there is an electronic noise independent of the number of detected X-ray photons. Consequently, the measured signal y follows a mixed Poisson-Gaussian distribution

$$y = p + b \sim \lambda\mathcal{P}(\frac{I}{\lambda}) + \mathcal{N}(d_\mathrm{e}, \sigma_\mathrm{e}), \tag{5.5}$$

where λ denotes the detector gain and converts the number of X-ray photons following the Poisson distribution to the measured intensity units, d_e is often referred to as the darkfield signal (image intensity without X-ray irradiation), and σ_e is a standard deviation of the darkfield signal. Furthermore, we add a convolution with a Gaussian kernel

$$y_{blur} = y \circledast h, \qquad h(x,y) = ae^{-\frac{x^2+y^2}{2\sigma^2}} \tag{5.6}$$

to represent all effects influencing the Point Spread Function (e.g. focal spot of the tube and scintillator processes) [130]. This noise model represents a modified version of the model from Chapter 3 (Eq. 3.1), where Gaussian blur and darkfield were not included. Note that the parameter λ in Eq. 5.5 is the inverse of λ in Eq. 3.1.

The image formation model defined by Eq. 5.2 and 5.5 splits the contribution of object properties (V and $\mu(E)$) and imaging system design ($\Phi(E)$, I_0, and detector parameters $g(E)$, $D(E)$, λ, d_e, σ_e). While the former are constant and define the FOD problem, the latter can be optimized to provide the best image quality. All detector parameters in this model are inherent properties of the chosen equipment and can not be changed without replacing a detector. This type of optimization is beyond the scope of this chapter because it is mostly defined by the price of the

detector and the available budget. However, the incident beam intensity I_0 and the energy spectrum $\Phi(E)$ can be adjusted without changing the equipment.

The incident beam intensity I_0 depends on the geometry of the imaging system and the tube settings. X-ray tubes produce radiation uniformly distributed in a cone of height d - the distance between the source and the pixel. The flux j depends slightly on the voltage and is proportional to the tube current i. The X-ray radiation is collected in a square pixel of size a over the exposure time t. Thus, the incident beam intensity I_0 is given by

$$I_0 \approx j \frac{a^2}{4\pi d^2} t \propto it \frac{a^2}{d^2},$$ (5.7)

if the pixel is close to the center, and cone beam artifacts can be ignored. This equation offers many ways to control the incoming intensity: changing the imaging resolution (a), the exposure time (t), the geometric configuration of the system (d), and the current i.

5.3.3 Dual-energy image segmentation

If the inspected object consists of the main object with properties $(\mu_\mathrm{m}, V_\mathrm{m})$ and the foreign object with properties $(\mu_\mathrm{f}, V_\mathrm{f})$, Eq. 5.2 transforms into

$$I = I_0 \exp\left(-\mu_\mathrm{m} \int_l \mathbb{1}_{V_\mathrm{m}}(x)\,\mathrm{d}x - \mu_\mathrm{f} \int_l \mathbb{1}_{V_\mathrm{f}}(x)\,\mathrm{d}x\right).$$ (5.8)

The image is usually analyzed after post-processing according to the equation

$$M = -\log\frac{I}{I_0} = \mu_\mathrm{m}L_\mathrm{m} + \mu_\mathrm{f}L_\mathrm{f},$$ (5.9)

where M is the image after pre-processing, and $L = \int_l \mathbb{1}_{V_\mathrm{m}}(x)\,\mathrm{d}x$ is the thickness of the object V along the ray l. This equation highlights the problem of material separation for a single X-ray image. The same value of M can be obtained with different thicknesses of the main and foreign objects. However, as demonstrated in Chapter 2, this ambiguity can be reduced by performing a dual-energy acquisition.

In practice, dual-energy imaging can be implemented in two different ways [120]. In the first method, the voltage of the tube is rapidly switched to capture an X-ray projection with different X-ray energy properties. The second approach which is more commonly used in industrial environments uses a detector with two sensor media. The first sensor registers low-energy X-ray radiation, and the second sensor detect the high-energy signal. For both methods, two images are acquired with different energy spectra $\Phi(E)$. This leads to different flatfield values I_0^a and I_0^b, attenuation coefficients μ^a and μ^b, and image intensities after pre-processing M^a and M^b (superscripts a and b refer to different spectra). While M is proportional to the object thickness, the quotient R of corrected images is constant for small homogeneous materials:

$$\begin{cases} R_{\mathrm{m}} = \frac{M_{\mathrm{m}}^{\mathrm{a}}}{M_{\mathrm{m}}^{\mathrm{b}}} = \frac{\mu_{\mathrm{m}}^{\mathrm{a}} L_{\mathrm{m}}}{\mu_{\mathrm{m}}^{\mathrm{b}} L_{\mathrm{m}}} = \frac{\mu_{\mathrm{m}}^{\mathrm{a}}}{\mu_{\mathrm{m}}^{\mathrm{b}}} \\ R_{\mathrm{f}} = \frac{\mu_{\mathrm{f}}^{\mathrm{a}}}{\mu_{\mathrm{f}}^{\mathrm{b}}} \end{cases} . \tag{5.10}$$

Here the object with attenuation μ and thickness L is indicated by the subscript (m - for the main object, f - foreign object), R_{m} is the quotient value of the main object, and R_{f} – of the foreign object. For the main object with the present foreign object, the quotient image changes to

$$R = \frac{M^{\mathrm{a}}}{M^{\mathrm{b}}} = \frac{\alpha \beta R_{\mathrm{f}} + R_{\mathrm{m}}}{\alpha \beta + 1}, \tag{5.11}$$

where $\alpha = \frac{L_{\mathrm{f}}}{L_{\mathrm{m}}}$ and $\beta = \frac{\mu_{\mathrm{f}}^{\mathrm{b}}}{\mu_{\mathrm{m}}^{\mathrm{b}}}$. The presence of the foreign object can be detected by observing a difference ΔR between R and R_{m} given by the equation

$$\Delta R = \frac{\alpha \beta (R_{\mathrm{f}} - R_{\mathrm{m}})}{\alpha \beta + 1}. \tag{5.12}$$

Following Eq. 5.12, the FO can be found in a quotient image by locating a connected cluster of pixels with non-zero ΔR surrounded by noise. The magnitude of the background noise can be computed by estimating noise in Eq. 5.5 and propagating it through Eq. 5.9 and 5.12. However, Eq. 5.12 is written for a single pixel. A foreign object covers a region in the image, producing a distribution of ΔR. The exact value is different in each pixel of the FO region because the thicknesses of MO and FO are not constant. A computer vision algorithm is required to segment this distribution from noisy outliers by considering intensity gradients and average values for different scales. From an industrial point of view, this means that the detectability of the FO depends on its exact orientation at the time of acquisition. A large FO may have a small projection on the acquired image, and its quotient difference may be mistaken for noise.

Furthermore, there are many practical limitations hampering the viability of this approach for real data. Chapter 2 has demonstrated that Eq. 5.4 has a limited applicability when inspecting agricultural products. The value of R_{m} is not constant and depends slightly on the thickness. A special correction (Eq. 2.7) was required to mitigate this effect. If the main object is inhomogeneous, its R_{m} will depend on the thickness ratio of different materials, and the foreign object may be difficult to distinguish from a particular material combination of the main object. Finally, the quotient image has a high noise level, especially near the boundaries where both M^{a} and M^{b} are small.

Similar to other industrial computer vision problems, we use deep learning methods to solve the FOD problem. There is a large variety of deep learning methods, and we have chosen DCNNs because of their efficiency and wide adoption for different industrial tasks. Given the images M^a and M^b, the goal of the DCNN is to output the image mask of $L_{\mathrm{f}} > 0$. To train the DCNN, we provide a large set of training data using the manual segmentation serving as ground-truth.

Due to generalization from a large amount of data, DCNNs are able to achieve high segmentation accuracy despite all previously mentioned challenges. The exact performance level depends on many technical details, including the DCNN architecture, variety, and amount of training data. DCNNs as a deep learning approach can be replaced by other methods, such as the previously mentioned Vision Transformers. When choosing a DCNN architecture, we considered the accuracy level for well-known benchmarks (ImageNet) and the ability to train with limited data.

5.3.4 Performance analysis

Due to the large number of trainable parameters and low interpretability of DCNNs, it is difficult to predict how accurate the output would be given the image features. Eq. 5.12 gives an example of one feature ΔR that could be used to distinguish FO from the main object. Thus, we assume that the accuracy should be higher for large values of ΔR and lower for small ΔR. As a result, there should be a value of ΔR for which the algorithm is expected to work with high accuracy. Similar to Chapter 4, we use the POD analysis to characterize the performance of DCNNs and find out when they reach high accuracy.

A segmentation DCNN outputs a segmented image where every pixel is marked as corresponding to either the main object, foreign object or the background. We convert the segmentation into a binary result - detection. If the image contains no FO, the segmentation should not have any pixels marked as the FO to be considered correct. If there is an FO, the correct segmentation should have FO pixels that do not completely miss the correct FO location (Recall > 10%).

We assume that the Probability of Detection P depends on ΔR according to the equation

$$g(P) = c_0 + c_1 \Delta R, \tag{5.13}$$

where $g(P) = \log(-\log(1 - P))$ is an S-shaped link function. This equation corresponds to Eq. 4.10, with the defect size replaced by the quotient difference ΔR. Parameters c_0 and c_1 are determined using the Maximum Likelihood Estimation based on a sequence of ΔR and the corresponding detection outcomes (success with probability P and a failure with probability $1 - P$). The value of P describes the performance of the system. If the system design ensures that the value of ΔR is high enough for FOs of interest, then the value of P will be large enough for consistent detection. Thus, the value of ΔR can be seen as image quality in the context of detection.

We use the value of ΔR corresponding to $P = 90\%$ as the performance metric. ΔR characterizes all components of the inspection system: both physical properties of the studied object and experimental settings. The main disadvantage of using ΔR as a performance metric lies in the fact that a large variety of foreign objects could have the same contrast value. As shown in Eq. 5.12, the contrast depends on the FO thickness (L_f in α), its location in the main object (L_m in α). Furthermore, there is an implicit effect of the FO shape since ΔR is distributed

over the image region. This problem is not unique to our approach of performance evaluation and stems from the X-ray image formation.

POD curves provide a way to evaluate the accuracy of the image generator. Two real images of the same object are never the same due to noise fluctuations. Thus, a direct comparison between real and generated images would require to check whether the noise distribution is correct. However, we only need generated images to estimate the performance of DCNNs. Thus, if the POD curve based on real data is indistinguishable from one derived from generated data, the generator is sufficiently accurate.

5.3.5 Description of the experimental setup

All experiments were performed at the FleX-ray laboratory [36] of Centrum Wiskunde en Informatica in Amsterdam, the Netherlands (Fig. 5.3). The X-ray system in the laboratory was typically used for micrometer resolution computed tomography. Therefore, we had to make several adjustments to imitate industrial X-ray imaging with this system. Unlike line detectors commonly used in industrial applications, FleX-ray contained a planar detector. The area of the detector was $143\,\text{mm} \times 114\,\text{mm}$, the projection size was $956\,\text{px} \times 760\,\text{px}$ with a resolution of $150\,\mu\text{m}$. We performed the detector calibration [82] to determine parameters from Eq. 5.5 characterizing the setup.
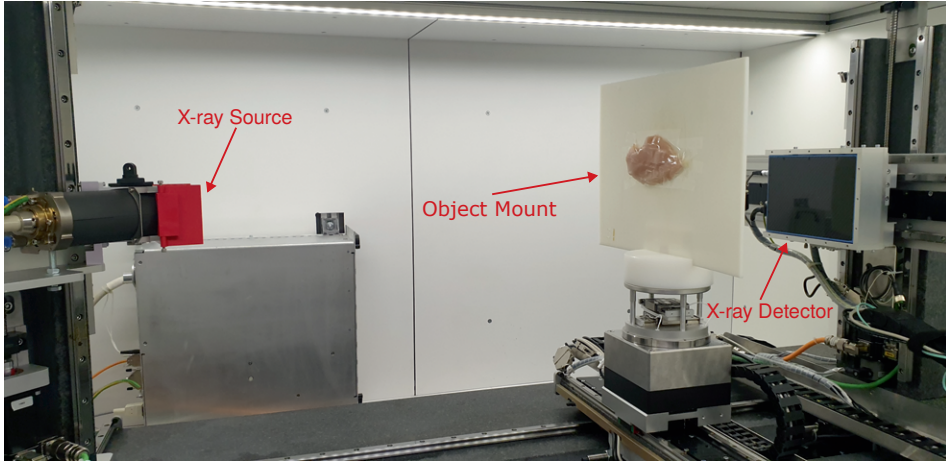

Figure 5.3: Photo of the experimental setup depicting the X-ray tube, object mounting system, and the detector. Chicken fillets are placed parallel to the detector to imitate projections in the industrial environment. For better visualization, the mount is placed closer to the tube than during the experiments

The image generation approach was tested on the problem of detecting bone fragments in chicken fillets. In this task, full chicken fillets were imaged with X-ray, and some of them contained pieces of rib bone of different lengths. The goal was to

detect as small bones as possible to satisfy health safety guidelines. Furthermore, the inspection could not use a long exposure since it would limit the conveyor belt throughput.

The X-ray inspection in the industrial environment would commonly use a detector below the conveyor belt and a tube above it, so a chicken fillet is parallel to the detector surface. We used a laboratory setup to acquire the data, and to achieve similar images we have attached fillets vertically to a plastic board (as shown on Fig. 5.3). Pieces of rib bones were cut from a chicken carcass. Bone fragments were placed on top and inside the fillet manually and randomly.

We have acquired 2 datasets with a total of 338 X-ray projections of chicken fillets. The experimental geometry and tube properties were the same as during calibration. The goal of the first dataset was to capture a variety of different products and FOs to train a DCNN. Furthermore, the first dataset was used to measure the fraction of false positives (products without FO where an FO was incorrectly detected). The second dataset presented a narrower task where two small bone fragments were placed in different locations of two fillets. The purpose of this dataset was to find the threshold value of ΔR that guarantees reliable detection.

The data acquisition was performed in two stages. First, we made a dataset consisting of 163 projections containing a bone fragment and 91 projections of boneless fillets. There were 14 different chicken fillets and 44 bone fragments of different sizes. All projections were recorded with an exposure time of 1 s. The size of the bone fragments ranged from 1.5 mm to 11 mm. The second dataset was made with two chicken fillets and two small bone fragments (2 mm and 3 mm) that were placed in different regions of the same fillet. This dataset contained 84 projections (two boneless and 82 with a bone fragment) acquired with 4 different values of exposure time: 1 s, 100 ms, 50 ms, and 20 ms.

It is important to note that our equipment is not specialized for high-throughput imaging. Our goal was to use a general-purpose scanner and produce data with a resolution and noise level comparable to the data from the conveyor line. In the industrial environment, line detectors with lower resolution (0.3-0.5 mm) and lower exposure time (0.1-10 ms) are used on conveyor belts running at tens of $^m/_{min}$. While our data are recorded with an exposure time higher than 20 ms, the X-ray intensity in a 150 μm pixel would be similar to the measurements with lower exposure time and larger pixel area (as illustrated by Eq. 5.7)

5.4 Results

5.4.1 System calibration

The calibration was performed with tube voltages of 40 kV and 90 kV with an additional filter of 50 μm copper. The Source-to-Object distance was 990 mm and the Source-to-Detector was 1059 mm leading to a small magnification of 1.07 and close to the maximum field of view. We acquired a series of 5000 flatfield images

with the same experimental settings. The series was converted into a distribution of the mean intensity value $\overline{y}$ and its standard deviation σ_y over the detector plane. We repeated the flatfield measurement for different values of the tube current leading to a sequence of $\{\overline{y}; \sigma_y\}$ for every pixel of the detector. Using Eq. 5.5 and linear regression, we extracted the values of the noise parameters:

$$\begin{cases} \overline{y} = I + d_{\mathrm{e}} \\ \sigma_y^2 = \lambda I + \sigma_{\mathrm{e}}^2 \end{cases} \Rightarrow \sigma_y^2 = \lambda(\overline{y} - d_{\mathrm{e}}) + \sigma_{\mathrm{e}}^2. \tag{5.14}$$

The distributions of the extracted parameters are shown on Fig. 5.4. These values are unique to our experimental setup, but the procedure can be applied to any detector. Furthermore, we computed an approximate value of the blur radius $\sigma = 0.8\,\mathrm{px}$ (Eq. 5.6). To obtain the estimate, we shifted a rectangular window around the image and computed the covariance matrix between the central pixel and the neighboring pixels. We made a fit of the covariance as a function of distance between pixels with a Gaussian function and used its RMS width as an approximation for σ.
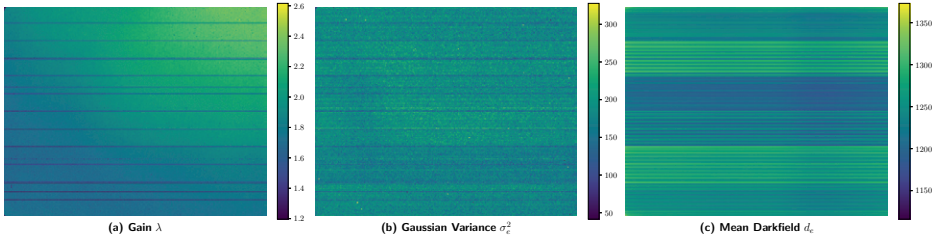

(a) Gain λ (b) Gaussian Variance σ_{ε}^2 (c) Mean Darkfield d_e

Figure 5.4: Distributions of the extracted noise parameters over the detector plane: (a) gain λ, (b) variance of darkfield σ_{e}^2, and (c) the mean values of darkfield d_{e}

Following Eq. 5.7, we used the flatfield measurement to determine the connection between the incident beam intensity I_0 and the exposure time t in a form

$$I_0 = kt. \tag{5.15}$$

It was computed that $k = 0.58\,^1/_{\mathrm{ms}}$ for $40\,\mathrm{kV}$ and $40\,\mathrm{W}$, and $k = 3.86\ ^1/_{\mathrm{ms}}$ for $90\,\mathrm{kV}$ and $45\,\mathrm{W}$.

After calibration, a noisy image corresponding to a certain exposure time t could be produced using a reference high-quality image. We considered an image to be high-quality if the intensity values are large enough so that $\sigma_y \ll \overline{y}$, and the exposure time of 1 s was considered to be sufficient. To generate a noisy image for the exposure time t, we estimated the beam intensity I_0 using Eq. 5.15. Then according to Beer's law (Eq. 5.2) and the reference image we computed the mean estimated values of intensity I. Finally, the noise was added using a random number generator according to the distribution from Eq. 5.5. An example of the high-quality reference image ($t = 1\,\mathrm{s}$) and noisy real and generated images

for different values of exposure time (100 ms, 50 ms, 20 ms) is shown on Fig. 5.5. According to our observations, including the blurring step from Eq. 5.6 improved visual similarity between the real and the generated images.
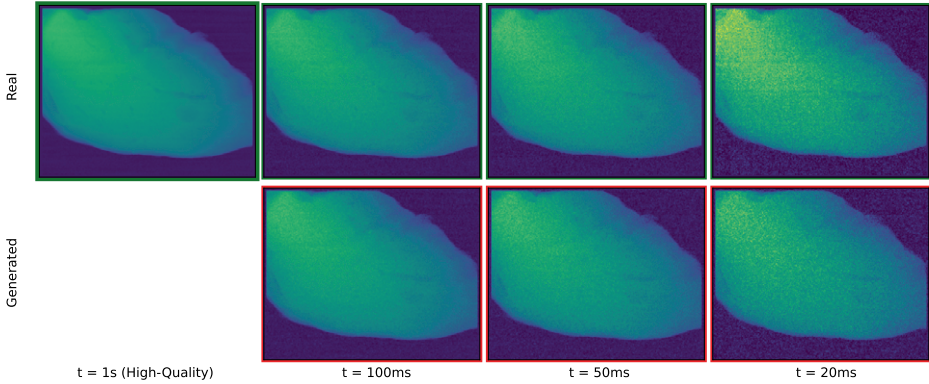

Figure 5.5: Comparison of the generated and real images corresponding to different values of exposure time: 1 s, 100 ms, 50 ms, 20 ms. The inspected object is a chicken fillet containing a bone fragment

5.4.2   Foreign object detection

To solve the FOD problem with DCNNss, we have used Segmentation Models package [73], as in the Appendix of Chapter 4. DeepLabV3Plus with EfficientNet-B0 encoder and encoder depth of 5 (4.5 M trainable parameters) was chosen as the DCNN architecture. We expected this architecture to provide higher segmentation accuracy than the MSD network used in Chapters 2 and 4. The network was trained on the first dataset with a split between training and validation data. The stopping criterion was achieving the minimum Dice's loss on the validation subset. To account for the random nature of the training process, the network was trained 100 times with the same data and different random seeds.

When tested on the projections containing bone fragments, the network was segmenting foreign objects with a mean recall value of 75%. An example of accurate segmentation is shown on Fig. 5.6. To assess the performance on boneless cases, we have converted the segmentation results into the detection results. Under this definition, the network scored 96% accuracy. Thus, with long exposure time, the selected DCNN architecture was able to solve the FOD problem despite the complicated morphological structure of projections and the limited amount of training data.

The generation approach and the setup parameters extracted from calibration were used to train DCNNs for FO segmentation with an exposure time of 100 ms, 50 ms, and 20 ms. We tested these networks on the second dataset containing small bone fragments in different locations. The POD curves as a function of FO

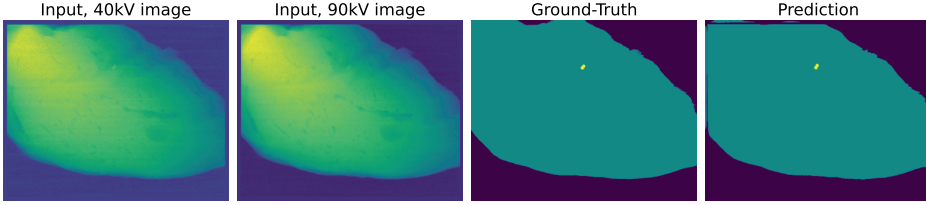| Input, 40kV image | Input, 90kV image | Ground-Truth | Prediction |

Figure 5.6: Example of an accurate segmentation. The DCNN takes an image with two channels (40kV and 90kV) as input. The segmentation mask is compared to the ground-truth to evaluate accuracy

contrast defined by 5.12 are shown on Fig. 5.7. We have observed that the value of contrast at which the probability of detection reaches 90% is similar for tests on real and generated data taking into account the variance from deep learning training. These values are shown in Table 5.1. Thus, the generation algorithms produced sufficiently accurate images, so the training on generated data was possible and the test results were similar for real and generated data.

Exposure time, [ms]	Contrast value for POD=90%	
	Generated test data	Real test data
1000	-	0.03 ± 0.03
100	0.10 ± 0.02	0.09 ± 0.02
50	0.13 ± 0.02	0.14 ± 0.17
20	0.18 ± 0.02	0.21 ± 0.04

Table 5.1: Comparison of performance metrics for different values of exposure time with tests on real and generated data. Values of contrast corresponding to the high accuracy are similar between real and generated test data with respect to the uncertainty

Fig. 5.8a compares POD curves corresponding to different values of exposure time. Every curve corresponds to tests on generated data (except for 1 s where real data would be indistinguishable from generated), and the fit parameters are averaged over 100 iterations of DCNN training. The figure shows the expected effect of exposure time on detection: for the same value of contrast the performance gets worse with less exposure. Following the Methods section, this can be explained by a higher noise level. Using the image generator, Fig. 5.8a could be extended to other values of exposure time and used to find the optimal system design. Safety guidelines lead to a threshold for contrast that characterizes FOs that have to be detected consistently. POD curves show which exposure times lead to a sufficiently high detection probability for that contrast, and the smallest time could be chosen to balance high throughput and product safety. Assuming that the probability of detection $P = 90\%$ is enough for reliable detection, we show on Fig. 5.8b
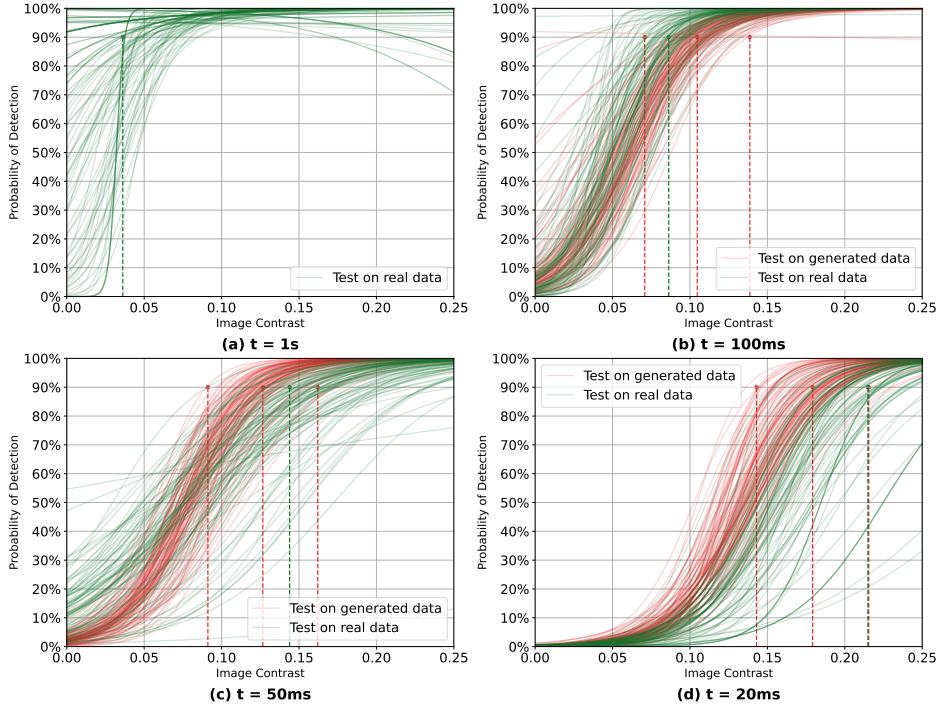
Figure 5.7: POD curves representing the performance of the DCNNs trained on generated data for 4 different values of exposure time: 1 s (a), 100 ms (b), 50 ms (c), 20 ms (d). Each line corresponds to an instance of DCNNs, 100 instances were trained for every exposure time. Red lines correspond to performance on generated data, green lines - on real data. The value of image contrast corresponding to the detection probability $P = 90\%$ is similar for real and generated data, the difference lies withing the confidence interval.

how the necessary value of contrast changes with the exposure time. For Fig. 5.8b we have estimated the performance level for 200 ms and 75 ms, where no experimental data are available. We have also experimented with smaller values of exposure time between 20 ms and 50 ms, but a high variance of the performance estimate was observed. This might indicate that the accuracy of the generation method is not sufficient around 20 ms. Due to the high variance of performance, it is not clear whether the linear dependence observed for a subset of points on Fig. 5.8b is a coincidence or a representation of DCNN properties. Given the low interpretability of deep learning methods, we do not attempt a theoretical explanation for this relation.
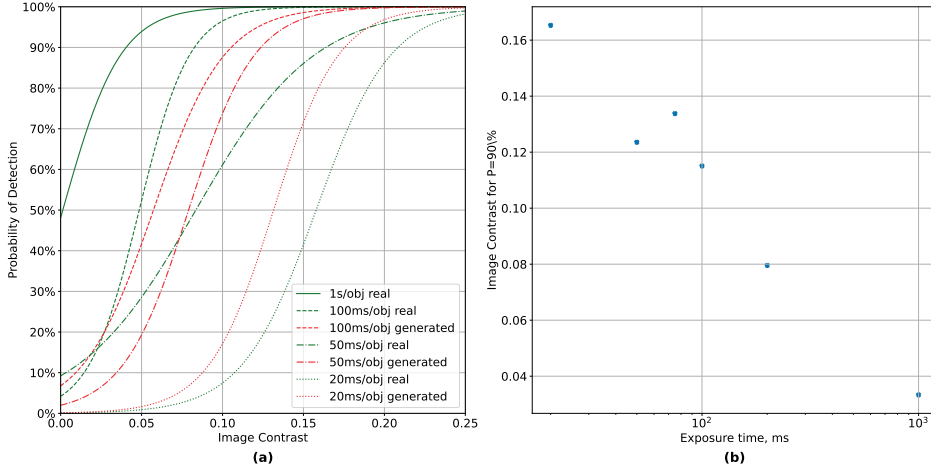
Figure 5.8: Comparison of POD curves corresponding to different values of exposure time (a). Each curve uses the average values of the fit parameters for curves shown on Fig. 5.7. Lower exposure time moves the POD curve to the right, indicating a worse detection rate at the same values of contrast. Image contrast is necessary for a good detection rate ($P = 90\%$) as a function of the exposure time (the plot uses the logarithmic scale for the exposure time) (b).

5.5   Discussion

Using image generation to predict inspection performance relies on the assumption that generated images could be similar enough to real ones. While we have checked this using POD curves, the problem of similarity between real and generated data could be addressed directly. Under low noise levels, even a difference between two images could be used to conclude if they are similar. However, high-throughput X-ray inspection is likely to work with noisy data. Thus, even two real images of the same object under the same conditions are never the same. To estimate similarity, it is necessary to check whether the noise pattern in two images corresponds to the same distribution. However, perfect correspondence is only expected for highly accurate Monte-Carlo algorithms which are not practical when the goal is to accelerate performance evaluation. Furthermore, it is unknown what level of similarity is necessary to get the same image analysis response. Thus, even if the accuracy of the generation model was quantitatively estimated, it would not be sufficient to conclude that it leads to the correct performance estimate.

The proposed image formation model has the advantage of being computationally efficient at the cost of not accounting for many experimental effects. Eq. 5.2 and 5.5 assume monochromatic X-ray radiation, and real data should be seen as a weighted sum of contributions of X-ray photons of different energy. The noise properties and the X-ray detection depend on the photon energy and change when

the attenuation in the inspected object affects the spectrum (beam-hardening). Fig. 5.4 should be seen as a map of average values, which are only accurate for the flatfield X-ray spectrum. In the object projection, the noise parameters should be corrected based on the thickness of the object corresponding to each pixel. However, extracting these corrections from the data is a challenging task, especially without the ability to measure the X-ray spectrum. This difference in noise parameters and the consequent inaccuracy of the generated noise may be one of the reasons why the proposed method loses accuracy for higher noise levels (20 ms as seen on Fig. 5.8). Furthermore, the noise is generated as uncorrelated and the correlation is added only with a Gaussian blur. In the generator, accurate simulation of the correlation would require modeling the scintillator and the electronics of the detector. Many issues of our model can be solved by Monte-Carlo simulation with an accurate model of the X-ray interactions (e.g. including scattering), X-ray detection, and signal formation. In practice, however, a balance between accuracy and computational cost is necessary to make data generation useful.

We have chosen a data-driven definition of image similarity - matching the POD curves. While this approach does not characterize the accuracy of the generator in all possible problems under different system settings, there is a direct confirmation that the generated data lead to the same performance estimate as the real data in a particular problem of interest. We assume that interpolation in terms of system settings is possible. If the generator is accurate for a number of values of different settings, it could be used for the values in-between. Thus, with a small number of real datasets under different settings, a wider range could be explored with the image generator.

The results show that a complete match between the POD curves for the real and generated data is unlikely. While the inaccuracy of the generator contributes to the difference between the POD curves, there is an inherent uncertainty in the proposed performance evaluation method. We indicate two main sources: the uncertainty of the DCNN performance and the uncertainty of the POD curves as a method of performance evaluation. To account for the random nature of DCNN training, we repeated the process 100 times and used the variance of the POD coefficients as a measure of the variance between the trained networks. Additional DCNN training was stopped when the standard deviation of the coefficients converged. Furthermore, the DCNN uncertainty is implicitly affected by biases in training and test data. If an image in the test dataset contains features not present during training (e.g. different shapes of bone or fillet), the network is more likely to fail. To ensure a good coverage of features, we used a large variety of bones and chicken fillets to make experimental datasets. Furthermore, the results on high-quality data showed that the network was able to perform the FOD with high accuracy, which would not be possible with significant biases in feature coverage.

The goal of the POD curve in our methodology is to provide a criterion (dual-energy quotient contrast ΔR) that guarantees high detection accuracy. As POD curves are computed using a log-likelihood fit, the resulting criterion could only be computed accurately with a proper test dataset. It should fully capture the transition from POD = 0% to POD = 100% and have enough undetectable and

always detectable cases. We observed that even a single outlier could significantly increase the uncertainty estimate for a POD curve. In the experimental data, it was crucial to have a wide range of ΔR in the test dataset, so even for $t = 20\,\mathrm{ms}$ it was possible to achieve a high detection rate.

The main downside of the proposed methodology is that the high performance criterion is linked to the image property ΔR. Industrial requirements would prefer to connect the performance to the physical properties of the foreign object, such as size, shape, and location in the main object. Since the POD curve is only a method of statistical analysis, it can be performed with respect to any feature. However, there is no guarantee that consistently high performance could be achieved with a simple constraint on a test sample. The image formation subsection shows how foreign object visibility depends on many parameters. The inspection system does not control the orientation of the FO with respect to the detector and the morphology of the main object around the FO. These issues are not unique to our methodology. The inability to control the view of the object affects all single projection inspection systems regardless of the image analysis and performance evaluation method. The second experimental dataset further highlights that foreign objects of the same size could be consistently detected or missed based on their location in the main object.

In the current implementation, our data generation method can be used to accelerate experimental design but there are still many challenges to overcome. We use the dataset of high-quality projections to simplify the image generation, and the results depend significantly on the size and variety of this dataset. To further reduce the number of experimental acquisitions, the generation method should be able to create new objects based on the known real data [5, 145]. While our generation model has been shown as sufficiently accurate for a particular bone detection task, it is not guaranteed in advance. Therefore, it is necessary to verify the generation accuracy before using the method for predictions. It is possible to apply this method to other problems, not only 2D inspection, but the required generation accuracy may depend significantly on the inspection technique. This problem be solved by making the image formation model more robust and accurate. The image formation model can also be improved by allowing the object volumes to be non-static during the measurement. This would allow the simulation of motion blur and the effect of conveyor belt vibration.

5.6   Conclusion

We have presented a computationally efficient model of X-ray image generation and used it to predict inspection accuracy for different X-ray system settings. Our method is a faster approach to system design than the traditional extensive experimental testing. The proposed methodology is aimed at industrial tasks and does not require specialized equipment to characterize the object and the inspection system. Due to several assumptions, our approach decreases in accuracy at higher noise levels and relies on a large amount of real high-quality X-ray data.

In the future, this method can be improved with a more robust image formation model, the generation of new objects, and extended to other X-ray inspection techniques.

6

CONCLUSION AND OUTLOOK

6.1 Contributions and limitations

The goal of the research presented in this thesis was to develop methods of X-ray image generation to improve industrial inspection of agricultural products. This thesis started from the assumption that there was sufficient fundamental knowledge for accurate X-ray simulation. However, two significant challenges limited its use for industrial tasks. First, there was a lack of detailed models describing the structure and variety of agricultural products. To address this challenge, it was proposed to extract the necessary information from the data. Second, accurate image generation was too computationally expensive to be widely used. Therefore, simplifications and approximations of the imaging model were considered to make image generation faster and easier to implement.

Chapters 2 to 5 focused on different aspects of product inspection and image generation. Chapter 2 provided an introduction to the problem of foreign object detection. It discussed the formulation of the problem and proposed different methods to solve it. Deep learning image processing was highlighted as a general method that only required application-specific data to achieve high accuracy. Chapters 3 to 5 addressed different challenges of image generation. The main contributions of the chapters corresponded to different elements of the image generator: the digital description of the objects, the model of imaging, and the system settings parameterization. In this chapter, the contributions and limitations of the methods proposed in this thesis are discussed.

Chapter 2 was centered around the difficulty of agricultural inspection due to unknown shape. Instead of relying on morphological properties, the possibility of separating objects based on material composition was considered. Dual-energy X-ray imaging was used to obtain more information about material properties. The pre-processing method developed in this chapter transformed raw data into images

in which the signal from the product was significantly reduced and the signal from the foreign object was enhanced. Using such images, foreign object segmentation and detection could be formulated as a simpler segmentation problem. While the method was validated on the problem of detecting bone fragments in chicken fillets, it was a general method that did not rely on task-specific assumptions. This pre-processing could be applied to a wide range of inspection problems where the product and the foreign object are approximately homogeneous and have different material compositions. The necessary information about material properties could then be extracted from the data without requiring prior knowledge.

The results of segmentation using a deep learning method were included in the chapter to illustrate the difference between morphology-based and material-based detection. The active contour segmentation was mostly based on material properties. The detection was based on dual-energy data and did not rely on shape knowledge. This approach resulted in high but not perfect ($\approx 90\%$) accuracy. The deep learning segmentation with a single-channel image could not fully rely on the difference in material properties due to the object overlap. However, it achieved similar accuracy by using extensive prior knowledge about the morphology of the objects. Finally, the deep learning method with a 2-channel image was able to reach almost perfect detection accuracy by combining morphological and material information. This high impact of morphological knowledge motivated the use of deep learning methods in later chapters.

This chapter illustrated the limited applicability of dual-energy X-ray imaging to agricultural inspection problems. In agriculture, both products and foreign objects often consist of organic materials. In terms of X-ray attenuation, their chemical composition is similar, and this similarity results in low material contrast relative to noise. However, this is not a problem of a specific image analysis method, but a general limitation of the imaging technique. The developed pre-processing method was shown to have little impact on the performance of the deep learning method. It was suggested that deep learning methods could learn similar data-driven pre-processing during training.

The main result of **Chapter 3** was that deep learning methods could be trained exclusively on generated images and then reach similarly high accuracy when applied to real X-ray images. Furthermore, two test cases presented in the chapter supported the claim that the generation method does not need to be perfectly accurate for successful training: scattering was not simulated, and the noise model and material properties were approximately estimated. The generalization of this claim is a challenging problem that was explored in later chapters, but not fundamentally explained. All successful applications of approximate generation methods in this thesis were limited to radiography (possibly dual-energy radiography). In radiography, inaccuracies in the generation model may be insignificant compared to the effects of object overlap and high noise levels. It is possible that more accurate inspection techniques (e.g., computed tomography) require more accurate generation methods.

A significant contribution of Chapter 3 was the methodology for generating a large number of images based on a single real object. This possibility started a

discussion about coverage - a variety of all possible shapes of products and foreign objects and their relative positions. From a deep learning perspective, coverage was related to a variety of images and was challenging to describe due to object overlap. However, with the method proposed in this chapter, it was possible to associate X-ray images with corresponding 3D volumes and define coverage at the volume level. As a result, coverage could be described by a variety of methods that do not necessarily rely on X-ray scans. For manufactured products, it would be possible to define coverage using blueprints and prior knowledge of foreign objects. Then even a single CT scan would not be necessary. However, for agricultural inspection problems without blueprints, a data-driven approach is more suitable. Therefore, this chapter proposed a method to define coverage using computational deformations of a single real object. The results of the chapter showed that even an approximate deformation model could be used to achieve high accuracy. Even if the generated volume could not exist in reality, the corresponding generated image did not necessarily reduce the accuracy of the deep learning method during training. However, simplifications and approximations of the deformation model are a limiting factor of the proposed method. To achieve higher detection accuracy (e.g., $> 99.9\%$), more realistic deformations and better coverage may be crucial.

The material properties of the objects used in this chapter were extracted from a segmented CT scan. The segmentation assumed that the object consisted of a limited set of materials and that it was possible to divide the object into homogeneous regions. This assumption was a limiting factor of the methodology, and it is important to note that homogeneity is not only related to the object but also to the resolution of the imaging. With a sufficiently high resolution, it would be possible to see fibers and even individual cells in agricultural products. However, at low resolution, the assumption of homogeneity can be seen as an approximation that different tissues have the same X-ray attenuation properties.

Chapter 4 explored the limitations of approximate generation models. The main contribution was a methodology to characterize the effect of various simplifications in a quantitative application-specific way. The results showed that the exclusion of X-ray scattering simulation could often be acceptable for the purpose of training an inspection method. As shown by the Probability of Detection curves, the absence of scattering had a significant effect only on the smallest detectable foreign objects, located in regions with a high scattering-to-primary ratio. This dependency on the properties of the foreign objects explained how inaccurate generators could be useful in practical applications. As long as the system is not used at the limits of detectability, the image processing method could be trained on images generated with simplifications and approximations.

This chapter left open the question of the extent to which the results can be generalized to different applications. In practice, it would be desirable to be able to predict the effect of scattering in advance without performing an application-specific study. This could be achieved by a general Probability of Detection analysis, where the frequency of detection in different applications is correlated with the same set of image features. However, finding such features would be a challenging problem. As shown in the previous chapters, morphology has a significant impact

on the detection problem, so it should be represented by the Probability of Detection arguments. While morphology was described by a cavity size in Chapter 4, this description did not work in Chapter 5, and a more complex measure of contrast was used instead. Therefore, generalization of the scattering influence results should first be performed for applications with similar object morphology. Further generalization would require a deeper understanding of the coverage and features used by deep learning methods to detect foreign objects.

Chapter 5 proposed a methodology for performing design optimization of inspection systems using generated images. The main contribution was the method of relating inspection performance to imaging settings. The image generator was implemented as a parameterizable method, where the parameters were related to the properties of the physical system. Similar to Chapter 3, this parameterization was achieved with a data-driven approach by performing a calibration of the system. As a result, accurate predictions of inspection accuracy were achieved without extensive prior knowledge of the inspection system.

In the case study presented in this chapter, only the exposure time was optimized. However, many other settings of interest could be related to the parameters of the generator without significant changes in the model of imaging. As pointed out in the chapter, some settings mainly affect the attenuation properties, while other settings influence the noise distribution. With the proposed method, optimizing the voltage of the X-ray source would be a challenging problem, since the X-ray energy affects the contrast between the product and the foreign object. However, imaging resolution, source current, and magnification could be represented in the parameters of the proposed model and optimized with the same approach.

The Probability of Detection curves in Chapter 5 highlighted the problem of statistical analysis of inspection performance. When applied to a real problem, it was difficult to determine which properties of the object were significantly correlated with the foreign object detection rate. Furthermore, these properties were image-based, not object-based. When the inspection system is analyzed from an industrial perspective, its reliability needs to be expressed in terms of object properties (e.g., what size of foreign object can be detected). However, the same foreign object may be detected or missed based on a number of additional, effectively random factors (e.g., product pose, foreign object rotation). As a result, even for large foreign objects, the probability of detection may be lower than expected. Therefore, the coverage of training data is even more important for accurate probability estimation.

Deep learning was an important aspect of the research presented in this thesis. The goal of image generation is motivated by deep learning methods that require large amounts of data to train. This raises a general question: what are the limitations of the deep learning approach for product inspection? As pointed out in this chapter, many reported results have limited generalizability due to the use of data-driven methods. Reliable detection and system design would benefit greatly from interpretable image analysis and decision-making. Does this mean that a data-driven approach should be replaced by an interpretable method created by experts in the field?

In many applications, deep learning image analysis is used as a cost-efficient alternative to classical image processing. While classical methods can reach high accuracy, they should be implemented by computer vision experts with deep knowledge of a specific application. Their efforts are not always generalizable. Some insights can be used in different applications, but many corner cases and small optimizations are task-specific. On the other hand, accurate deep learning methods can be obtained without a deep understanding of the task. An expert is needed to design the general architecture, and task-specific insights are gained by training on large amounts of data. The deep learning approach has the advantage of easier fine-tuning to a specific problem at the expense of explainability and reliability.

The low interpretability of deep learning stems from a large number of parameters. However, as shown in this thesis, the complicated morphology of objects makes the detection problem challenging in general. In high-throughput X-ray imaging, object overlap and high noise levels create uncertainty: different products with and without foreign objects could have indistinguishable images. As a result, accurate detection is unlikely to be achieved by a classical method with a handful of parameters. Both experts manually creating an image analysis procedure and deep learning methods require a variety of data to make accurate decisions in the context of the specific problem. Therefore, a robust and accurate image generator would be helpful for both training a deep learning method and implementing an interpretable classical method.

6.2   Outlook

Chapters 3 to 5 focused on different aspects of implementing an image generator. While Chapter 3 focused on the creation of artificial objects, there were no artificial objects in Chapter 5 because the focus was on simulating different settings of the system. In practice, the methods from all chapters should be combined into a single image generator. While combining them is not a fundamental problem, many technical improvements would be necessary to keep the accuracy of the general generator at a high level. In each chapter, it was observed that the accuracy decreased as the generator deviated from experimental measurements and predicted images based on interpolations from known data. To create a sufficiently accurate general method, it may be necessary to improve the quality of each element, from the deformation model to the noise calibration. In this section, different ways to improve the methods proposed in this thesis are discussed. Furthermore, possible applications of the new research developments are considered.

In recent years, there has been significant progress in the development of deep learning methods. Transformer architecture [43, 150] has become state-of-the-art in image processing and outperformed the deep convolutional neural networks that were used in this thesis. However, this change has a limited impact on the developed methods. In this thesis, deep learning methods were used as a black box. Both transformers and convolutional networks are supervised machine learning algorithms and can be switched without changing the methods proposed in

the thesis. Furthermore, it is argued that the convolutional architecture has the potential for improvement to compete with newer methods [90]. The research on Explainable AI [47] could have a significant impact on the problems of system design and definition of coverage discussed in this thesis.

Another question raised by deep learning developments is whether an image generator can be implemented as a learned method. There has been significant progress in text-to-image generation [22]. Generative adversarial networks have been successfully used to generate new images to train deep learning methods [132]. A similar approach is domain transfer - a method of generating images of objects based on their images made with different acquisition methods (e.g., generating CT images based on MRI measurements) [114]. While these methods can produce images that look realistic, they could also mislead the inspection system by creating non-existent details and model-specific artifacts [160]. Learned methods are more difficult to interpret and debug than conventional image generators. In the methods implemented in this thesis, there is a clear distinction between the model of the object, the method of X-ray simulation, the noise model, etc. There is a lot of prior knowledge about each subsystem, and the missing information can be obtained with an appropriate calibration. On the other hand, this prior knowledge does not benefit learned methods that rely on building complex system models solely from the data.

The definition of coverage is one of the fundamental problems of this thesis. The use of deep learning methods was motivated by the complicated morphology and variance of agricultural objects. While Chapter 3 proposed a solution for approximating the coverage with volume deformations, the proposed model had its limitations. The perfect solution would be to create a parameterizable mathematical model describing the shape of agricultural products. Its parameters would correspond to the variance in the biological processes that determine the resulting shape. While there is some progress in this direction [152], it is not clear whether such models could be implemented in the near future. An alternative solution is to use data-driven methods that are trained on a large variety of volumes corresponding to the same product to generate new volumes. The possibility of 3D volume generation has been shown with diffusion models [75]. However, it is not guaranteed that this method would be able to capture the morphological variety of agricultural products without significant artifacts. This data-driven approach can be seen as an improvement over the affine transform deformations proposed in Chapter 3.

In this thesis, information required for image generation was usually extracted from the data due to the lack of prior knowledge. In medical imaging, a lot of information about X-ray attenuation properties of different human tissues is available [64]. Medical phantoms with properties similar to human tissues are created to calibrate imaging systems experimentally [40]. While similar expertise is not currently available for agricultural problems and products, it could be achieved with time. Now dual-energy inspection systems for many tasks (e.g. detecting contaminants in packaged food) are commercially available [24, 159]. If X-ray inspection continues to extend to more challenging problems, it would be possible to

justify more in-depth studies of X-ray properties of different agricultural products and create standard imaging protocols. More information about products and standard setups would simplify the development of accurate image generators.

To achieve wider use, image generation methods should be computationally efficient. Deep learning methods can be used to accelerate X-ray simulation and noise models without significant loss of accuracy. Convolutional neural networks have been successfully used to imitate scattering distributions [94]. These methods help to balance generation accuracy and computational cost. While deep learning is less accurate than a Monte-Carlo method, it is significantly faster and provides higher accuracy than a model that completely excludes scattering. Similarly, deep noise estimation methods [164] can be used as more accurate models of the detector noise. A significant challenge for both approaches is the implicit influence of the studied object. Both scattering and noise properties depend on the energy of the X-ray photons. While their energy distribution is approximately known from the source, it is affected by the absorption in the studied object. As a result, the generalizability of the learned methods needs to be regularly verified due to the implicit assumptions about the problem and the object.

This thesis has focused on the inspection of a single X-ray image. It is important to note that this type of imaging has fundamental limitations in the ability to detect foreign objects, even if a perfect image generator is constructed. In a perfect scenario, the generated images are indistinguishable from real images, and the speed of generation is sufficiently fast to have the dataset as large as necessary to train a deep learning method. Furthermore, the generator has complete coverage for a particular problem and can create images of all possible combinations of products and foreign objects. In this case, the probability of detection is no longer influenced by the data and is completely determined by the imaging system. Foreign objects could still be missed due to low contrast relative to noise or insufficient resolution. Furthermore, if there is a large morphological variance, it may be impossible to distinguish two products based on a single X-ray image.

Many X-ray techniques provide better contrast between organic materials and can be used in agricultural inspections. Examples include phase-contrast imaging [49], small-angle scattering [115], and, to a lesser extent, spectral imaging with photon counting detectors [113]. While these techniques can be as fast as radiography, they require significantly more expensive equipment. As a result, more advanced methods are usually tested in the laboratory and not often used in the factory. However, if increased food quality requirements motivate the use of more advanced imaging techniques, the generation methods will need to be adapted. With the approach developed in this thesis, different acquisition methods can be implemented by changing the model of imaging while all existing information about the products, detectors, X-ray source, and performance metrics can be reused from similar applications.

BIBLIOGRAPHY

[1] S. Agostinelli, J. Allison, K. al Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. J. N. I. Barrand, et al. «GEANT4 — a Simulation Toolkit». *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303 (cit. on p. 64).

[2] I. Ahmed, M. Ahmad, and G. Jeon. «Integrating Digital Twins and Deep Learning for Medical Image Analysis in the Era of COVID-19». *Virtual Reality & Intelligent Hardware* 4.4 (2022), pp. 292–305 (cit. on p. 86).

[3] J. Alles and R. F. Mudde. «Beam Hardening: Analytical Considerations of the Effective Attenuation Coefficient of X-ray Tomography». *Medical physics* 34.7 (2007), pp. 2882–2889 (cit. on p. 92).

[4] V. Andriiashen and D. Kozhevnikov. «Development of the Projection-Based Material Decomposition Algorithm for Multienergy CT». *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.4 (2020), pp. 517–527 (cit. on p. 131).

[5] V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. «CT-based Data Generation for Foreign Object Detection on a Single X-ray Projection». *Scientific Reports* 13.1 (2023), p. 1881 (cit. on pp. 45, 66, 104, 131).

[6] V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. «X-Ray Image Generation as a Method of Performance Prediction for Real-Time Inspection: A Case Study». *Journal of Nondestructive Evaluation* 43.3 (2024) (cit. on pp. 85, 131).

[7] V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. «Quantifying the Effect of X-ray Scattering for Data Generation in Real-Time Defect Detection». *Journal of X-Ray Science and Technology* 43 (2024), pp. 1099–1119 (cit. on pp. 61, 131).

[8] V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. «Unsupervised Foreign Object Detection Based on Dual-Energy Absorptiometry in the Food Industry». *Journal of Imaging* 7.7 (2021), p. 104 (cit. on pp. 21, 131).

[9] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang. «MedGAN: Medical Image Translation Using GANs». *Computerized medical imaging and graphics* 79 (2020), p. 101684 (cit. on p. 64).

[10] M. Arnold and A. Gramza-Michalowska. «Recent Development on the Chemical Composition and Phenolic Extraction Methods of Apple (Malus Domestica)—A Review». *Food and Bioprocess Technology* (2023), pp. 1–42 (cit. on p. 11).

[11]   I. Aydin, M. Karakose, and A. K. I. N. Erhan. «A New Approach for Baggage Inspection by Using Deep Convolutional Neural Networks». In: *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*. IEEE, 2018, pp. 1–6 (cit. on p. 63).

[12]   A. Badal, D. Sharma, C. G. Graff, R. Zeng, and A. Badano. «Mammography and Breast Tomosynthesis Simulator for Virtual Clinical Trials». *Computer Physics Communications* 261 (2021), p. 107779 (cit. on p. 13).

[13]   C. R. Baldo, T. L. Fernandes, G. D. Donatelli, and W. Dewulf. «Digital Twin as a Tool to Select CT Scan Parameters». In: *Brazilian Technology Symposium*. Springer, 2020, pp. 561–569 (cit. on p. 86).

[14]   G. T. Barnes. «Contrast and Scatter in X-ray Imaging.» *Radiographics : a review publication of the Radiological Society of North America, Inc* 11.2 (1991), pp. 307–323 (cit. on p. 64).

[15]   C. Bellon and G.-R. Jaenisch. «aRTist–Analytical RT Inspection Simulation Tool». In: *Proc DIR*. 2007, pp. 25–27 (cit. on pp. 62, 64, 86, 87).

[16]   E. Bergbäck Knudsen, A. Prodi, J. Baltser, M. Thomsen, P. Kjær Willendrup, M. Sanchez del Rio, C. Ferrero, E. Farhi, K. Haldrup, A. Vickery, et al. «McXtrace: A Monte Carlo Software Package for Simulating X-ray Optics, Beamlines and Experiments». *Journal of Applied Crystallography* 46.3 (2013), pp. 679–696 (cit. on pp. 86, 87).

[17]   N. Bhatia, D. Tisseur, F. Buyens, and J. M. Létang. «Scattering Correction Using Continuously Thickness-Adapted Kernels». *NDT & E International* 78 (2016), pp. 52–60 (cit. on p. 64).

[18]   X. Bian, S. N. Lim, and N. Zhou. «Multiscale Fully Convolutional Network with Application to Industrial Inspection». In: *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2016, pp. 1–8 (cit. on p. 63).

[19]   B. A. Bircher, S. Wyss, D. Gage, A. Küng, C. Körner, and F. Meli. «High-Resolution X-ray Computed Tomography for Additive Manufacturing: Towards Traceable Porosity Defect Measurements Using Digital Twins». In: *Joint Special Interest Group Meeting between Euspen and ASPE Advancing Precision in Additive Manufacturing Inspire AG (St. Gallen, Switzerland)*. 2021 (cit. on p. 86).

[20]   J. M. Boone and J. A. Seibert. «An Accurate Method for Computer-Generating Tungsten Anode x-Ray Spectra from 30 to 140 kV». *Medical physics* 24.11 (1997), pp. 1661–1670 (cit. on p. 26).

[21]   J. M. Boone and J. A. Seibert. «An Analytical Model of the Scattered Radiation Distribution in Diagnostic Radiology». *Medical physics* 15.5 (1988), pp. 721–725 (cit. on p. 65).

[22]   A. Borji. «Generated Faces in the Wild: Quantitative Comparison of Stable Diffusion, Midjourney and Dall-e 2». 2022. arXiv: 2210.00586 (cit. on p. 112).

[23]   S. Cardoso, O. Gonçalves, and H. Schechter. «Evaluation of Scatter-to-Primary Ratio in Radiological Conditions». *Applied Radiation and Isotopes* 67.4 (2009), pp. 544–548 (cit. on p. 65).

[24]   S. Carmignato, W. Dewulf, and R. Leach. *Industrial X-ray Computed Tomography*. Vol. 10. Springer, 2018 (cit. on p. 112).

[25]   J.-F. Carrier, L. Archambault, L. Beaulieu, and R. Roy. «Validation of GEANT4, an Object-Oriented Monte Carlo Toolkit, for Simulations in Medical Physics». *Medical physics* 31.3 (2004), pp. 484–492 (cit. on p. 10).

[26]   H. Cen and R. Lu. «Optimization of the Hyperspectral Imaging-Based Spatially-Resolved System for Measuring the Optical Properties of Biological Materials». *Optics express* 18.16 (2010), pp. 17412–17432 (cit. on p. 5).

[27]   T. F. Chan and L. A. Vese. «Active Contours without Edges». *IEEE Transactions on image processing* 10.2 (2001), pp. 266–277 (cit. on pp. 23, 30).

[28]   V. Chelladurai, K. Karuppiah, D. S. Jayas, P. G. Fields, and N. D. G. White. «Detection of Callosobruchus Maculatus (F.) Infestation in Soybean Using Soft X-ray and NIR Hyperspectral Imaging Techniques». *Journal of Stored Products Research* 57 (2014), pp. 43–48 (cit. on p. 46).

[29]   G. Chen, Y. Guo, T. Katagiri, H. Song, T. Tomizawa, N. Yusa, and H. Hashizume. «Multivariate Probability of Detection (POD) Analysis Considering the Defect Location for Long-Range, Non-Destructive Pipe Inspection Using Electromagnetic Guided Wave Testing». *NDT & E International* 124 (2021), p. 102539 (cit. on p. 73).

[30]   H. Chen, X. Nie, S. Gan, Y. Zhao, and H. Qiu. «Interfacial Imperfection Detection for Steel-Concrete Composite Structures Using NDT Techniques: A State-of-the-Art Review». *Engineering Structures* 245 (2021), p. 112778 (cit. on p. 61).

[31]   X. Chen, H. Jing, Y. Tao, and X. Cheng. «Pattern Classification for Boneless Poultry Inspection Using Combined X-ray/Laser 3D Imaging». In: *Optical Sensors and Sensing Systems for Natural Resources and Food Safety and Quality*. Vol. 5996. International Society for Optics and Photonics, 2005, 59960T (cit. on pp. 45, 46).

[32]   Z. Chen, Y. Tao, and X. Chen. «Multiresolution Local Contrast Enhancement of X-ray Images for Poultry Meat Inspection». *Applied Optics* 40.8 (2001), pp. 1195–1200 (cit. on pp. 21, 23).

[33]   M. Chmiel, M. Roszko, L. Adamczak, T. Florowski, and D. Pietrzak. «Influence of Storage and Packaging Method on Chicken Breast Meat Chemical Composition and Fat Oxidation». *Poultry science* 98.6 (2019), pp. 2679–2690 (cit. on p. 11).

[34] C.-L. Chuang, C.-S. Ouyang, T.-T. Lin, M.-M. Yang, E.-C. Yang, T.-W. Huang, C.-F. Kuei, A. Luke, and J.-A. Jiang. «Automatic X-ray Quarantine Scanner and Pest Infestation Detector for Agricultural Products». *Computers and electronics in agriculture* 77.1 (2011), pp. 41–59 (cit. on p. 21).

[35] S. B. Coban, V. Andriiashen, P. S. Ganguly, M. van Eijnatten, and K. J. Batenburg. «Parallel-Beam X-ray CT Datasets of Apples with Internal Defects and Label Balancing for Machine Learning». 2020. arXiv: 2012.13346 (cit. on p. 8).

[36] S. B. Coban, F. Lucka, W. J. Palenstijn, D. Van Loo, and K. J. Batenburg. «Explorative Imaging and Its Implementation at the FleX-ray Laboratory». *Journal of Imaging* 6.4 (Apr. 2020), p. 18 (cit. on pp. 28, 52, 79, 96).

[37] E. Commission. «Commission Implementing Regulation (EU) 2019/627 of 15 March 2019 Laying down Uniform Practical Arrangements for the Performance of Official Controls on Products of Animal Origin Intended for Human Consumption in Accordance with Regulation (EU) 2017/625 of the European Parliament and of the Council and Amending Commission Regulation (EC) No 2074/2005 as Regards Official Controls». *Off. J. Eur. Union L* 131 (2019), pp. 51–100 (cit. on p. 1).

[38] E.-E. Commission et al. «Commission Regulation (EC) No. 543/2008 of 16 June 2008 Laying down Detailed Rules for the Application of Council Regulation (EC) No. 1234/2007 as Regards the Marketing Standards for Poultry Meat». *Off J Europ Comm* 157 (2008), pp. 46–87 (cit. on p. 2).

[39] D. M. Cunha, A. Tomal, and M. E. Poletti. «Evaluation of Scatter-to-Primary Ratio, Grid Performance and Normalized Average Glandular Dose in Mammography by Monte Carlo Simulation Including Interference and Energy Broadening Effects». *Physics in Medicine & Biology* 55.15 (2010), p. 4335 (cit. on p. 65).

[40] L. A. DeWerd and M. Kissick. *The Phantoms of Medical and Health Physics.* Springer, 2014 (cit. on p. 112).

[41] F. di Franco, A. Sarno, G. Mettivier, A. M. Hernandez, K. Bliznakova, J. M. Boone, and P. Russo. «GEANT4 Monte Carlo Simulations for Virtual Clinical Trials in Breast X-ray Imaging: Proof of Concept». *Physica Medica* 74 (2020), pp. 133–142 (cit. on pp. 13, 47, 79).

[42] L. G. Divyanth, V. Chelladurai, M. Loganathan, D. S. Jayas, and P. Soni. «Identification of Green Gram (Vigna Radiata) Grains Infested by Callosobruchus Maculatus through X-ray Imaging and GAN-Based Image Augmentation». *Journal of Biosystems Engineering* (2022), pp. 1–16 (cit. on p. 45).

[43]   A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Un-
       terthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al. «An Image
       Is Worth 16x16 Words: Transformers for Image Recognition at Scale». 2020.
       arXiv: 2010.11929 (cit. on pp. 80, 87, 111).

[44]   Z. Du, Y. Hu, N. Ali Buttar, and A. Mahmood. «X-Ray Computed Tomog-
       raphy for Quality Inspection of Agricultural Products: A Review». *Food
       science & nutrition* 7.10 (2019), pp. 3146–3160 (cit. on pp. 22, 85).

[45]   *Dual Energy X Ray Absorptiometry for Bone Mineral Density and Body
       Composition Assessment.* Human Health Series 15. Vienna: INTERNA-
       TIONAL ATOMIC ENERGY AGENCY, 2011 (cit. on p. 23).

[46]   F. Duan, S. Yin, P. Song, W. Zhang, C. Zhu, and H. Yokoi. «Automatic
       Welding Defect Detection of X-Ray Images by Using Cascade Adaboost
       with Penalty Term». *IEEE access : practical innovations, open solutions* 7
       (2019), pp. 125929–125938 (cit. on pp. 45, 47).

[47]   R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian,
       Z. Wen, T. Shah, G. Morgan, et al. «Explainable AI (XAI): Core Ideas,
       Techniques, and Solutions». *ACM Computing Surveys* 55.9 (2023), pp. 1–
       33 (cit. on p. 112).

[48]   T. Eelbode, P. Sinonquel, F. Maes, and R. Bisschops. «Pitfalls in Train-
       ing and Validation of Deep Learning Systems». *Best Practice & Research
       Clinical Gastroenterology* 52 (2021), p. 101712 (cit. on p. 34).

[49]   H. Einarsdóttir, M. J. Emerson, L. H. Clemmensen, K. Scherer, K. Willer,
       M. Bech, R. Larsen, B. K. Ersbøll, and F. Pfeiffer. «Novelty Detection of
       Foreign Objects in Food Using Multi-Modal X-ray Imaging». *Food control*
       67 (2016), pp. 39–47 (cit. on p. 113).

[50]   A. Erasmus, T. D. Brunet, and E. Fisher. «What Is Interpretability?» *Phi-
       losophy & Technology* 34.4 (2021), pp. 833–862 (cit. on p. 4).

[51]   L. A. Feldkamp, L. C. Davis, and J. W. Kress. «Practical Cone-Beam Al-
       gorithm». *Josa a* 1.6 (1984), pp. 612–619 (cit. on p. 54).

[52]   M. Ferguson, R. Ak, Y.-T. T. Lee, and K. H. Law. «Automatic Localization
       of Casting Defects with Convolutional Neural Networks». In: *2017 IEEE
       International Conference on Big Data (Big Data)*. IEEE, 2017, pp. 1726–
       1735 (cit. on pp. 45, 47).

[53]   E. Fredenberg. «Spectral and Dual-Energy X-ray Imaging for Medical Ap-
       plications». *Nuclear Instruments and Methods in Physics Research Section
       A: Accelerators, Spectrometers, Detectors and Associated Equipment* 878
       (2018), pp. 74–87 (cit. on p. 90).

[54]   N. Freud, J.-M. Létang, and D. Babot. «A Hybrid Approach to Simulate
       Multiple Photon Scattering in X-ray Imaging». *Nuclear Instruments and
       Methods in Physics Research Section B: Beam Interactions with Materials
       and Atoms* 227.4 (2005), pp. 551–558 (cit. on p. 69).

[55]   G. A. Georgiou. «PoD Curves, Their Derivation, Applications and Limitations». *Insight-Non-Destructive Testing and Condition Monitoring* 49.7 (2007), pp. 409–414 (cit. on pp. 15, 63, 86, 87).

[56]   P. Getreuer. «Chan-Vese Segmentation». *Image Processing On Line* 2 (2012), pp. 214–224 (cit. on p. 36).

[57]   L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter, and L. Kagal. «Explaining Explanations: An Overview of Interpretability of Machine Learning». In: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. IEEE, 2018, pp. 80–89 (cit. on p. 4).

[58]   A. J. Gislason, A. G. Davies, and A. R. Cowen. «Dose Optimization in Pediatric Cardiac X-ray Imaging». *Medical physics* 37.10 (2010), pp. 5258–5269 (cit. on p. 87).

[59]   S. S. Gleason, M. J. Paulus, and J. A. Mullens. *Automatic Detection of Bone Fragments in Poultry Using Multi-Energy x-Rays*. Google Patents, Apr. 9, 2002 (cit. on p. 23).

[60]   Q. Gong, J. A. Greenberg, R.-I. Stoian, D. Coccarelli, E. Vera, and M. E. Gehm. «Rapid Simulation of X-ray Scatter Measurements for Threat Detection via GPU-based Ray-Tracing». *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 449 (2019), pp. 86–93 (cit. on pp. 62, 64).

[61]   Q. Gong, R.-I. Stoian, D. S. Coccarelli, J. A. Greenberg, E. Vera, and M. E. Gehm. «Rapid Simulation of X-ray Transmission Imaging for Baggage Inspection via GPU-based Ray-Tracing». *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* 415 (2018), pp. 100–109 (cit. on pp. 47, 62, 64, 86, 87).

[62]   L. Grevillot, D. J. Boersma, H. Fuchs, A. Aitkenhead, A. Elia, M. Bolsa, C. Winterhalter, M. Vidal, S. Jan, U. Pietrzyk, et al. «GATE-RTion: A GATE/Geant4 Release for Clinical Applications in Scanned Ion Beam Therapy». *Medical Physics* 47.8 (2020), pp. 3675–3681 (cit. on p. 11).

[63]   C. J. Griffith. «Food Safety: Where from and Where To?» *British Food Journal* 108.1 (2006), pp. 6–15 (cit. on p. 1).

[64]   H. J. Griffiths. «Tissue Substitutes in Radiation Dosimetry and Measurement. No. 4». *Radiology* 173.1 (1989), pp. 202–202 (cit. on pp. 26, 112).

[65]   P. Gupta, Z. Sinno, J. L. Glover, N. G. Paulter, and A. C. Bovik. «Predicting Detection Performance on Security X-ray Images as a Function of Image Quality». *IEEE Transactions on Image Processing* 28.7 (2019), pp. 3328–3342 (cit. on p. 88).

[66]   R. P. Haff and N. Toyofuku. «X-Ray Detection of Defects and Contaminants in the Food Industry». *Sensing and Instrumentation for Food Quality and Safety* 2.4 (2008), pp. 262–273 (cit. on pp. 6, 45).

[67]   A. M. Hafiz and G. M. Bhat. «A Survey on Instance Segmentation: State of the Art». *International journal of multimedia information retrieval* 9.3 (2020), pp. 171–189 (cit. on p. 23).

[68]   B. J. Heismann, J. Leppert, and K. Stierstorfer. «Density and Atomic Number Measurements with Spectral X-Ray Attenuation Method». *Journal of applied physics* 94.3 (2003), pp. 2073–2079 (cit. on p. 26).

[69]   A. A. Hendriksen. *Ahendriksenh/Msd_pytorch: V0.7.2.* Version v0.7.2. Zenodo, Dec. 2019 (cit. on pp. 39, 71).

[70]   D. Hendrycks, S. Basart, N. Mu, S. Kadavath, F. Wang, E. Dorundo, R. Desai, T. Zhu, S. Parajuli, M. Guo, et al. «The Many Faces of Robustness: A Critical Analysis of out-of-Distribution Generalization». In: *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 2021, pp. 8340–8349 (cit. on p. 66).

[71]   D. Hendrycks and K. Gimpel. «A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks». In: *International Conference on Learning Representations.* 2017 (cit. on p. 66).

[72]   G. Hernández and F. Fernández. «Xpecgen: A Program to Calculate x-Ray Spectra Generated in Tungsten Anodes.» *J. Open Source Softw.* 1.7 (2016), p. 62 (cit. on p. 69).

[73]   P. Iakubovskii. *Segmentation Models Pytorch.* GitHub, 2019 (cit. on pp. 82, 99).

[74]   *Non-Destructive Testing — Industrial Computed Radiography with Storage Phosphor Imaging Plates — Part 2: General Principles for Testing of Metallic Materials Using X-rays and Gamma Rays.* Standard. Geneva, CH: International Organization for Standardization, Sept. 2017 (cit. on p. 61).

[75]   A. Jain, B. Mildenhall, J. T. Barron, P. Abbeel, and B. Poole. «Zero-Shot Text-Guided Object Generation with Dream Fields». In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2022, pp. 867–876 (cit. on p. 112).

[76]   S. Jan, D. Benoit, E. Becheva, T. Carlier, F. Cassol, P. Descourt, T. Frisson, L. Grevillot, L. Guigues, L. Maigne, et al. «GATE V6: A Major Enhancement of the GATE Simulation Platform Enabling Modelling of CT and Radiotherapy». *Physics in Medicine & Biology* 56.4 (2011), p. 881 (cit. on pp. 64, 68, 86, 87).

[77]   K. A. Jessen. «Balancing Image Quality and Dose in Diagnostic Radiology». *European Radiology Supplements* 14 (2004), pp. 9–18 (cit. on p. 65).

[78]   D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks. «Characterising the Digital Twin: A Systematic Literature Review». *CIRP journal of manufacturing science and technology* 29 (2020), pp. 36–52 (cit. on p. 86).

[79]   R. Kakarala and A. O. Hero. «On Achievable Accuracy in Edge Localization». *IEEE Transactions on Pattern Analysis & Machine Intelligence* 14.07 (1992), pp. 777–781 (cit. on p. 87).

[80]   D. Kaur and Y. Kaur. «Various Image Segmentation Techniques: A Review». *International Journal of Computer Science and Mobile Computing* 3.5 (2014), pp. 809–814 (cit. on p. 23).

[81]   A. Khan, A. D. Vibhute, S. Mali, and C. H. Patil. «A Systematic Review on Hyperspectral Imaging Technology with a Machine and Deep Learning Methodology for Agricultural Applications». *Ecological Informatics* 69 (2022), p. 101678 (cit. on p. 5).

[82]   A. C. Konstantinidis, M. B. Szafraniec, R. D. Speller, and A. Olivo. «The Dexela 2923 CMOS X-ray Detector: A Flat Panel Detector Based on CMOS Active Pixel Sensors for Medical Imaging Applications». *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 689 (2012), pp. 12–21 (cit. on pp. 50, 96).

[83]   A. Kostenko, V. Andriiashen, and K. J. Batenburg. «Registration-Based Multi-Orientation Tomography». *Optics express* 26.22 (2018), pp. 28982–28995 (cit. on p. 131).

[84]   N. Kotwaliwale, K. Singh, A. Kalne, S. N. Jha, N. Seth, and A. Kar. «X-Ray Imaging Methods for Internal Quality Evaluation of Agricultural Produce». *Journal of food science and technology* 51.1 (2014), pp. 1–15 (cit. on pp. 21, 45).

[85]   J.-S. Kwon, J.-M. Lee, and W.-Y. Kim. «Real-Time Detection of Foreign Objects Using X-ray Imaging for Dry Food Manufacturing Line». In: *2008 IEEE International Symposium on Consumer Electronics*. IEEE, 2008, pp. 1–4 (cit. on p. 21).

[86]   D.-H. Lee, E.-S. Kim, J.-S. Cho, J.-H. Ryu, and B.-S. Min. «A Two-Stage Automatic Labeling Method for Detecting Abnormal Food Items in X-ray Images». *Journal of Food Measurement and Characterization* (2022), pp. 1–11 (cit. on p. 47).

[87]   K. Lee, S. Yi, S. Hyun, and C. Kim. «Review on the Recent Welding Research with Application of CNN-based Deep Learning Part I: Models and Applications». *Journal of Welding and Joining* 39.1 (2021), pp. 10–19 (cit. on p. 62).

[88]   J. Leuschner, M. Schmidt, P. S. Ganguly, V. Andriiashen, S. B. Coban, A. Denker, D. Bauer, A. Hadjifaradji, K. J. Batenburg, P. Maass, et al. «Quantitative Comparison of Deep Learning-Based Image Reconstruction Methods for Low-Dose and Sparse-Angle CT Applications». *Journal of Imaging* 7.3 (2021), p. 44 (cit. on p. 131).

[89]   H. Lim, J. Lee, S. Lee, H. Cho, H. Lee, and D. Jeon. «Low-Density Foreign Body Detection in Food Products Using Single-Shot Grid-Based Dark-Field X-ray Imaging». *Journal of Food Engineering* 335 (2022), p. 111189 (cit. on p. 6).

[90]   Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. «A Convnet for the 2020s». In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, pp. 11976–11986 (cit. on pp. 80, 112).

[91]   Ó. López-Campos, J. C. Roberts, I. L. Larsen, N. Prieto, M. Juárez, M. E. Dugan, and J. L. Aalhus. «Rapid and Non-Destructive Determination of Lean Fat and Bone Content in Beef Using Dual Energy X-ray Absorptiometry». *Meat science* 146 (2018), pp. 140–146 (cit. on p. 21).

[92]   J. Ma, D.-W. Sun, H. Pu, J.-H. Cheng, and Q. Wei. «Advanced Techniques for Hyperspectral Imaging in the Food Industry: Principles and Recent Applications». *Annual review of food science and technology* 10.1 (2019), pp. 197–220 (cit. on p. 5).

[93]   J. Ma, Z. Liang, Y. Fan, Y. Liu, J. Huang, W. Chen, and H. Lu. «Variance Analysis of X-ray CT Sinograms in the Presence of Electronic Noise Background». *Medical physics* 39 (7 Part 1 2012), pp. 4051–4065 (cit. on pp. 12, 87).

[94]   J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß. «Deep Scatter Estimation (DSE): Accurate Real-Time Scatter Estimation for X-ray CT Using a Deep Convolutional Neural Network». *Journal of Nondestructive Evaluation* 37.3 (2018), pp. 1–9 (cit. on pp. 64, 113).

[95]   C. J. Martin, P. F. Sharp, and D. G. Sutton. «Measurement of Image Quality in Diagnostic Radiology». *Applied radiation and isotopes* 50.1 (1999), pp. 21–38 (cit. on p. 65).

[96]   L. Martin, A. Tuysuzoglu, W. C. Karl, and P. Ishwar. «Learning-Based Object Identification and Segmentation Using Dual-Energy CT Images for Security». *IEEE Transactions on Image Processing* 24.11 (2015), pp. 4069–4081 (cit. on p. 23).

[97]   J. Masci, U. Meier, D. Ciresan, J. Schmidhuber, and G. Fricout. «Steel Defect Classification with Max-Pooling Convolutional Neural Networks». In: *The 2012 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2012, pp. 1–6 (cit. on p. 63).

[98]   S. K. Mathanker, P. R. Weckler, N. Wang, T. Bowser, and N. O. Maness. «Local Adaptive Thresholding of Pecan X-ray Images: Reverse Water Flow Method». *Transactions of the ASABE* 53.3 (2010), pp. 961–969 (cit. on p. 23).

[99]   S. K. Mathanker, P. R. Weckler, and T. J. Bowser. «X-Ray Applications in Food and Agriculture: A Review». *Transactions of the ASABE* 56.3 (2013), pp. 1227–1239 (cit. on pp. 45, 85).

[100] J. R. Mathiassen, E. Misimi, M. Bondø, E. Veliyulin, and S. O. Østvik. «Trends in Application of Imaging Technologies to Inspection of Fish and Fish Products». *Trends in Food Science & Technology* 22.6 (2011), pp. 257–275 (cit. on p. 85).

[101] A. M. Matser, B. Krebbers, R. W. van den Berg, and P. V. Bartels. «Advantages of High Pressure Sterilisation on Quality of Food Products». *Trends in Food Science & Technology* 15.2 (2004), pp. 79–85 (cit. on p. 2).

[102] D. Mery. «Computer Vision for X-ray Testing». *Switzerland: Springer International Publishing* 10 (2015), pp. 978–3 (cit. on pp. 45, 47, 62).

[103] D. Mery, I. Lillo, H. Loebel, V. Riffo, A. Soto, A. Cipriano, and J. M. Aguilera. «Automated Fish Bone Detection Using X-ray Imaging». *Journal of Food Engineering* 105.3 (2011), pp. 485–492 (cit. on pp. 23, 24, 45, 46).

[104] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, and M. Carrasco. «GDXray: The Database of X-ray Images for Nondestructive Testing». *Journal of Nondestructive Evaluation* 34.4 (2015), pp. 1–12 (cit. on pp. 8, 47, 62, 64).

[105] H. S. El-Mesery, H. Mao, and A. E.-F. Abomohra. «Applications of Non-Destructive Technologies for Agricultural and Food Products Quality Inspection». *Sensors* 19.4 (2019), p. 846 (cit. on p. 21).

[106] C. Miao, L. Xie, F. Wan, C. Su, H. Liu, J. Jiao, and Q. Ye. «Sixray: A Large-Scale Security Inspection x-Ray Benchmark for Prohibited Item Discovery in Overlapping Images». In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, pp. 2119–2128 (cit. on p. 7).

[107] V. I. Mikla and V. V. Mikla. *Medical Imaging Technology*. Elsevier, 2013 (cit. on p. 5).

[108] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. «Image Segmentation Using Deep Learning: A Survey». *IEEE transactions on pattern analysis and machine intelligence* 44.7 (2021), pp. 3523–3542 (cit. on p. 23).

[109] A. Mitra. *Fundamentals of Quality Control and Improvement*. John Wiley & Sons, 2016 (cit. on p. 1).

[110] Y. Mo, Y. Wu, X. Yang, F. Liu, and Y. Liao. «Review the State-of-the-Art Technologies of Semantic Segmentation Based on Deep Learning». *Neurocomputing* 493 (2022), pp. 626–646 (cit. on p. 23).

[111] D. B. Mumford and J. Shah. «Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems». *Communications on pure and applied mathematics* (1989) (cit. on p. 30).

[112] K. Naresh, K. A. Khan, R. Umer, and W. J. Cantwell. «The Use of X-ray Computed Tomography for Design and Process Modeling of Aerospace Composites: A Review». *Materials & Design* 190 (2020), p. 108553 (cit. on p. 85).

[113] J. Nguyen, P.-A. Rodesch, D. Richtsmeier, K. Iniewski, and M. Bazalova-Carter. «Optimization of a CZT Photon Counting Detector for Contaminant Detection». *Journal of Instrumentation* 16.11 (2021), P11015 (cit. on p. 113).

[114] S. Niu, M. Liu, Y. Liu, J. Wang, and H. Song. «Distant Domain Transfer Learning for Medical Imaging». *IEEE Journal of Biomedical and Health Informatics* 25.10 (2021), pp. 3784–3793 (cit. on p. 112).

[115] S. Olakanmi, C. Karunakaran, and D. Jayas. «Applications of X-ray Micro-Computed Tomography and Small-Angle X-ray Scattering Techniques in Food Systems: A Concise Review». *Journal of Food Engineering* 342 (2023), p. 111355 (cit. on pp. 85, 113).

[116] İ. E. Parlak and E. Emel. «Deep Learning-Based Detection of Aluminum Casting Defects and Their Types». *Engineering Applications of Artificial Intelligence* 118 (2023), p. 105636 (cit. on p. 63).

[117] D. M. Pelt and J. A. Sethian. «A Mixed-Scale Dense Convolutional Neural Network for Image Analysis». *Proceedings of the National Academy of Sciences* 115.2 (2018), pp. 254–259 (cit. on pp. 24, 33, 71).

[118] L. F. A. Pereira, E. Janssens, G. D. Cavalcanti, R. Tsang, M. Van Dael, P. Verboven, B. Nicolai, and J. Sijbers. «Inline Discrete Tomography System: Application to Agricultural Product Inspection». *Computers and electronics in agriculture* 138 (2017), pp. 117–126 (cit. on p. 22).

[119] W. Rawat and Z. Wang. «Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review». *Neural computation* 29.9 (2017), pp. 2352–2449 (cit. on p. 45).

[120] V. Rebuffel and J.-M. Dinten. «Dual-Energy X-ray Imaging: Benefits and Limits». *Insight-non-destructive testing and condition monitoring* 49.10 (2007), pp. 589–594 (cit. on pp. 90, 93).

[121] Á. Rodríguez-Sánchez, A. Thompson, L. Körner, N. Brierley, and R. Leach. «Review of the Influence of Noise in X-ray Computed Tomography Measurement Uncertainty». *Precision Engineering* 66 (2020), pp. 382–391 (cit. on p. 87).

[122] O. Ronneberger, P. Fischer, and T. Brox. «U-Net: Convolutional Networks for Biomedical Image Segmentation». In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18.* Springer, 2015, pp. 234–241 (cit. on p. 33).

[123] E.-P. Rührnschopf and K. Klingenbeck. «A General Framework and Review of Scatter Correction Methods in X-Ray Cone-Beam Computerized Tomography. Part 1: Scatter Compensation Approaches». *Medical physics* 38.7 (2011), pp. 4296–4311 (cit. on p. 64).

[124] A. Sarno, G. Mettivier, F. Di Lillo, R. M. Tucciariello, K. Bliznakova, and P. Russo. «Normalized Glandular Dose Coefficients in Mammography, Digital Breast Tomosynthesis and Dedicated Breast CT». *Physica Medica* 55 (2018), pp. 142–148 (cit. on p. 79).

[125] D. Sarrut, M. Bardiès, N. Boussion, N. Freud, S. Jan, J.-M. Létang, G. Loudos, L. Maigne, S. Marcatili, T. Mauxion, et al. «A Review of the Use and Potential of the GATE Monte Carlo Simulation Code for Radiation Therapy and Dosimetry Applications». *Medical physics* 41 (6Part1 2014), p. 064301 (cit. on p. 11).

[126] R. Schoonhoven, J.-W. Buurlage, D. M. Pelt, and K. J. Batenburg. «Real-Time Segmentation for Tomographic Imaging». In: *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2020, pp. 1–6 (cit. on p. 24).

[127] S. Seabold and J. Perktold. «Statsmodels: Econometric and Statistical Modeling with Python». In: *9th Python in Science Conference*. 2010 (cit. on p. 72).

[128] J. H. Siewerdsen and D. A. Jaffray. «Optimization of X-ray Imaging Geometry (with Specific Application to Flat-Panel Cone-Beam Computed Tomography)». *Medical physics* 27.8 (2000), pp. 1903–1914 (cit. on p. 87).

[129] K. Simonyan and A. Zisserman. «Very Deep Convolutional Networks for Large-Scale Image Recognition». 2014. arXiv: `1409.1556` (cit. on p. 33).

[130] D. Smalley, S. Baker, B. Baldonado, J. Castaneda, A. Corredor, J. H. Clayton, L. Fegenbush, C. Gautier, A. Gehring, T. Haines, et al. «Image Restoration of High-Energy x-Ray Radiography with a Scintillator Blur Model». *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 968 (2020), p. 163910 (cit. on pp. 69, 92).

[131] X. Song, B. W. Pogue, S. Jiang, M. M. Doyley, H. Dehghani, T. D. Tosteson, and K. D. Paulsen. «Automated Region Detection Based on the Contrast-to-Noise Ratio in near-Infrared Tomography». *Applied optics* 43.5 (2004), pp. 1053–1062 (cit. on p. 87).

[132] V. Sorin, Y. Barash, E. Konen, and E. Klang. «Creating Artificial Images for Radiology Applications Using Generative Adversarial Networks (GANs)–a Systematic Review». *Academic radiology* 27.8 (2020), pp. 1175–1185 (cit. on p. 112).

[133] M. Sun and J. M. Star-Lack. «Improved Scatter Correction Using Adaptive Scatter Kernel Superposition». *Physics in Medicine & Biology* 55.22 (2010), p. 6695 (cit. on pp. 11, 64).

[134] M. Tan and Q. Le. «Efficientnet: Rethinking Model Scaling for Convolutional Neural Networks». In: *International Conference on Machine Learning*. PMLR, 2019, pp. 6105–6114 (cit. on p. 50).

[135] Y. Tao and J. G. Ibarra. «Thickness-Compensated X-ray Imaging Detection of Bone Fragments in Deboned Poultry—Model Analysis». *Transactions of the ASAE* 43.2 (2000), p. 453 (cit. on p. 23).

[136]   A. Tempelaere, T. Van De Looverbosch, K. Kelchtermans, P. Verboven, T. Tuytelaars, and B. Nicolai. «Synthetic Data for X-ray CT of Healthy and Disordered Pear Fruit Using Deep Learning». *Postharvest Biology and Technology* 200 (2023), p. 112342 (cit. on p. 64).

[137]   H. Tian, T. Wang, Y. Liu, X. Qiao, and Y. Li. «Computer Vision Technology in Agricultural Automation—A Review». *Information Processing in Agriculture* 7.1 (2020), pp. 1–19 (cit. on p. 5).

[138]   A.-A. Tulbure, A.-A. Tulbure, and E.-H. Dulf. «A Review on Modern Defect Detection Models Using DCNNs–Deep Convolutional Neural Networks». *Journal of Advanced Research* 35 (2022), pp. 33–48 (cit. on p. 87).

[139]   T. Tyystjärvi, I. Virkkunen, P. Fridolf, A. Rosell, and Z. Barsoum. «Automated Defect Detection in Digital Radiography of Aerospace Welds Using Deep Learning». *Welding in the World* 66.4 (2022), pp. 643–671 (cit. on p. 63).

[140]   K. Urazoe, N. Kuroki, A. Maenaka, H. Tsutsumi, M. Iwabuchi, K. Fuchuya, T. Hirose, and M. Numa. «Automated Fish Bone Detection in X-ray Images with Convolutional Neural Network and Synthetic Image Generation». *IEEJ Transactions on Electrical and Electronic Engineering* 16.11 (2021), pp. 1510–1517 (cit. on pp. 46, 47).

[141]   G. Vachtsevanos, W. D. Daley, B. S. Heck, A. Yezzi, and Y. Ding. «Fusion of Visible and X-ray Sensing Modalities for the Enhancement of Bone Detection in Poultry Products». In: *Biological Quality and Precision Agriculture II*. Vol. 4203. International Society for Optics and Photonics, 2000, pp. 102–110 (cit. on p. 23).

[142]   W. Van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers. «Fast and Flexible X-ray Tomography Using the ASTRA Toolbox». *Optics express* 24.22 (2016), pp. 25129–25147 (cit. on p. 54).

[143]   M. Van Dael, P. Verboven, A. Zanella, J. Sijbers, and B. Nicolai. «Combination of Shape and X-ray Inspection for Apple Internal Quality Control: In Silico Analysis of the Methodology Based on X-ray Computed Tomography». *Postharvest Biology and Technology* 148 (2019), pp. 218–227 (cit. on p. 23).

[144]   T. Van De Looverbosch, M. H. R. Bhuiyan, P. Verboven, M. Dierick, D. Van Loo, J. De Beenbouwer, J. Sijbers, and B. Nicolaï. «Nondestructive Internal Quality Inspection of Pear Fruit by X-ray CT Using Machine Learning». *Food Control* 113 (2020), p. 107170 (cit. on p. 24).

[145]   T. Van De Looverbosch, J. He, A. Tempelaere, K. Kelchtermans, P. Verboven, T. Tuytelaars, J. Sijbers, and B. Nicolai. «Inline Nondestructive Internal Disorder Detection in Pear Fruit Using Explainable Deep Anomaly Detection on X-ray Images». *Computers and Electronics in Agriculture* 197 (2022), p. 106962 (cit. on pp. 8, 47, 64, 86, 104).

[146]   T. Van De Looverbosch, E. Raeymaekers, P. Verboven, J. Sijbers, and B. Nicolai. «Non-Destructive Internal Disorder Detection of Conference Pears by Semantic Segmentation of X-ray CT Scans Using Deep Learning». *Expert Systems with Applications* 176 (2021), p. 114925 (cit. on pp. 23, 47).

[147]   S. Van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. «Scikit-Image: Image Processing in Python». *PeerJ* 2 (2014), e453 (cit. on p. 55).

[148]   M. van Dael, S. Lebotsa, E. Herremans, P. Verboven, J. Sijbers, U. L. Opara, P. J. Cronje, and B. M. Nicolaï. «A Segmentation and Classification Algorithm for Online Detection of Internal Disorders in Citrus Using X-ray Radiographs». *Postharvest Biology and Technology* 112 (2016), pp. 205–214 (cit. on p. 21).

[149]   S. K. Vasan, C. Osmond, D. Canoy, C. Christodoulides, M. J. Neville, C. Di Gravio, C. H. Fall, and F. Karpe. «Comparison of Regional Fat Measurements by Dual-Energy X-ray Absorptiometry and Conventional Anthropometry and Their Association with Markers of Diabetes and Cardiovascular Disease Risk». *International journal of obesity* 42.4 (2018), pp. 850–857 (cit. on p. 23).

[150]   A. Vaswani. «Attention Is All You Need». *Advances in Neural Information Processing Systems* (2017) (cit. on p. 111).

[151]   P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors. «SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python». *Nature Methods* 17 (2020), pp. 261–272 (cit. on p. 55).

[152]   S. Visa, C. Cao, B. M. Gardener, and E. van der Knaap. «Modeling of Tomato Fruits into Nine Shape Categories Using Elliptic Fourier Shape Modeling and Bayesian Classification of Contour Morphometric Data». *Euphytica: Netherlands journal of plant breeding* 200 (2014), pp. 429–439 (cit. on pp. 4, 112).

[153]   Ž. Vujović et al. «Classification Model Evaluation Metrics». *International Journal of Advanced Computer Science and Applications* 12.6 (2021), pp. 599–606 (cit. on p. 15).

[154]   J. Wang, C. Miao, W. Wang, and X. Lu. «Research of X-ray Nondestructive Detector for High-Speed Running Conveyor Belt with Steel Wire Ropes». In: *Electronic Imaging and Multimedia Technology V*. Vol. 6833. SPIE, 2007, pp. 482–490 (cit. on p. 91).

[155] B. R. Whiting. «Signal Statistics in X-ray Computed Tomography». In: *Medical Imaging 2002: Physics of Medical Imaging*. Vol. 4682. SPIE, 2002, pp. 53–60 (cit. on pp. 69, 92).

[156] B. R. Whiting, P. Massoumzadeh, O. A. Earl, J. A. O'Sullivan, D. L. Snyder, and J. F. Williamson. «Properties of Preprocessed Sinogram Data in X-ray Computed Tomography». *Medical physics* 33.9 (2006), pp. 3290–3303 (cit. on pp. 69, 92).

[157] J. Yang, S. Li, Z. Wang, H. Dong, J. Wang, and S. Tang. «Using Deep Learning to Detect Defects in Manufacturing: A Comprehensive Survey and Current Challenges». *Materials* 13.24 (2020), p. 5755 (cit. on p. 62).

[158] M. Yosifov, M. Reiter, S. Heupl, C. Gusenbauer, B. Fröhler, R. Fernández-Gutiérrez, J. De Beenhouwer, J. Sijbers, J. Kastner, and C. Heinzl. «Probability of Detection Applied to X-ray Inspection Using Numerical Simulations». *Nondestructive Testing and Evaluation* 37.5 (2022), pp. 536–551 (cit. on p. 63).

[159] A. Yousefi, H. Jafari, M. Khorsandi, and A. Faezmehr. «Assessment of a Low-Cost Commercial CCD for Use in X-ray Imaging». *Applied Radiation and Isotopes* 190 (2022), p. 110476 (cit. on p. 112).

[160] N. Yu, L. S. Davis, and M. Fritz. «Attributing Fake Images to Gans: Learning and Analyzing Gan Fingerprints». In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2019, pp. 7556–7566 (cit. on p. 112).

[161] M. T. Zeegers, T. van Leeuwen, D. M. Pelt, S. B. Coban, R. van Liere, and K. J. Batenburg. «A Tomographic Workflow to Enable Deep Learning for X-ray Based Foreign Object Detection». 2022. arXiv: 2201.12184 (cit. on pp. 24, 47, 51).

[162] M. T. Zeegers. *A Collection of 131 CT Datasets of Pieces of Modeling Clay Containing Stones - Part 1 of 5*. Zenodo, Jan. 2022 (cit. on pp. 46, 51).

[163] D. Zeng, J. Huang, Z. Bian, S. Niu, H. Zhang, Q. Feng, Z. Liang, and J. Ma. «A Simple Low-Dose x-Ray CT Simulation from High-Dose Scan». *IEEE transactions on nuclear science* 62.5 (2015), pp. 2226–2233 (cit. on p. 13).

[164] Q. Zhang, A. Nicolson, M. Wang, K. K. Paliwal, and C. Wang. «DeepMMSE: A Deep Learning Approach to MMSE-based Noise Power Spectral Density Estimation». *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 28 (2020), pp. 1404–1415 (cit. on p. 113).

[165] X. Zhu, Y. Wei, Y. Lu, M. Zhao, K. Yang, S. Wu, H. Zhang, and K. K. Wong. «Comparative Analysis of Active Contour and Convolutional Neural Network in Rapid Left-Ventricle Volume Quantification Using Echocardiographic Imaging». *Computer Methods and Programs in Biomedicine* 199 (2021), p. 105914 (cit. on p. 23).

[166] F. T. Zohora and K. C. Santosh. «Circular Foreign Object Detection in Chest X-ray Images». In: *Recent Trends in Image Processing and Pattern Recognition: First International Conference, RTIP2R 2016, Bidar, India, December 16–17, 2016, Revised Selected Papers 1.* Springer, 2017, pp. 391–401 (cit. on p. 87).

LIST OF PUBLICATIONS

Publications that are part of this thesis:

1. V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. «Unsupervised Foreign Object Detection Based on Dual-Energy Absorptiometry in the Food Industry». *Journal of Imaging* 7.7 (2021), p. 104.

2. V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. «CT-based Data Generation for Foreign Object Detection on a Single X-ray Projection». *Scientific Reports* 13.1 (2023), p. 1881.

3. V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. «Quantifying the Effect of X-ray Scattering for Data Generation in Real-Time Defect Detection». *Journal of X-Ray Science and Technology* 43 (2024), pp. 1099–1119.

4. V. Andriiashen, R. van Liere, T. van Leeuwen, and K. J. Batenburg. «X-Ray Image Generation as a Method of Performance Prediction for Real-Time Inspection: A Case Study». *Journal of Nondestructive Evaluation* 43.3 (2024).

Publications that are not part of this thesis:

1. A. Kostenko, V. Andriiashen, and K. J. Batenburg. «Registration-Based Multi-Orientation Tomography». *Optics express* 26.22 (2018), pp. 28982–28995.

2. V. Andriiashen and D. Kozhevnikov. «Development of the Projection-Based Material Decomposition Algorithm for Multienergy CT». *IEEE Transactions on Radiation and Plasma Medical Sciences* 5.4 (2020), pp. 517–527.

3. J. Leuschner, M. Schmidt, P. S. Ganguly, V. Andriiashen, S. B. Coban, A. Denker, D. Bauer, A. Hadjifaradji, K. J. Batenburg, P. Maass, et al. «Quantitative Comparison of Deep Learning-Based Image Reconstruction Methods for Low-Dose and Sparse-Angle CT Applications». *Journal of Imaging* 7.3 (2021), p. 44.

SUMMARY

Applying X-ray image generation to design inspection systems

Product inspection is an important part of modern industries. In particular, the agricultural industry requires inspection systems to maintain high quality while producing large volumes of food products. A common task is the detection of foreign objects - unwanted objects that may be present in products and compromise their quality. Automating detection is a challenging problem because food products have a natural variety of shapes and material compositions due to the variance in biological processes. Unlike manufactured products, food items do not have a blueprint. Recent advances in computer vision and deep learning have been instrumental in creating robust and accurate automated foreign object detection methods. These methods have been particularly successful with digital color photographs. In many modern factories, sets of cameras are installed on the conveyor belt, and specialized software analyzes photographs taken by these cameras to detect foreign objects. However, digital photography is limited to surface inspection, and other techniques are considered to detect foreign objects inside products.

There is rising interest in the application of X-ray imaging for food inspection. X-ray radiation is frequently used in security and medical applications to inspect the internal structure of objects. X-ray radiography - the acquisition of a single X-ray image - is well suited for the inspection of food products on the conveyor belt because it is a fast method with a limited dose of radiation. There are successful applications of X-ray systems in the inspection of packaged food (detecting plastic and metal residue), fish, and meat (bone fragments). Two challenges limit the wider use of X-ray imaging. First, many food products and foreign objects are made of organic materials and have low contrast on X-ray images. Therefore, many imaging settings, such as acquisition time, voltage, and resolution, need to be optimized to produce images with the best visibility of the foreign object. Second, X-ray images are often challenging to analyze, even for a trained human expert (e.g. a hospital radiologist). X-rays penetrate through the object, and the images capture all the features of the object located at different depths simultaneously. Food products have a complex internal structure that results in many intensity gradients in the X-ray images. With this structure as a background, small foreign objects could be missed.

Despite these challenges, deep learning image analysis methods could achieve high accuracy of foreign object detection on X-ray images. However, large datasets of images are required to train these methods, and such datasets are scarce in industrial applications. There is a limited number of publicly available datasets containing images of agricultural products. These datasets are usually general, and they do not specialize in a particular agricultural product to cover a variety of internal structures (e.g., images of different apples of the same cultivar). Furthermore, there are no standard settings for X-ray imaging of agricultural products.

Available data often correspond to different settings of imaging, and comparing images of the same object under different settings is rarely possible. As a result, existing data are often not large enough to train accurate detection methods and do not explore different imaging settings to optimize the design of the inspection system.

In this thesis, image generation is proposed to mitigate the lack of real data for the purpose of deep learning training and system design. The image generator is used as a prediction method (the general structure is shown in Fig. S1). Given information about the object and the imaging system, it predicts what an X-ray image would look like without performing an experiment. Such a prediction can be highly accurate because the physics of X-ray radiation and imaging are well known. Generation does not solve a fundamental problem of deep learning methods, which require large volumes of data to be accurate. Instead, the goal is to make image generation easier and faster than experimental data collection.
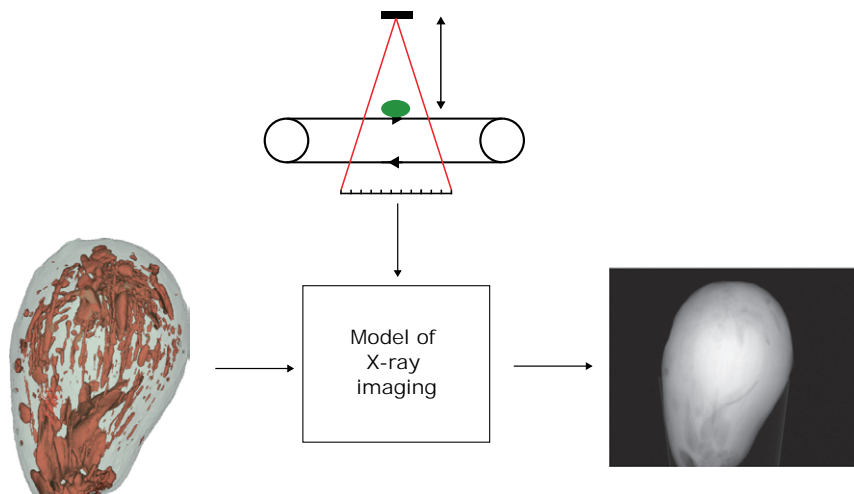
Figure S1: The structure of an image generator as a prediction method. It receives information about the object as input and computes the X-ray image according to the model of imaging. Different properties of the imaging system are represented by the parameters of the model and affect the generated image.

A significant challenge of image generation for industrial tasks is to provide the necessary information about the object and the system. Highly accurate generation techniques require the chemical composition of the object and the spatial distribution of different chemical elements. This information is rarely available for agricultural products. Most generation methods need a description of the shape of the object. While obtaining a digital volume description is feasible for a single product, large-scale generation requires these data for a large variety of objects. Finally, a sophisticated model of the X-ray source and detector is required to pre-

dict how the image is influenced by changes in imaging settings. In this thesis, an alternative data-driven approach is used to obtain the necessary information without complicated and expensive analysis. A set of calibrations and measurements with an X-ray system is used to extract the necessary properties from the data. This approach relies on assumptions and simplifications that reduce the resulting accuracy, but it is easier to implement under industrial constraints.

When real images are replaced with generated ones for training deep learning methods, it is important to validate the image generator. The problem of defining the accuracy of generation is not trivial. On one hand, real X-ray images of the same object acquired with the same system are never the same due to noise. Thus, images cannot be compared pixel by pixel, a metric of similarity is required. On the other hand, there is no prior knowledge of what level of accuracy (the value of the similarity metric) is sufficient for a particular application, and there have been successful applications of approximate image generation methods to practical problems. In this thesis, an application-specific definition of similarity is proposed and used. To compare two sets of images, multiple instances of the same deep learning method are trained on these datasets and applied to the same test dataset. If the performance on the test data is similar (statistical analysis is used to quantitatively characterize similarity), the images are considered similar in the context of a particular task and image analysis method.

Problem formulation and summary of chapters

This thesis proposes X-ray image generation as a method to improve image processing and design X-ray inspection systems. The chapters of this thesis focus on the different challenges of designing an image generator that can accurately and quickly produce a large number of images. The necessary information is extracted from the data without relying on rarely available prior knowledge.

Chapter 2 introduces the problem of foreign object detection using the example of detecting bone fragments in chicken fillets. While human observers see large bones on X-ray images with relative ease, automated detection is a challenging computer vision problem. It is necessary to distinguish bone fragments from thickness gradients of the chicken fillet, despite not knowing the shapes in advance. In this chapter, this problem is solved by applying a more sophisticated image processing approach. It is shown that with dual-energy X-ray imaging, bone fragments can be separated based on their material properties rather than their morphological properties. An image pre-processing method is proposed to effectively "subtract" chicken fillets from the images. As a result, foreign objects can be detected by classical image analysis algorithms without extensive adaptation to a particular object morphology. Their performance is compared with deep learning methods. While deep learning image analysis achieves higher accuracy, it requires significantly more prior knowledge in the form of training data.

Chapter 3 explores the possibility of training deep learning methods solely on generated images. The approach is tested on two problems: inspecting objects made of modeling clay and detecting signs of ripeness in avocados. The results

show that a deep learning method could achieve high accuracy on real test images without seeing a single real image during training. At the same time, the chosen generation method uses many approximations. Many experimental effects present during a real acquisition are not simulated, the internal structure of the inspection object is modeled with several simplifications. For practical applications, Chapter 3 proposes a method of creating large datasets of images based on a single real object. The method uses a CT scan of a real avocado to create "artificial" avocados. While "artificial" fruits are much more similar to each other than real ones, they provide enough variety of generated X-ray images to train an accurate deep learning method.

Chapter 4 explores why the simplified model of X-ray imaging used in the previous chapter was sufficient for training deep learning methods. The focus of this chapter is X-ray scattering. This effect is present in real X-ray images but was not simulated in generated images. However, the results indicate that the presence of scattering is not necessary to train a neural network. In this chapter, the influence of scattering is studied using highly accurate software for X-ray simulation. With this software, it is possible to remove scattering from the list of interactions between X-ray radiation and matter. The influence of missing scattering is then studied on a model problem of detecting cavities in cylinders made of different materials. The results show that excluding scattering has a small effect on detectability. Furthermore, the effect is only significant for small cavities that are barely detectable, since large cavities are detected regardless of simulating scattering in training data. Thus, the importance of scattering in the generator is characterized by a single number - the size of the smallest detectable foreign object. This methodology can be adapted to other more practical tasks to determine whether scattering should be simulated.

Chapter 5 presents a method for optimizing the design of the X-ray inspection system using image generation. The focus of this chapter is on making the generator parameterizable so that it is possible to generate images under different imaging settings. This approach is applied to the same problem as in Chapter 2 - detecting bone fragments in chicken fillets. However, the goal of Chapter 5 is to study the accuracy of inspection as a function of the image acquisition time. In the factory environment, inspection is performed on a conveyor belt, and slow imaging slows the flow of products. Therefore, it is crucial to know how fast the inspection can be while maintaining the desirable accuracy. A calibration procedure is proposed to connect the values of the X-ray exposure time and parameters of the image generator. By comparing real images acquired with different frame rates and generated images with the corresponding generation parameters, it is verified that the generator is sufficiently accurate in predicting the inspection performance. The same performance metric is used to show how the inspection gets worse when the exposure time decreases. With the methods developed in this chapter, industrial constraints of the inspection problem can be translated into ranges of generator parameters, and the parameters leading to the most suitable performance can be chosen as optimal.

SAMENVATTING

Het toepassen van röntgenbeeldgeneratie bij het ontwerpen van inspectiesystemen

Productinspectie is een belangrijk onderdeel van moderne industrieën. De landbouwindustrie heeft met name inspectiesystemen nodig om de kwaliteit hoog te houden bij het produceren van grote hoeveelheden voedsel. Een veelvoorkomende taak is de detectie van vreemde objecten - ongewenste voorwerpen die aanwezig kunnen zijn in producten en de kwaliteit ervan kunnen verminderen. Het automatiseren van deze detectie is een uitdagend probleem omdat voedingsmiddelen een natuurlijke variatie aan vormen en materiaalsamenstellingen hebben door de variatie in biologische processen. In tegenstelling tot gefabriceerde producten hebben voedingsmiddelen geen blauwdruk. Recente ontwikkelingen in computervisie en deep learning hebben geholpen om robuuste en nauwkeurige methoden te ontwerpen om vreemde objecten te detecteren. Deze methoden zijn vooral succesvol met digitale kleurenfoto's. In veel moderne fabrieken worden meerdere camera's op de transportband geïnstalleerd en analyseert gespecialiseerde software de foto's die door deze camera's worden genomen om vreemde objecten te detecteren. Digitale fotografie is echter beperkt tot inspectie van het oppervlak van voorwerpen waardoor er interesse is in andere technieken om vreemde objecten binnenin producten te detecteren.

Er is een toenemende interesse in de toepassing van röntgen(X-ray)-gebaseerde beeldvormende technieken voor voedselinspectie. Röntgenstraling wordt vaak gebruikt op luchthavens en ziekenhuizen om de interne structuur van objecten te inspecteren. Röntgenradiografie - het verkrijgen van een enkele röntgenfoto - is geschikt voor de inspectie van voedselproducten op de lopende band omdat het een snelle methode is met een lage stralingsdosis. Röntgensystemen worden al succesvol toegepast bij de inspectie van verpakt voedsel (detectie van plastic en metaalresten), vis en vlees (botfragmenten). Twee uitdagingen beperken het algemener toepassen van röntgenbeeldvorming. Ten eerste zijn veel voedingsmiddelen en ongewenste voorwerpen gemaakt van organische materialen en hebben ze een laag contrast op röntgenbeelden. Daarom moeten veel beeldvormingsinstellingen, zoals acquisitietijd, röntgenbronspanning en resolutie, geoptimaliseerd worden om beelden te produceren om de zichtbaarheid van het vreemde voorwerp te optimaliseren. Ten tweede zijn röntgenbeelden vaak moeilijk te analyseren, zelfs voor een expert (bijvoorbeeld een radioloog in een ziekenhuis). Röntgenstralen schijnen door het object heen en hierdoor leggen röntgenfoto's alle kenmerken van het object op verschillende diepten tegelijkertijd vast. Voedingsmiddelen hebben een complexe interne structuur die resulteert in vele grijstinten in de röntgenfoto's. Met deze structuur als achtergrond kunnen kleine ongewenste voorwerpen gemist worden.

Ondanks deze uitdagingen kunnen deep-learning-methoden een hoge nauwkeu-

righeid bereiken bij het detecteren van ongewenste voorwerpen op röntgenbeelden. Er zijn echter grote datasets van afbeeldingen nodig om deze methoden te trainen en dergelijke datasets zijn zeldzaam in industriële toepassingen. Er is een klein aantal openbaar beschikbare datasets met afbeeldingen van landbouwproducten. Deze datasets zijn meestal niet gespecialiseerd in een bepaald landbouwproduct om een verscheidenheid aan interne structuren te dekken (bijvoorbeeld afbeeldingen van verschillende appels van dezelfde soort). Bovendien zijn er geen standaardinstellingen voor röntgenbeeldvorming van landbouwproducten. Beschikbare beelden zijn vaak opgenomen met verschillende instellingen voor beeldvorming, en het vergelijken van beelden van hetzelfde object onder verschillende instellingen is zelden mogelijk. Als gevolg hiervan zijn de bestaande datasets vaak niet groot genoeg om nauwkeurige detectiemethoden te trainen en worden verschillende beeldvormingsinstellingen niet onderzocht om het ontwerp van het inspectiesysteem te optimaliseren.

In dit proefschrift wordt een voorstel beschreven om het gebrek aan echte gegevens voor deep learning-training en systeemontwerp te verhelpen doormiddel van het genereren van afbeeldingen. De beeldgenerator wordt gebruikt als een voorspellingsmethode (de algemene structuur wordt getoond in Fig. S2). Met behulp van informatie over het object en het beeldvormingssysteem, voorspelt het hoe een röntgenbeeld eruit zou zien zonder een experiment uit te voeren. Zo'n voorspelling kan zeer nauwkeurig zijn omdat de natuurkundige werking van röntgenstraling en beeldvorming goed bekend is. Het genereren van beelden lost geen fundamenteel probleem op van deep learning methoden, die grote hoeveelheden gegevens nodig hebben om nauwkeurig te zijn. In plaats daarvan is het doel om het generaren van beelden makelijker en sneller te maken dan het verzamelen van echte beelden.

Een belangrijke uitdaging bij het simuleren van beelden voor industriële taken is de hoeveelheid informatie die nodig is over het object en het systeem. Zeer nauwkeurige voorspellingsmethoden vereisen de chemische samenstelling van het object en de verdeling van verschillende chemische elementen binnen het object. Deze informatie is zelden beschikbaar voor landbouwproducten. De meeste voorspellingsmethoden hebben een beschrijving van de vorm van het object nodig. Hoewel het mogelijk is om een digitale volumebeschrijving te maken voor een enkel product, is dit niet realistisch voor de grote verscheidenheid aan objecten die vereist is voor het grootschalig genereren van afbeeldingen. Bovendien is een geavanceerd model van de röntgenbron en detector nodig om te voorspellen hoe het beeld wordt beïnvloed door veranderingen in de beeldinstellingen. In dit proefschrift wordt een alternatieve datagestuurde aanpak gebruikt om de benodigde informatie te verkrijgen zonder ingewikkelde en tijdsintensieve analyses. Kalibraties en metingen met een röntgensysteem worden gebruikt om de benodigde eigenschappen uit de gegevens te halen. Deze aanpak gaat uit van aannames en vereenvoudigingen die de resulterende nauwkeurigheid verminderen, maar is gemakkelijker te implementeren met de beperkingen die een industriële toepassing met zich mee brengt.

Wanneer echte afbeeldingen worden vervangen door gegenereerde afbeeldingen voor het trainen van deep learning methoden, is het belangrijk om de beeldgenerator te valideren. Het bepalen van de nauwkeurigheid van het voorspellingsmodel
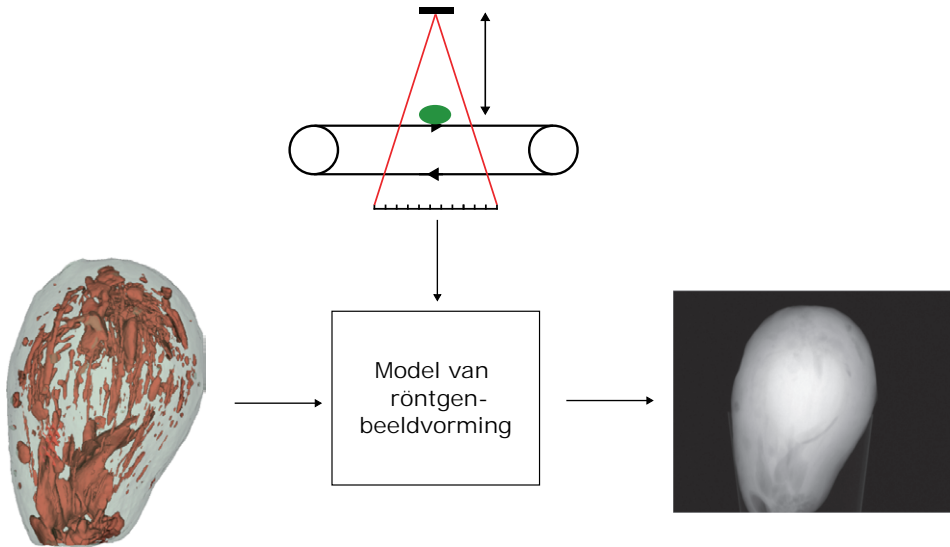
Figuur S2: De structuur van een beeldgenerator als voorspellingsmethode. Informatie over het object wordt ontvangen als invoer en het röntgenbeeld wordt berekent volgens het beeldvormingsmodel. Verschillende eigenschappen van het beeldvormingssysteem worden gerepresenteerd door de parameters van het model en beïnvloeden het gegenereerde beeld.

is niet triviaal. Aan de ene kant zijn echte röntgenbeelden van hetzelfde object, verkregen met hetzelfde systeem, nooit hetzelfde door ruis. Afbeeldingen kunnen dus niet pixel voor pixel met elkaar vergeleken worden, er is dus een maat nodig voor de vergelijkbaarheid van de beelden. Aan de andere kant is er geen voorkennis over welk niveau van nauwkeurigheid (de waarde van de vergelijkbaarheidsmaat) voldoende is voor een bepaalde toepassing en zijn er succesvolle toepassingen geweest van simulatiemethoden voor praktische problemen. In dit proefschrift wordt een toepassingsspecifieke definitie van vergelijkbaarheid voorgesteld en gebruikt. Om twee sets afbeeldingen te vergelijken, worden meerdere exemplaren van dezelfde deep learning methode getraind op deze datasets en toegepast op dezelfde testdataset. Als de prestaties op de testgegevens vergelijkbaar zijn (statistische analyse wordt gebruikt om de gelijkenis kwantitatief te karakteriseren), worden de afbeeldingen beschouwd als vergelijkbaar in de context van een bepaalde taak en beeldanalysemethode.

Probleemformulering en samenvatting van hoofdstukken

In dit proefschrift wordt röntgenbeeldgeneratie voorgesteld als een methode om beeldanalyse te verbeteren en röntgengebaseerde inspectiesystemen te ontwerpen.

De hoofdstukken van dit proefschrift richten zich op de verschillende uitdagingen van het ontwerp van een beeldgenerator die nauwkeurig en snel een groot aantal beelden kan produceren. De benodigde informatie wordt uit data afgeleid zonder dat voorkennis vereist is, aangezien deze zelden beschikbaar is.

In **Hoofdstuk 2** wordt het probleem van het detecteren van vreemde voorwerpen geïntroduceerd met een voorbeeld van de detectie van botfragmenten in kipfilets. Hoewel experts gemakkelijk grote botten zien op röntgenfoto's, is geautomatiseerde detectie geen triviaal computervisieprobleem. Het is noodzakelijk om botfragmenten te onderscheiden van de kipfilet, ondanks het feit dat de vormen niet van tevoren bekend zijn. In dit hoofdstuk wordt dit probleem opgelost door meer geavanceerde beeldverwerking toe te passen. Er wordt aangetoond dat door twee verschillende spanningen van de rontgenbron te gebruiken, botfragmenten kunnen worden gescheiden op basis van hun materiaaleigenschappen in plaats van hun morfologische eigenschappen. Er wordt een pre-processingtechniek voorgesteld om effectief kipfillet uit de beelden te halen. Hierdoor kunnen vreemde objecten worden gedetecteerd door klassieke beeldanalysealgoritmen, zonder uitgebreide aanpassing aan een bepaalde objectmorfologie. Hun prestaties worden vergeleken met deep learning methoden. Hoewel deep learning beeldanalyse een hogere nauwkeurigheid bereikt, vereist het aanzienlijk meer voorkennis in de vorm van trainingdata.

In **Hoofdstuk 3** wordt de mogelijkheid onderzocht om deep learning methoden te trainen op gegenereerde röntgenbeelden. De aanpak wordt getest op twee problemen: het inspecteren van objecten gemaakt van modelleerklei en (on)rijpe avocado's. De resultaten tonen aan dat een deep learning methode een hoge nauwkeurigheid kan bereiken op echte testafbeeldingen zonder ook maar één echt beeld te zien tijdens de training. Tegelijkertijd gebruikt de gekozen voorspellingsmodel veel benaderingen. Veel experimentele effecten die aanwezig zijn tijdens een echte acquisitie worden niet gesimuleerd, de interne structuur van het inspectieobject wordt gemodelleerd met verschillende vereenvoudigingen. Voor praktische toepassingen wordt in hoofdstuk 3 een methode voorgesteld om grote datasets van beelden te maken op basis van een enkel echt object. De methode gebruikt een CT-scan van een echte avocado om "kunstmatige" avocado's te maken. Hoewel "kunstmatige" vruchten veel meer op elkaar lijken dan echte, bieden ze voldoende variatie in gegenereerde röntgenbeelden om een nauwkeurige deep learning methode te trainen.

In **Hoofdstuk 4** wordt onderzocht waarom het vereenvoudigde model van röntgenbeeldvorming dat in het vorige hoofdstuk werd gebruikt, voldoende was voor het trainen van deep learning methoden. De focus van dit hoofdstuk is de verstrooiing van röntgenstralen. Dit effect is aanwezig in echte röntgenbeelden, maar werd niet gesimuleerd in de gegenereerde afbeeldingen. De resultaten geven echter aan dat de aanwezigheid van verstrooiing niet noodzakelijk is om een neuraal netwerk te trainen. In dit hoofdstuk wordt de invloed van verstrooiing bestudeerd met behulp van zeer nauwkeurige software voor röntgensimulatie. Met deze software is het mogelijk om verstrooiing te verwijderen uit de lijst van interacties tussen röntgenstraling en materie. De invloed van het weglaten van verstrooiing wordt

vervolgens bestudeerd op een modelprobleem: het detecteren van holtes in cilinders van verschillende materialen. De resultaten laten zien dat het weglaten van verstrooiing een klein effect heeft op de detecteerbaarheid. Bovendien is het effect alleen belangrijk voor kleine holtes die nauwelijks detecteerbaar zijn, aangezien grote holtes gedetecteerd worden ongeacht of verstrooiing gesimuleerd wordt in de trainingdata. Het belang van verstrooiing in de generator wordt dus gekarakteriseerd door een enkel getal - de grootte van het kleinste detecteerbare vreemde object. Deze methodologie kan worden aangepast aan andere, meer praktische taken om te bepalen of verstrooiing moet worden gesimuleerd.

Hoofdstuk 5 presenteert een methode om het röntgeninspectiesysteem te optimalizeren met behulp van beeldgeneratie. De focus van dit hoofdstuk is het maken van een parametriseerbare generator, zodat het mogelijk is om beelden te genereren onder verschillende beeldvormingsinstellingen. Deze aanpak wordt toegepast op hetzelfde probleem als in hoofdstuk 2 - de detectie van botfragmenten in kipfilets. Het doel van hoofdstuk 5 is echter om de nauwkeurigheid van de inspectie te bestuderen als functie van de acquisitietijd. In de fabrieks wordt de inspectie uitgevoerd op een lopende band en langzame beeldvorming vertraagt de productstroom. Daarom is het belangrijk om te weten hoe snel de inspectie kan worden gedaan, terwijl de gewenste nauwkeurigheid wordt behouden. Er wordt een kalibratieprocedure voorgesteld om de waarden van de belichtingstijd en parameters van de beeldgenerator met elkaar te verbinden. Door echte beelden te vergelijken die zijn verkregen met verschillende belichtingstijden en gegenereerde beelden met de corresponderende generatieparameters, is er geverifieerd dat de generator voldoende nauwkeurig is in het voorspellen van de inspectieprestaties. Dezelfde maat wordt gebruikt om te laten zien hoe de inspectie slechter wordt als de belichtingstijd afneemt. Met de methoden die in dit hoofdstuk zijn ontwikkeld, kunnen industriële beperkingen van het inspectieprobleem worden vertaald naar beperkingen voor generatorparameters en kunnen de parameters die leiden tot de meest geschikte prestaties worden gekozen.

CURRICULUM VITAE

Vladyslav Andriiashen was born in 1996 in Odesa, Ukraine. He completed his secondary education in 2013 at Richelieu Lyceum in Odesa, Ukraine (high school diploma with honors). He obtained his bachelor's degree in applied physics and mathematics (with honors) in 2017 and his master's degree in applied physics and mathematics (with honors) in 2019 from Moscow Institute of Physics and Technology. The master's thesis with the title "Development of the iterative reconstruction method for multi-energy CT" was supervised by dr. Alexey Zhemchugov. In 2019, he started as a PhD candidate under the supervision of prof.dr. Robert van Liere at Centrum Wiskunde & Informatica, the national research institute for mathematics and computer science in Amsterdam, the Netherlands.

ACKNOWLEDGMENTS

This dissertation is the result of many years of work. Challenges are not an uncommon part of scientific research. However, many of them have been amplified by the COVID-19 pandemic and War. In this section, I would like to acknowledge the support of many people who helped me during my PhD research.

I would like to thank my promotor Robert van Liere. Your energy and enthusiasm inspired many ideas that later became part of the publications. I am grateful for your support and your ability to find the right words in various challenging situations. I appreciate that you gave me a lot of freedom and independence, even when it lead to many disagreements.

I would like to thank Joost Batenburg for giving me the opportunity to work as a PhD student in the Computational Imaging group. This was a significant change in my life, which I value greatly. Conversations with you have always been a great source of insight into long-term project planning and science in general.

I would like to thank Tristan van Leeuwen. Your feedback and ideas have improved different aspects of my publications and taught me a lot. I appreciate your effort as a group leader in maintaining a great atmosphere in the group full of fruitful collaborations and interesting exchanges of ideas.

I thank the members of the Computational Imaging group. Certain aspects of this field were new to me and I had to learn a lot. I would like to thank Alexander Kostenko – for your help and guidance during my visit to CWI in 2018, even before I started my PhD. Sophia Coban – for teaching me how to operate the FleX-ray laboratory and for your great help at the beginning of my PhD. Willem Jan Palenstijn and Alexander Skorikov – for helping me with various computer science and software development questions. Allard Hendriksen and Richard Schoonhoven – for explaining many things about deep learning and how to implement it in the best way. Mathé Zeegers – for the opportunity to discuss many things with another PhD student working on a similar project during the first years of my PhD.

Thanks to the colleagues who shared an office with me – Rien Lagerwerf, Poulami Ganguly, Adriaan Graas, Hamid Fathi, Richard Schoonhoven, Francien Bossema, and Ezgi Demircan Türeyen. I hope that I have distracted you from work to about the same extent as you have distracted me. I am grateful to Sophia, Poulami, and Maureen for your effort and insights during our CodeSprint collaboration. I would like to thank Dzemila, Georgios, Daan, Felix, Rob, Roozbeh, Ajinkya, Jordi, Dirk, Maximilian, Tianyuan, Marcos, Xinyuan for contributing to the great research environment, providing interesting points of view, and collaborating in sometimes unpredictable ways.

The work required to create this dissertation would hardly be possible without many people working at CWI. In particular, I would like to thank Minnie Middelberg for great many things you had to do to maintain the insitute a building where people feel welcome. Remco Westra and Ramona Rijff – for your help with various technical problems. Nada Mitrovic – for making sure different events and