Expert Systems With Applications 206 (2022) 117768

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A tomographic workflow to enable deep learning for X-ray based foreign object detection

Mathé T. Zeegers [a,*], Tristan van Leeuwen [a,b], Daniël M. Pelt [a,c], Sophia Bethany Coban [a,d], Robert van Liere [a,e], Kees Joost Batenburg [a,c]

[a] Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands
[b] Mathematical Institute, Utrecht University, Budapestlaan 6, 3584 CD Utrecht, The Netherlands
[c] Leiden Institute of Advanced Computer Science, Niels Bohrweg 1, 2333 CA Leiden, The Netherlands
[d] Department of Mathematics, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom
[e] Faculteit Wiskunde en Informatica, Technical University Eindhoven, Groene Loper 5, 5612 AZ Eindhoven, The Netherlands

ARTICLE INFO

Keywords:
X-ray imaging
Foreign object detection
Segmentation
Computed tomography
Machine learning
Deep learning

ABSTRACT

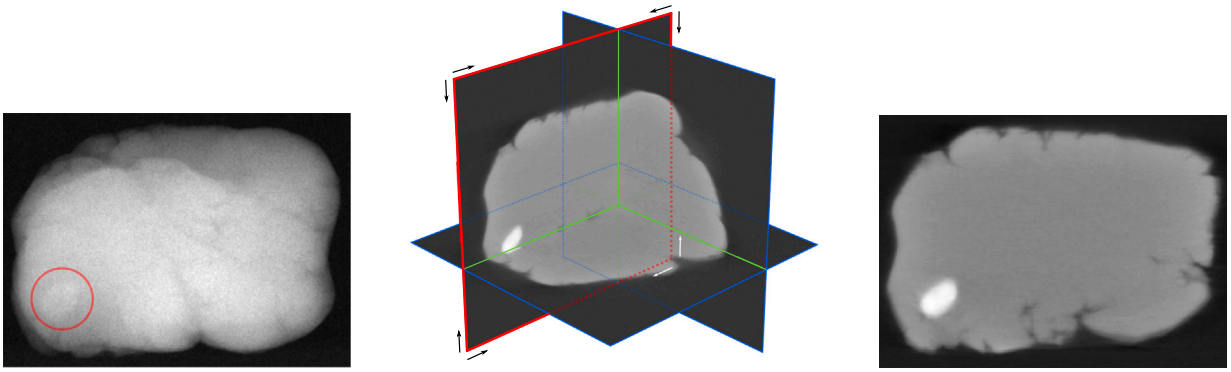
Detection of unwanted ('foreign') objects within products is a common procedure in many branches of industry for maintaining production quality. X-ray imaging is a fast, non-invasive and widely applicable method for foreign object detection. Deep learning has recently emerged as a powerful approach for recognizing patterns in radiographs (i.e., X-ray images), enabling automated X-ray based foreign object detection. However, these methods require a large number of training examples and manual annotation of these examples is a subjective and laborious task. In this work, we propose a Computed Tomography (CT) based method for producing training data for supervised learning of foreign object detection, with minimal labor requirements. In our approach, a few representative objects are CT scanned and reconstructed in 3D. The radiographs that are acquired as part of the CT-scan data serve as input for the machine learning method. High-quality ground truth locations of the foreign objects are obtained through accurate 3D reconstructions and segmentations. Using these segmented volumes, corresponding 2D segmentations are obtained by creating virtual projections. We outline the benefits of objectively and reproducibly generating training data in this way. In addition, we show how the accuracy depends on the number of objects used for the CT reconstructions. The results show that in this workflow generally only a relatively small number of representative objects (i.e., fewer than 10) are needed to achieve adequate detection performance in an industrial setting.

1. Introduction

Foreign object detection in an industrial high-throughput setting is essential for guaranteeing quality and safety of objects processed in factory lines. Foreign objects may, for example, appear in products such as meat, fish or vegetables as small pieces of glass, bones, plastic, wood or stone that could harm consumers (Andriiashen, van Liere, van Leeuwen, & Batenburg, 2021; Wilm, 2012; Zhu, Spachos, Pensini, & Plataniotis, 2021). Conventional nondestructive methods for detecting foreign objects include ultrasound imaging, X-ray imaging, magnetic resonance imaging, fluorescence imaging, (hyperspectral) spectroscopic imaging and thermal imaging (He et al., 2021; Li et al., 2019; Mohd Khairi, Ibrahim, Md Yunus, & Faramarzi, 2018; Narsaiah, Biswas, & Mandal, 2020; Nicolaï et al., 2014; Xiong, Sun, Pu, Gao, & Dai, 2017). X-ray imaging provides the unique opportunity to visualize the interior structure of an object in a fast, low-cost, and non-invasive manner. This enables *X-ray based foreign object detection*, in which the goal is to detect unwanted smaller objects inside base objects based on their distinct attenuation or attenuation patterns, as observed in generated *radiographs* (i.e. standard 2D X-ray images). The possibility to reveal hidden foreign objects in radiographs has lead to its extensive use in various industrial applications (Einarsdóttir et al., 2016; Haff & Toyofuku, 2008; Kwon, Lee, & Kim, 2008; Mathanker, Weckler, & Bowser, 2013; Mery et al., 2011; Narsaiah et al., 2020; Zhong, Zhang, Lu, Liu, & Wang, 2019), for which low-cost, adaptive and efficient image processing methods are essential (Mathanker et al., 2013; Xiong et al., 2017). One way to achieve better discrimination of foreign objects in radiographs is to use multispectral X-ray imaging detectors, simultaneously capturing radiographs at two or more energy levels (Si-Mohamed et al., 2017; Taguchi, Blevis, & Iniewski, 2020). As the

* Correspondence to: Science Park 123, 1098 XG Amsterdam, The Netherlands.
E-mail addresses: m.t.zeegers@cwi.nl (M.T. Zeegers), t.van.leeuwen@cwi.nl (T. van Leeuwen), d.m.pelt@liacs.leidenuniv.nl (D.M. Pelt), sophia.coban@manchester.ac.uk (S.B. Coban), robert.van.liere@cwi.nl (R. van Liere), k.j.batenburg@liacs.leidenuniv.nl (K.J. Batenburg).

https://doi.org/10.1016/j.eswa.2022.117768
Received 26 January 2022; Received in revised form 15 May 2022; Accepted 3 June 2022
Available online 11 June 2022



(a) 2D radiograph (b) 2D slices in the reconstructed 3D volume space (c) Red-bordered 2D slice of the 3D reconstruction

Fig. 1. Different views of an imaged product (Play-Doh) with a foreign object inserted (a piece of gravel). A 2D radiograph with the location of the foreign object (red circle) is shown (**a**), as well as multiple slices through the 3D volume of the reconstructed object (**b**), of which the slice with the red border is highlighted (**c**). The images show the difference in contrast: the foreign object is much easier to distinguish based on intensity values in the reconstructed 3D volume (**b** and **c**) than in the 2D radiograph (**a**).

attenuation properties of each material have their own characteristic dependence on the X-ray energy, these multispectral images can be analyzed to extract material composition information.

However, superposition of materials gives rise to similar levels of intensities for different objects in 2D radiographs. This problem limits the application of commonly used segmentation methods, such as threshold-based, clustering-based, and boundary-based or edge-based segmentation (Sezgin & Sankur, 2004; Silva, Oliveira, & Pithon, 2018), to extract different components of the object. Additionally, high-throughput acquisition may lead to high noise levels in radiographs, and this increases the difficulty of successful foreign object detection even further (Mathanker et al., 2013; Xiong et al., 2017). Commonly used segmentation methods can be unsuitable in case of poor image qualities caused by conditions such as noise, low contrast and homogeneity in regions close to foreign objects (Silva et al., 2018). Most conventional unsupervised methods can therefore not achieve high accuracies (Silva et al., 2018) without extensive manual parameter tuning to use a method for a specific problem (Al-Sarayreh, Reis, Yan, & Klette, 2019; Rong, Xie, & Ying, 2019).

Machine learning is a powerful tool for recognizing patterns in images (Zhao, Zheng, Xu, & Wu, 2019) and can potentially detect foreign objects in radiographs (Zhu et al., 2021). Recent machine learning methods address a wide variety of segmentation problems (Garcia-Garcia, Orts-Escolano, Oprea, Villena-Martinez, & Garcia-Rodriguez, 2017; Silva et al., 2018), and provide a remarkable improvement over more classical segmentation methods in many practical applications (Guo, Liu, Georgiou, & Lew, 2018). A key obstacle in the application of machine learning is the need for large datasets (Chartrand et al., 2017; Deshpande, Minai, & Kumar, 2020; Wu, Liu, & Liu, 2019), which is particularly prominent in machine learning for foreign object detection as each new combination of sample, foreign object, and imaging settings requires additional data. On top of that, supervised learning uses labeled datasets for training. However, manual annotation (as in e.g. Al-Sarayreh et al., 2019; Silva et al., 2018) requires tremendous efforts (Akcay & Breckon, 2022), is time consuming and tedious (Tajbakhsh et al., 2020), is subjective and can be prone to errors.

The key contribution of this paper is to propose a workflow based on 3D Computed Tomography (CT) for efficiently creating large training datasets, overcoming the aforementioned obstacle. CT scans of a relatively small number of objects are carried out with low exposure time – as in a high-throughput setting – yielding a large number of radiographs that are used as input for the supervised machine learning method. The same set of radiographs is also used offline for generating multiple high-quality tomographic 3D reconstructions, from which foreign objects can easily be segmented in 3D and projected back onto a virtual 2D detector to give the corresponding ground truth locations of the foreign objects in the radiographs. Without the effort of extensive manual labeling, this results in a large dataset with which deep learning can be carried out to detect foreign objects from fast-acquisition radiographs at a high rate. The example in Fig. 1 illustrates the difference in ease of segmentation for a CT reconstructed 3D volume versus a 2D radiograph. Whereas segmenting the foreign object in a radiograph is a challenging task, simple global thresholding can be applied to the CT volume to separate the foreign object from the base object. Additionally, more sophisticated and accurate segmentation and denoising rules can be imposed on 3D volumes (Garcia-Garcia et al., 2017; Pan, Zhou, Zhu, & Zheng, 2018; Van De Looverbosch, Raeymaekers, Verboven, Sijbers, & Nicolaï, 2021) than on 2D radiographs.

The structure of the paper is as follows. Section 2 provides the background of applying machine learning for foreign object detection, and Section 3 explains the proposed method of data generation to apply machine learning. In Section 4, the workflow is demonstrated in a laboratory experiment, and shows how the number of imaged objects affects the detection accuracies. Additionally, the robustness of the workflow is analyzed. Section 5 discusses various aspects of the results and the flexibility and modularity of the workflow. Section 6 presents the conclusions from this work.

2. Preliminaries

In this section, we introduce the X-ray foreign object detection problem and the machine learning concepts used in this paper.

2.1. Foreign object detection with X-ray imaging

We consider the problem of foreign object detection in an industrial high-throughput conveyor belt setting. The problem and the usage of X-ray imaging to solve this are schematically shown in Fig. 2. In foreign object detection, the aim is to correctly determine for each object whether a foreign object is contained in it or not, for instance a piece of bone within a meat sample.

For this problem, we focus on finding an accurate *segmentation* for each radiograph. A segmentation partitions an image into sets of pixels with the same label. In our case, the formed segmented image is binary and indicates on which detector pixels a foreign object is projected. The segmentation depends on the type of objects that are considered to be foreign (by for instance a manufacturer). Any further classification (based on the minimum size of a foreign object for example) can be carried out after the segmented image is produced.

Throughout this paper, we use the term *radiograph* for radiographs corrected using flatfield radiographs (without an object) and darkfield

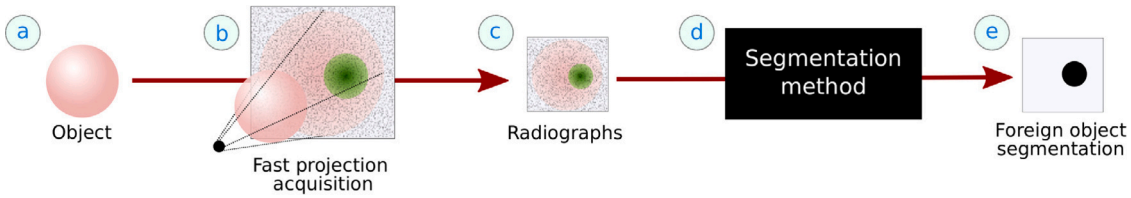


Fig. 2. A schematic overview of the foreign object detection problem and the segmentation-based approach to solving it. Each object (**a**) is assumed to have a correct segmentation (**e**). By using X-ray imaging (**b**), a radiograph of the object (**c**) can be acquired. Using a segmentation method (**d**), a segmented image (**e**) can be produced. The main challenge is to find a suitable segmentation method such that this approach to foreign object detection produces the correct results.

images (without the X-ray beam) that serve as input to the segmentation method. The quality of a radiograph depends on a number of properties of the scan, including exposure times, tube intensities, photon energy windows and the geometric setup (Russo, 2017). In a high-throughput setting, the steps in Fig. 2 should be fast to carry out, typically resulting in high noise levels and a challenging segmentation task.

2.2. Supervised learning

Machine learning is a widely used approach for difficult imaging tasks, as it can extract complicated patterns from complex images. In the foreign object detection problem, supervised machine learning can be used to learn the segmentation task such that it generalizes well for all possible fast-acquisition radiographs of similar objects with similar acquisition settings. To do so, a set of examples $\{(x_i, y_i)\}_{i=1}^N$ is used, where $\{x_i\}_{i=1}^N$ are acquired radiographs and $\{y_i\}_{i=1}^N$ are their corresponding foreign objects segmentations. The aim is to find the unknown segmentation function F that maps each radiograph x_i to its segmentation y_i. To find an approximate solution that generalizes well, the set of images is partitioned into a training set, a validation set and a test set. The training set is used to learn the function F_{train} that minimizes the loss L on the training set, which is the sum of errors between the segmented images $F_{\text{train}}(x_i)$ produced by the segmentation function and the true segmented images y_i. To find a suitable segmentation function, a (convolutional) neural network is often used as a model and parametrized using weights and biases that are optimized during the training process. While carrying out the training with a chosen loss function and optimization algorithm, the performance of the model is evaluated on the validation set. Several stopping criteria can be used for this, for example stopping the training when the error on the validation set increases, or training for a fixed time (and recording the network that gives the best results on the validation set). To avoid any bias towards the training and validation data, the accuracy of the trained model is finally assessed using the test set.

Since the introduction of Fully Convolutional Networks (Long, Shelhamer, & Darrell, 2015), in which successive contracting convolutional layers are utilized for pixel-wise semantic segmentation, many convolutional neural network (CNN) architectures have been proposed that can be used for the object segmentation task. U-Net changes the FCN architecture by – along with downsampling operators and skip connections – introducing upsampling operators instead of pooling operators, giving it an U-shaped appearance (Ronneberger, Fischer, & Brox, 2015). Similarly, Deconvnet (Noh, Hong, & Han, 2015) also introduces an auto-encoder structure with deconvolution and unpooling operations (without skip connections). The success of these methods on medical image segmentation and object detection spawned other commonly used CNN architectures for segmentation such as SegNet (Badrinarayanan, Kendall, & Cipolla, 2017), RefineNet (Lin, Milan, Shen, & Reid, 2017), PSPNet (Zhao, Shi, Qi, Wang, & Jia, 2017), and Mask R-CNN (He, Gkioxari, Dollár, & Girshick, 2017) for instance segmentation. Although some of the listed architectures need relatively few training examples for successful segmentation, the annotation of these examples still requires considerable efforts.

3. Proposed method for training data acquisition

When attempting to perform machine learning for X-ray based foreign object detection, the major obstacle is to acquire (manually) annotated training data. In this section, we explain the methodology of our proposed CT-based workflow for efficiently creating this annotated training data.

Our proposed workflow for using CT to obtain annotated training images is schematically displayed in Fig. 3. First, we select a set of representative objects as training objects (Fig. 3a). For each object, a set of fast-acquisition radiographs is collected from a set of predefined angles (Fig. 3b). These fast-acquisition radiographs will form the input set of the intended training dataset (Fig. 3c). The total number of examples in the resulting dataset is the number of training objects multiplied by the number of selected angles.

The same set of radiographs is used to carry out a tomographic reconstruction of the object and acquire high-quality CT volumetric data (Fig. 3d and e). The next step is to segment the reconstructed volume such that a possible foreign object is separated from the base object (Fig. 3f). This segmentation step can be automated and many methods are available to implement this (Lenchik et al., 2019). Here, we consider volumetric segmentation methods that consist of a global thresholding step. Binary segmentation by global thresholding is defined by the following function $S : \mathbb{R} \to \{0, 1\}$ that acts on every voxel z_{ijk} in reconstruction volume z:

$$S(z_{ijk}) = \begin{cases} 1 & z_{ijk} \geq \theta, \\ 0 & z_{ijk} < \theta. \end{cases}$$

Here, θ is the segmentation threshold. The more angles and other high-quality settings are used to obtain projection data, the easier it is to accurately segment the foreign object. Easier segmentation can also be accomplished by carrying out a separate high-quality scan of the same object and making a reconstruction with these high-quality radiographs. Additionally, for segmentation, prior information about the objects can be used, such as bounding boxes on the foreign object location (Kern & Mastmeyer, 2021). Also, 3D denoising (Diwakar & Kumar, 2018; Hendriksen, Pelt, & Batenburg, 2020) can be used to remove non-foreign object pixels captured by the thresholding operation.

From the constructed foreign object segmentation, virtual ground truth projections are generated by simulating projections of the foreign objects onto a virtual detector (Fig. 3g). This results in the set of ground truth images, which will serve as target images in the machine learning procedure (Fig. 3h). These virtual projections need to be taken under the same angles as in the fast-acquisition scan (Fig. 3b). When this procedure is repeated for all objects, this results in a large dataset with annotated training examples with which supervised machine learning can be carried out (Fig. 3c and f). The trained model can then be applied to similar new objects scanned in the same fast-acquisition setting, without the need for acquisition of high-quality radiographs or CT scans.

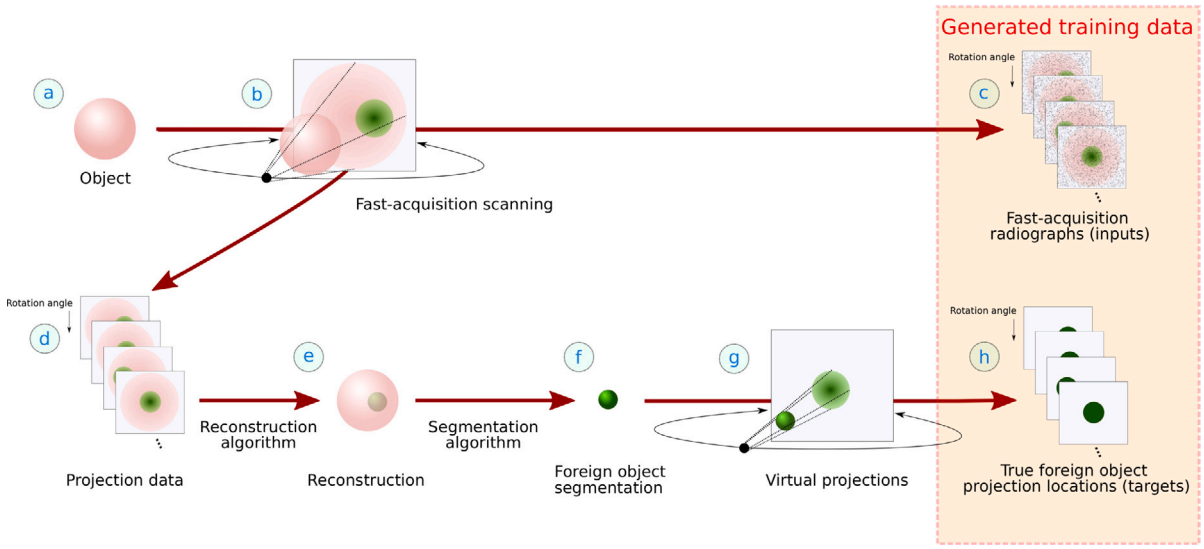


Fig. 3. The complete workflow of data acquisition (**a**,**b**) and the generation of training data (**c**,**h**) for deep learning driven foreign object detection, through 3D reconstruction from the CT scan (**d**, **e**), segmentation (**f**), and virtual projections (**g**). The reconstruction reveals the hidden foreign objects inside the main object. Note that the projection data (**d**) is usually just the set of fast-acquisition radiographs (**d**).

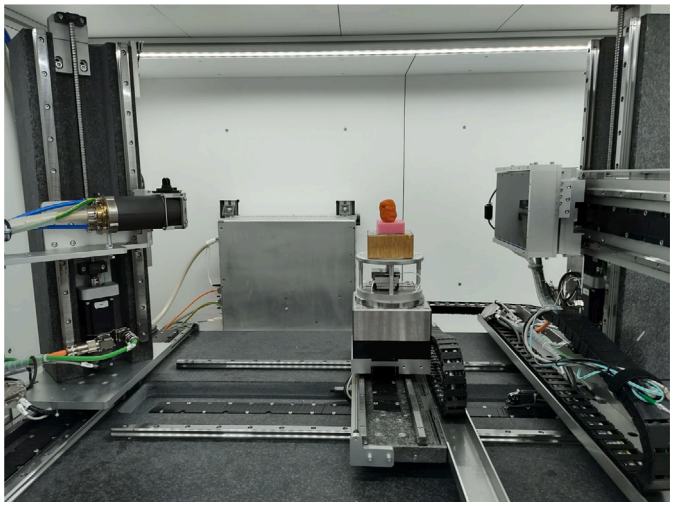


Fig. 4. The scanning setup in the FleX-ray laboratory with the X-ray source on the left and the detector on the right.

4. Experiments and results

In this section, we demonstrate the proposed workflow using the in-house FleX-ray CT system at CWI (Coban, Lucka, Palenstijn, Van Loo, & Batenburg, 2020) (Fig. 4), and investigate the relation between machine learning performance and the number of training objects used.

4.1. Base objects and foreign objects

As test objects, we use base objects that are created from a fixed amount of modeling clay (Play-Doh, Hasbro, RI, USA). Play-Doh is primarily made of a mixture of water, salt and flour and we therefore consider it to be a representative example of products in the food industry, where foreign objects may be pieces of stone, plastic, or metal. A basic shape is deformed and remolded for every object instance (Fig. 5(a)) in such a way that they are similar from object to object, but still exhibit some natural variation. For the foreign objects choose to use gravel (Fig. 5(b)), with the stones having an average diameter of ca. 7 mm (ranging from 3 mm to 11 mm). These stones have slight variations in shape and material. We create 3 objects with three inserted stones, 35 with two stones, 62 with one stone (Fig. 5(c)) and 11 without a stone.

4.2. CT scanning and data preparation

A fast CT-scan is made for each of the objects, which yields both a series of radiographs (i.e. the X-ray projections) and a reconstructed 3D volume of the object. The objects are scanned in the FleX-ray laboratory (Coban et al., 2020) (Fig. 4). The FleX-ray CT-scanner has a cone-beam microfocus X-ray point source with a focal spot size of 17 μm, and a Dexela1512NDT detector. The source, object and detector positions can be configured flexibly, and are arranged such that the distance between the source and detector is 69.80 cm, and the distance between the source and the object 44.14 cm. For the radiographs a voltage of 90 kV with a power of 20 W is used, while the exposure time is kept low at 20 ms, with the intention to emulate the imaging conditions of in-line industrial systems and produce sufficiently noisy radiographs. To achieve high-quality reconstructions, 1800 projections of each object are obtained at equidistant angles over a full 360° rotation. All projection angles are precisely recorded during the scan for the later stages of the workflow. Before and after each scan, 10 darkfield images and 10 flatfield projections are obtained. Each object is positioned in a random manner, and the cylinders may therefore be standing upright or be laying down on the long edge. The generated CT data are made available at Zenodo (Zeegers, 2022a). Example radiographs are shown in Fig. 6. Separating the projected foreign objects from the base object in these radiographs is not a trivial task, illustrating the problem of obtaining annotated training data for automated segmentation using machine learning directly from these images.

The Simultaneous Iterative Reconstruction Technique (SIRT) (Kak, Slaney, & Wang, 2002; Van der Sluis & Van der Vorst, 1990) algorithm (100 iterations) as implemented in the ASTRA toolbox (Van Aarle et al., 2016, 2015) is used to compute the reconstructed 3D CT volume of the object. A visualization of the reconstruction from the third object in Fig. 6 and its foreign object is shown in Fig. 7. The CT reconstruction allows to slice the object along different axes. As the CT voxel intensity is directly related to the attenuation coefficient of the material in a voxel, the segmentation task for the 3D CT volume is, in this case, much more straightforward and can be carried out by global thresholding (see Appendix for additional details on intensity value distributions).

(a) Play-Doh (b) Stones (c) Play-Doh with stone inserted

Fig. 5. An example of a base object (**a**) and examples of foreign objects (**b**) used in the laboratory experiments, as well as an example of a combined object (**c**).

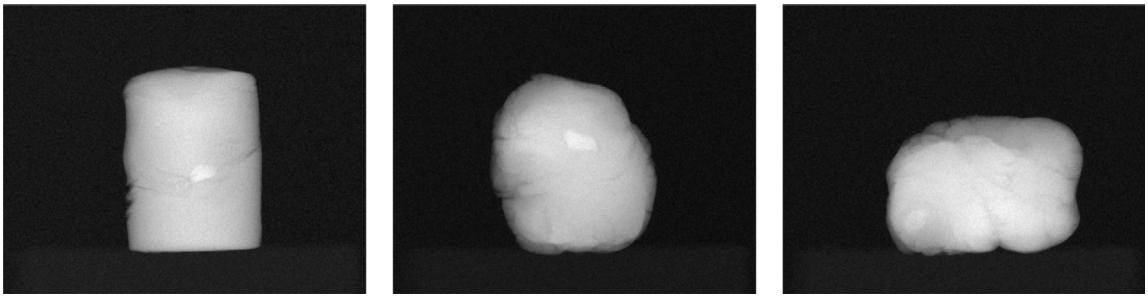


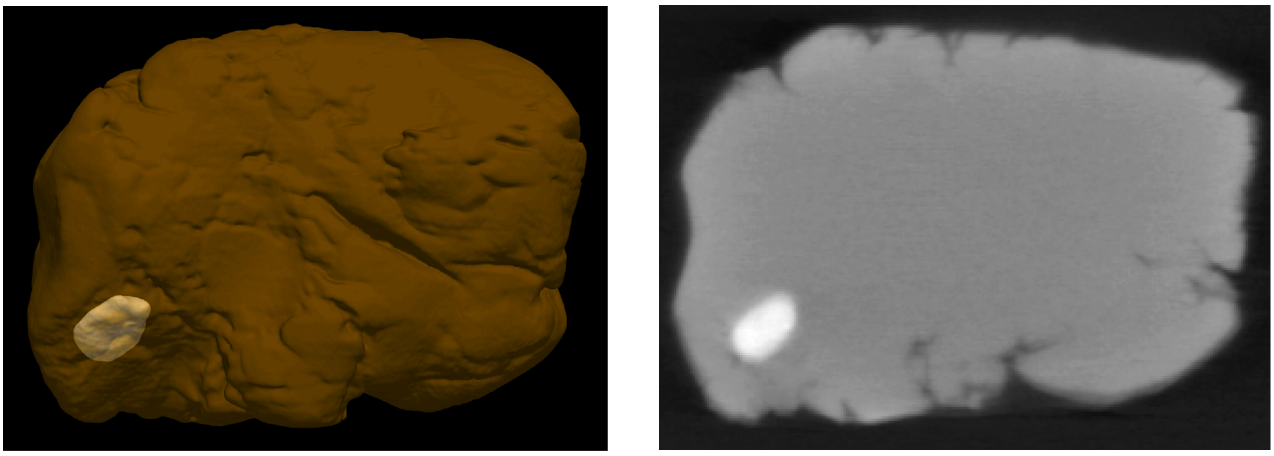
Fig. 6. Example of radiographs (size 965 × 760 pixels) of three objects that are scanned. In the first and second radiographs the foreign object is clearly visible, but in the third it is more difficult to distinguish it from the base object, even though it is visible on the bottom left.

Therefore, a simple global threshold based on Otsu's method (Otsu, 1979) is sufficient to segment the foreign objects.

From the 3D segmented objects obtained from the CT-scans, 2D segmentations for the individual radiographs are computed. This is done by computing the projections of the segmented parts with the ASTRA toolbox using the same geometric properties and recorded angles as in the radiograph acquisition of the actual CT-scan, to ensure geometric consistency in the training examples. Every non-zero pixel on the detector is marked as a projected foreign object location. The result is a dataset containing 1800 radiographs and corresponding segmented images for each object. This dataset is made available at Zenodo as well (Zeegers, 2022b).

4.3. Machine learning

We use the Mixed-Scale Dense (MSD) (Pelt & Sethian, 2018) and the common U-Net convolutional neural network architectures (Ronneberger et al., 2015) to train the task of image segmentation. For our experiments with U-Net, we have slightly changed the architecture, as we observed this improved performance in the experiments compared to the standard version. We downsample twice, with a stride of 2. The initial number of feature maps is set to 128, and the number of feature maps doubles for each downsampling layer. For upsampling, bilinear interpolation is used. A spatial 3 × 3 convolution operation with zero padding and a ReLU activation function are carried out before and after all downsampling and upsampling operations. The biases and convolution weights are initialized by sampling from $\mathcal{U}(-\sqrt{k}, \sqrt{k})$, with $k = 1/c_{\text{in}} \cdot a^2$ being the range, c_{in} the number of input channels and a the kernel size. ADAM optimization on the average of the binary cross entropy loss and the dice loss (Jadon, 2020; Sudre, Li, Vercauteren, Ourselin, & Cardoso, 2017) between the data and the predictions is used for training. The network is implemented with PyTorch (Paszke et al., 2017, 2019). For comparison between architectures, we also use the MSD network for training. MSD is a compact network architecture that has been demonstrated to be suitable for real-time segmentation of X-ray and CT images using relatively few training examples compared to larger networks (Pelt & Sethian, 2018), including the U-Net architecture. We use a depth of 100 intermediate layers and width of 1 channel per intermediate layer and increase the dilation parameter repeatedly from 1 to 10 dilations in each layer, which are common settings for the MSD network (Lagerwerf, Pelt, Palenstijn, & Batenburg, 2020; Pelt, Batenburg, & Sethian, 2018; Pelt & Sethian, 2018). Xavier initialization is used for the convolution weights. ADAM optimization (Kingma & Ba, 2015) is used during training on the cross-entropy loss between the ground truth and the segmented images, and the batch size of training examples is set to 10. We use the GPU implementations in Python that are available (Pelt, 2019; Pelt & Sethian, 2018). For both architectures, the learning rate is set to 0.001 and all networks are trained on a GeForce GTX TITAN X GPU with CUDA version 10.1.243. All hyperparameters of both network architectures are kept the same during all experiments. Data augmentation is applied by rotation and flipping of the input examples. All networks are trained for 9 h, and the



(a) Reconstructed 3D volume (b) Slice of the 3D reconstruction

Fig. 7. Visualization of reconstructions by 3D rendering (**a**) and slicing (**b**) of the third object in Fig. 6.

network with parameters resulting in the lowest error on the validation set is used for testing.

With these networks, we carry out an image-to-image training from radiographs (Fig. 3c) to their corresponding foreign objects segmentations (Fig. 3f). For training, 60 randomly chosen base objects containing a foreign object are used. The remaining 51 objects are used for testing. All images are resized using cubic interpolation to 128×128 to speed up the training process (global thresholding with parameter $\theta = 0.5$ is applied to the resized ground truth images to make these binary again). We test the performance of the trained networks for different numbers of objects included in the training scheme. To compare the workflow with labor-intensive 2D data annotation, we compare the following training strategies:

- **Workflow approach**: For each network, we fix the total number of training examples to 1800. A random but fixed order of the 60 training objects is created and the first i objects among these are used for the training set. The training examples are selected from the set of radiographs and ground truths created by the workflow from these i training objects in equal amounts. Every 10th example is used for validation during training.
- **Classical approach**: For each network, only one randomly chosen training radiograph with the corresponding ground truth (generated using the workflow) is selected for each of the first i included training objects. The resulting set of training examples is separated such that $^9/_{10}$ part is used for training (rounded down to the nearest integer) and $^1/_{10}$ part is used for validation (rounded up).

4.4. Quality measures

To evaluate the accuracy of the trained networks on the test set, in which the target images are generated using the workflow on the test objects. Three different measures on the segmented images and the corresponding target images are computed. The collection of these measures both assess the image segmentation accuracy and the object detection accuracy. An image segmentation accuracy is based on the classification of each pixel in the segmented image, and there are standardized ways to measure this that do not depend on any parameters (Grandini, Bagli, & Visani, 2020). An object detection accuracy compares connected components (groups of pixels connected by their edges) in the segmented image with the ground truth images. Although these accuracy measures require additional parameters to define the notion of detection, they are more relevant to the foreign object detection application.

The first measure is an *image-based average class accuracy* (also called *balanced accuracy* Grandini et al., 2020) to assess the accuracy of a produced segmentation. The average class accuracy of a segmented image relative to the target image is given by the sum of the true positives divided by the true positives and false negatives (the recall) of each class, averaged over the number of classes. In the binary case this becomes

$$\frac{1}{2}\left(\frac{\mathrm{TP}_{\mathrm{FO}}}{\mathrm{TP}_{\mathrm{FO}}+\mathrm{FN}_{\mathrm{FO}}}+\frac{\mathrm{TP}_{\mathrm{BG}}}{\mathrm{TP}_{\mathrm{BG}}+\mathrm{FN}_{\mathrm{BG}}}\right). \tag{1}$$

Here, $\mathrm{TP}_{\mathrm{FO}}, \mathrm{FN}_{\mathrm{FO}}, \mathrm{TP}_{\mathrm{BG}}$ and $\mathrm{FN}_{\mathrm{BG}}$ are the true positive and false negative rates of the foreign object and the combined base object and background pixel classifications respectively over the entire segmented image relative to the target image. The average class accuracy as given in (1) is averaged over all target images.

The second measure is an *object based detection rate*. A *connected component* is a maximal set of nonzero-valued pixels such that each pixel is reachable from another pixel in the set via a sequence of neighboring pixels in the set. Each connected component in the target image with a minimum size of 8 pixels (0.05% of the image size) is considered as an object that should be detected. We define such an object as *detected* if its pixel-wise recall relative to the segmented image is higher than a certain threshold η:

$$\frac{\mathrm{TP}_{\mathrm{obj}}^{\mathrm{tar}}}{\mathrm{TP}_{\mathrm{obj}}^{\mathrm{tar}}+\mathrm{FN}_{\mathrm{obj}}^{\mathrm{tar}}} > \eta. \tag{2}$$

Here, $\mathrm{TP}_{\mathrm{obj}}^{\mathrm{tar}}$ and $\mathrm{TP}_{\mathrm{obj}}^{\mathrm{tar}}$ are the true positive and false negative pixels in the target object relative to the segmented image. The threshold indicates the percentage of pixels of a projected foreign object in the target image that should be indicated as foreign object pixels in the segmented image produced by the network to be marked as a detected object. In our experiments, we set $\eta = 0.3$. We define the *detection rate* as the percentage of components in all target images for which condition (2) holds.

The third measure is an *object based false positive detection rate*. Each connected component in the segmented image with a minimum size of 8 pixels is considered as a potentially detected object. We define such a potentially detected object as a *false positive* if its pixel-wise recall relative to the target image is lower than a certain threshold δ:

$$\frac{\mathrm{TP}_{\mathrm{obj}}^{\mathrm{seg}}}{\mathrm{TP}_{\mathrm{obj}}^{\mathrm{seg}}+\mathrm{FN}_{\mathrm{obj}}^{\mathrm{seg}}} < \delta. \tag{3}$$

Here, $\mathrm{TP}_{\mathrm{obj}}^{\mathrm{seg}}$ and $\mathrm{TP}_{\mathrm{obj}}^{\mathrm{seg}}$ are the true positive and false negative pixels in the segmented object relative to the target image. The threshold indicates the percentage of pixels of a foreign object in the segmented

image produced by the network that are correctly labeled as foreign objects compared to the foreign object in the target image. In our experiments, we set $\delta = 0.3$. We define the *false positive detection rate* as the percentage of potential objects in all segmented images for which condition (3) holds.

4.5. Results

For the test set, we select a random angle and an orthogonal one for each test object, making the total number of testing radiographs 102. We measure the average class accuracy, the object-based detection rate and the object-based false positive detection rate of segmentations created by the network on the projections from the test set. The results are given in Fig. 8.

For all measures, the quality of the foreign object segmentations in the radiographs using networks trained with the workflow data is low for a few training objects. This initially improves with the addition of relatively few training objects, but this improvement stagnates beyond 20 objects. However, the detection accuracy still shows slight improvements beyond this point, but almost completely stabilizes from 40 objects onwards. Based on a decided accuracy goal, a certain number of objects need to be scanned and used for training to achieve that accuracy. The false positive rate decreases strongly and maintains a low level value from including 3 objects in the training onwards. Note that the results between the U-Net and MSD architectures agree well with each other.

When we compare the usage of a fixed number of training radiographs among all training objects with the classical approach of using only one radiograph per object, we see that this leads to inferior results in all aspects. The average class accuracies and the object based detection rates are lower for all numbers of included training objects, while the false positive rates are higher. The difference between architectures only shows for the false positive detection rate, which is generally higher with the U-Net architecture.

4.6. Laboratory experiments with many foreign objects

A natural way to reduce the number of objects used for training that need to be scanned for obtaining accurate segmentations may be to include more foreign object in the imaged objects. To test this, we repeat the experiments of the previous section, but we insert 5 to 8 foreign objects instead of 0 to 2. The foreign objects are placed within the base object such that overlapping of foreign objects in the radiographs is minimized. We have scanned an additional set of 20 objects with these characteristics. An example of a radiograph of an object with many foreign objects is shown in Fig. 9(a). We compare the following training strategies in which the workflow data comes from the following sets of training objects:

- **Few foreign objects**: Base objects with 0 to 2 foreign objects
- **Many foreign objects**: Base objects with 5 to 8 foreign objects
- **Mixed**: 50% – 50% mix of base objects with 0 to 2 foreign objects and base objects with 5 to 8 foreign objects.

All networks are evaluated on the testing set from the previous section (with test objects containing few foreign objects). The average class accuracies, detection accuracies and the false positive rates of the trained neural networks with these schemes on the test set are shown in Fig. 9. From the graphs in Figs. 9(b) and 9(c) we see that the average class accuracies and detection accuracies are higher for the many foreign object training scheme, but Fig. 9(d) indicates that false positive rate is also roughly 5 times higher. The mixed approach appears to find middle ground between the two other approaches for all measures. We see that from 20 objects onwards the mixed approach is as good as the approach with a few foreign objects in terms of the false positive rate, while being superior in terms of average class accuracy and detection accuracy for up to 40 training objects. This shows that including many foreign objects in the training set for detecting few to no foreign objects in the test set has limited additional value, but mixing these with examples with objects containing a few foreign objects may result in higher detection quality while maintaining a similar false positive detection rate.

4.7. Robustness of the workflow

In the previous experiments, the trained networks are tested on a set of projections that are generated using the same 3D segmentation threshold parameter in the workflow as in the generation of the data for the training and validation sets. To assess the robustness of the workflow to different segmentation parameters, we generate the training datasets with different values of the segmentation parameter θ (see Fig. 10(a)). For each of these values, networks are trained and assessed on the test set from the previous sections. The number of training objects that are included in the workflow is fixed to 10 (which has led to equivalent results in the previous experiments as with 60 objects in the classical approach).

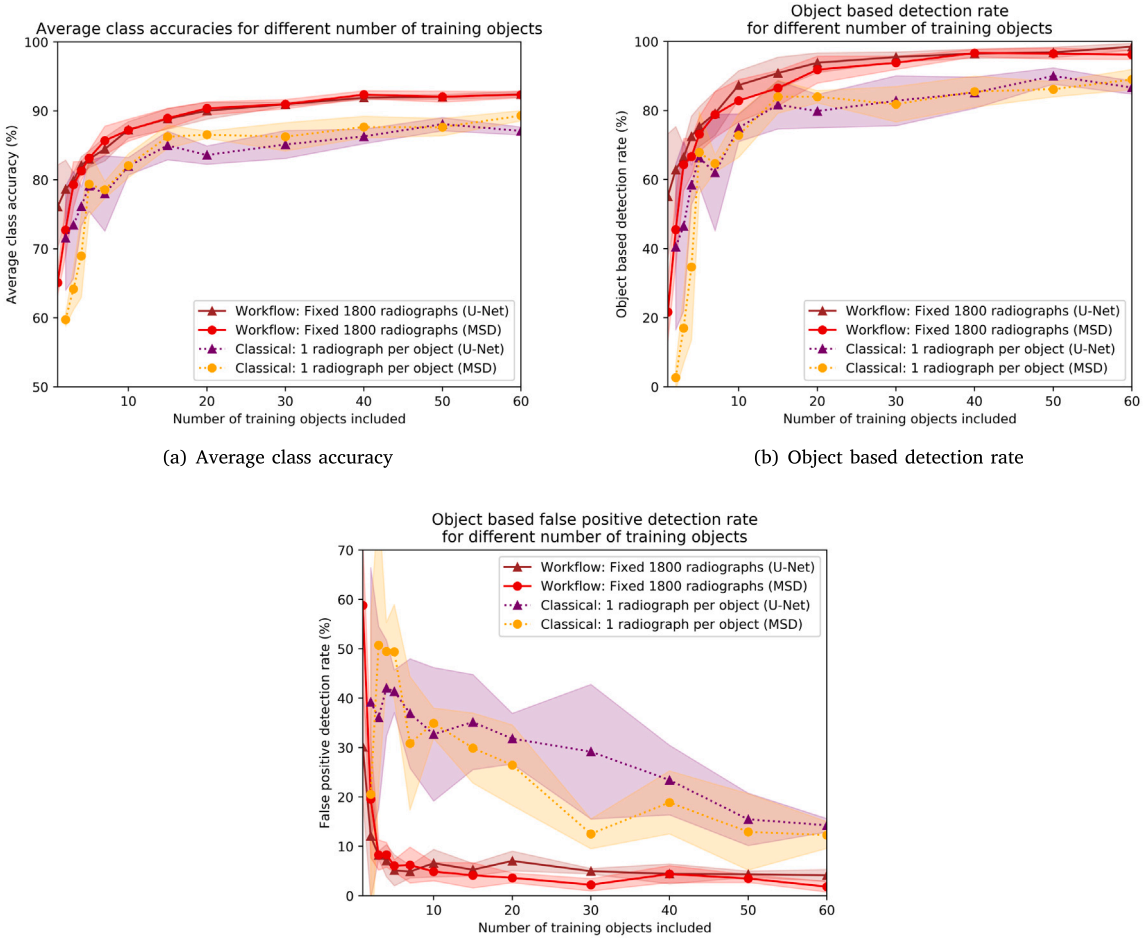
In Fig. 10, the average class accuracies, detection accuracies and the false positive rates of the trained neural networks are shown for the different thresholds. The results for U-Net and MSD are very similar. As the threshold value increases, the average class accuracy decreases, with significantly lower values for $\theta = 0.014$ and $\theta = 0.015$. The same holds for the detection rate, but it reaches a plateau between $\theta = 0.009$ and $\theta = 0.013$ where this accuracy measure gives similar values. For low values of the threshold parameter, the false positive values are high, and from $\theta = 0.011$ and higher these are low and similar to each other. Taken together, threshold parameters between $\theta = 0.011$ and $\theta = 0.013$ lead to very similar results. We conclude that for the class of objects considered in these experiments, the workflow is robust against moderate variation of the segmentation parameter and that suboptimal segmentation methods can also be used in the workflow.

4.8. Simulation experiments

In this section, we will demonstrate the workflow in a controlled simulated setting. In this way, we can verify the results with larger training and test sets when more objects are available. Furthermore, the test set previously consisted of data generated with the workflow, but in a simulated setting 'absolute' ground truth can be created for the test set by directly projecting the simulated foreign objects (see Fig. 11). We verify that the proposed workflow (with CT scanning, reconstruction and segmentation) results in segmented foreign objects of which the projections are similar to absolute ground truth projections, which further supports the confidence we can have in the experimental test results.

We have generated a set of 500 objects, each in an object space of 128^3 voxels. Each object is a cube of size 64^3 voxels, which is placed in the center of the volume. To create sufficient variety among the objects, the cube is cut off by eight planes. For each corner of the cube, a plane is created by selecting points on each of the three outgoing edges of the corner, randomly between the corner point and the midpoint of that edge. The pixels are cut off whose location is on side of the plane opposite to the center of the cube. See Fig. 12 for a visualization. Additionally, we rotate the resulting object with random angles around all axes. After that we include a foreign object as an ellipsoid with a radius randomly chosen between 3 and 7 voxels at a random location within or on the edge of the base object. These ellipsoids have a random orientation as well. As a result, the foreign objects vary in shape, size, orientation and location. With 50% probability, we include two of these foreign objects instead of one in the base object.

Based on the spectral properties of the assigned materials, we create simulated radiographs (Fig. 11b). Details of the computation can be found in the Appendix of Zeegers, Pelt, van Leeuwen, van Liere, and Batenburg (2020). First, we make projections of each material



(a) Average class accuracy

(b) Object based detection rate

(c) Object based false positive detection rate

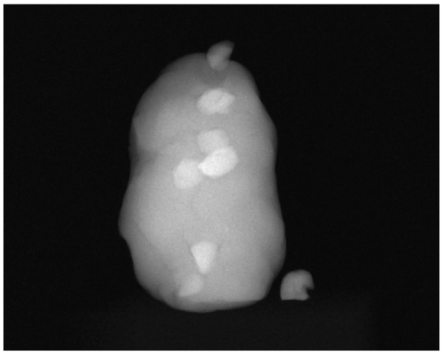
Fig. 8. The average class accuracy (**a**), the object based detection rate (**b**), and the object based false positive rate (**c**) of segmentations with trained U-Net and MSD networks on laboratory data for different number of training objects. The results are shown for the fixed number of training radiograph approach (workflow) and the one training radiograph per object approach (classical). The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the respective standard deviations.

separately by computing cone beam forward projections using the ASTRA toolbox (Van Aarle et al., 2016, 2015). From this, the simulated radiographs are computed by taking the spectral properties of each material into account (taken from the National Institute for Standards and Technology (NIST) Hubbell & Seltzer, 1995). We model the foreign objects as bone and the base object as tissue for each object. We take the spectral material characteristics between 15 KeV and 90 KeV into account, and use an exposure time of 0.002 s for each radiograph, for which the Poisson noise that is applied is relatively high. These settings are chosen such that there is sufficient contrast in the radiographs, but not as much that it can be very easily identified with simple segmentation methods. The simulated detector size – and therefore the projection image size – is 128 × 128 pixels. Examples of radiographs from five objects are given in Fig. 13.

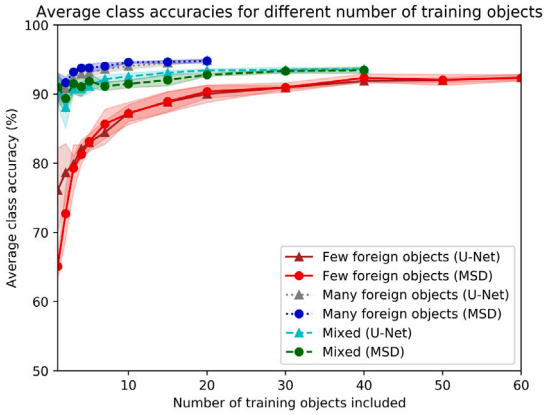
A total of 100 objects are reserved as training objects, while the other 400 objects are reserved for testing. For each training object, the ground truth corresponding to each radiograph is generated with the workflow, with the same strategy and parameters as in Section 4.2. Global thresholding with parameter value $\theta = 0.04$ is used for the reconstructions. For each test object, the 'absolute' ground truth corresponding to each radiograph is generated by directly projecting the virtual foreign objects (Fig. 11a and f), thereby skipping the reconstruction and segmentation steps. The projections are segmented such that every non-zero pixel on the detector is a projected foreign object location.

To verify that the direct use of the generated 3D volumes results in very similar ground truth projections compared to when to workflow is followed, the resulting ground truth projections are compared for the training set. The Jaccard index between the resulting ground truth pairs, averaged over all projection angles for all 100 training objects, is 0.961 for SIRT with 100 iterations. This result indicates that the resulting ground truth projections resulting from both approaches are very similar, and both are likely to yield the same quality measures.

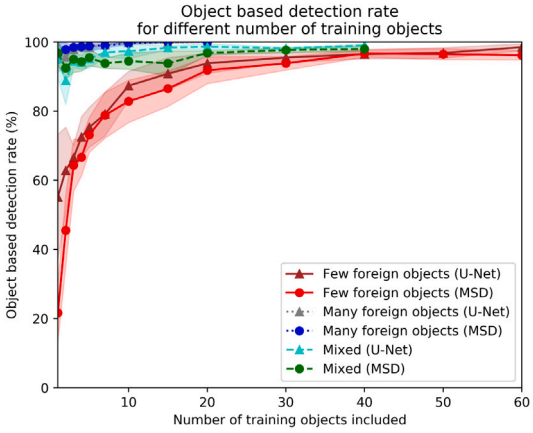
To further confirm this, the training of networks as described in Section 4.3 is repeated with the simulated projections, with the trained networks this time being evaluated on the test set with 'absolute' ground truth. The results for the three measures are given in Fig. 14, and are in accordance with the experiments with the laboratory data. A notable difference is that the average class accuracy and detection accuracy reach their maximum values for a relatively lower number of training objects (and the same goes for the minimum value of the false positive rate). This is most likely because the simulated objects are less complex, resulting in radiographs with less complicated structures. Nevertheless, the results again show inferior results for the approach where one radiograph per training object is used, since 100 objects are needed to reach similar quality measure values as for the workflow with only 7 objects.



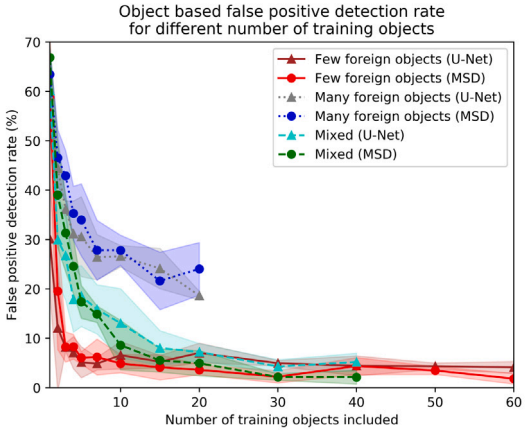
(a) Base object with many foreign objects



(b) Average class accuracy



(c) Object based detection rate



(d) Object based false positive detection rate

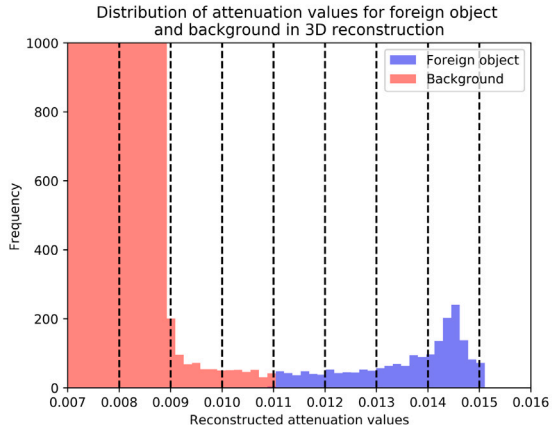
Fig. 9. Example of a radiograph of a base object with many foreign objects (**a**). The eight foreign objects are vertically placed in the object, although there still may be some overlap. The average class accuracy (**b**), the object based detection rate (**c**), and the object based false positive detection rate (**d**) of segmentations with trained U-Net and MSD networks for different number of training objects and training strategies are shown. The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the standard deviations.

5. Discussion

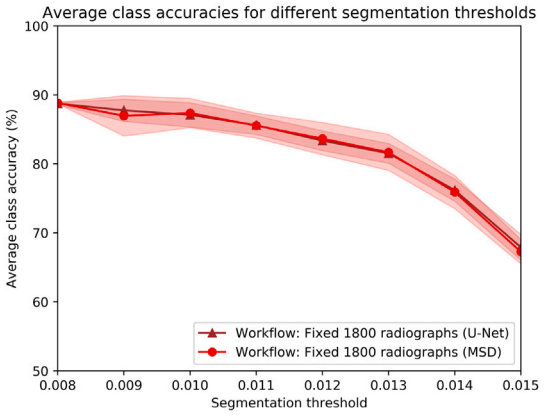
Regarding the difficulty of transferring the method to industrial settings, practical issues such as differences in the CT setup can be expected, but we expect that these are straightforward to overcome. In general, we expect that the proposed workflow can be transferred to industrial applications if the following three conditions are met. First of all, the materials should have different absorptions in the energy range of acquired radiographs (determined by the peak voltage of the source and the energetic detection range of the detector) to be distinguishable at all. Secondly, it should be possible to carry out a segmentation in the reconstruction domain, and the foreign objects should therefore not be too small. Lastly, the foreign objects should be detectable in the radiographs.

In our experiments, the Play-Doh is selected to be representative of many example products in the food industry, and the stones for the related foreign objects. This particular case meets the conditions stated above, and therefore stone can be detected in our examples. Moreover, we expect no problems with metal detection because of its higher X-ray attenuation and visibility in the radiographs, and this high attenuation may allow for even smaller metal foreign object sizes as they are more likely to still appear in the reconstruction and radiographs. On the other hand, if the objects contain large metal pieces and other materials need to be detected, it could lead to artifacts in the reconstructions, but there are many artifact reduction methods available that can be used to mitigate this (Gjesteby et al., 2016). Regarding other less dense materials such as plastics, successful application of the workflow strongly depends on the visibility of the foreign objects in the radiographs in the first place. If foreign objects are impossible to discern in radiographs, creating training data by manual annotation is also not possible. Of course, many solutions can be proposed for this invisibility problem, but this discussion is independent of the workflow for training data generation. However, even without advanced imaging methods, applying the workflow and generating 2D ground truth could lead to networks retrieving patterns in the radiographs that are difficult to find by human inspection. Additionally, we have shown in the experiments that multiple foreign objects in a sample do not pose a problem. When the material types among these are varied, we do not expect any problem as long as they are distinguishable in the radiographs. Finally, when transferring the workflow to an industrial setup, we may expect practical issues regarding the CT setup, but this is not a fundamental problem of the proposed method.

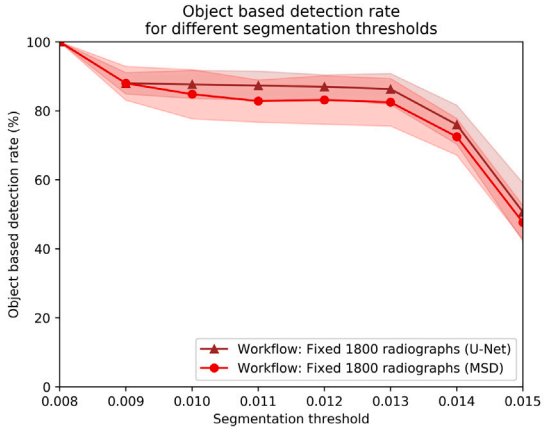
Overall, the graphs presenting the foreign object detection accuracies of stones in Play-Doh in Section 4 indicate an increase of segmentation and detection accuracy with increasing the number of objects from which the training data is created. The accuracies initially increase strongly with the number of training objects but this increase decays when the number of training objects is further increased. The detection rates and false positives rate introduced in Section 4.4 depend on the thresholds η and δ, respectively. If a higher threshold η is used in condition (2), the detection rate will decrease, because fewer objects in the target images will meet this condition. Conversely, if the threshold



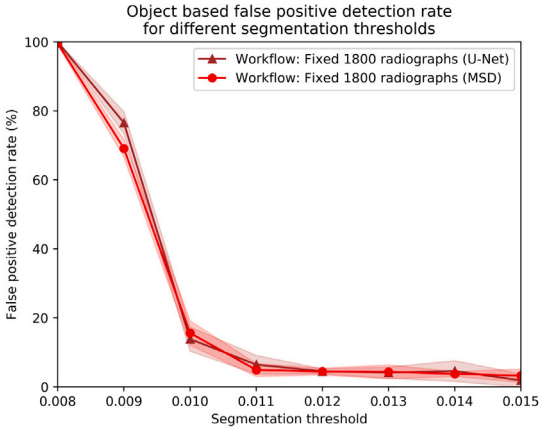
(a) Attenuation value histogram with thresholds



(b) Average class accuracy



(c) Object based detection rate



(d) Object based false positive detection rate

Fig. 10. The eight considered thresholds for the generation of the training datasets in the workflow, drawn in the histogram of attenuation values of the third object in Fig. 6, and the average class accuracy (**b**), the object based detection rate (**c**), and the object based false positive detection rate (**d**) of segmentations with trained U-Net and MSD networks on data resulting from these segmentation thresholds. The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the standard deviations.

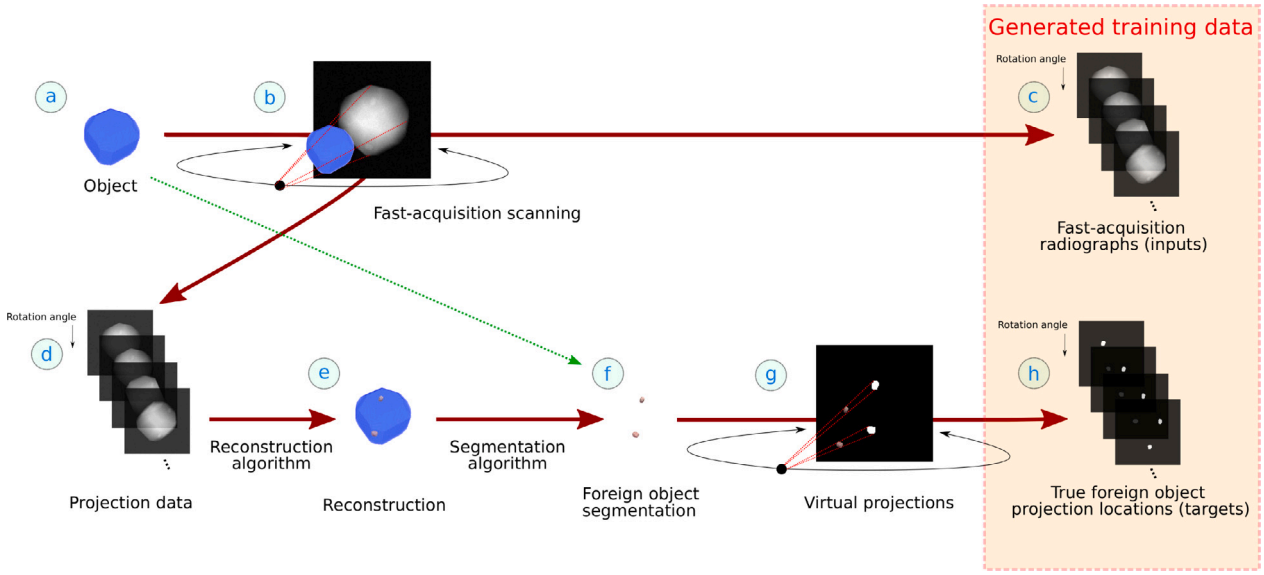


Fig. 11. The complete workflow of data acquisition (**a**,**b**) and the generation of training data (**c**,**h**) for the simulation experiment by 3D reconstruction from the CT scan (**d**, **e**), segmentation (**f**), and virtual projections (**g**). The 3D reconstruction reveals the hidden foreign objects inside the object. The dotted green arrow (**a** to **f**) indicates that because of the simulated nature of the objects, the reconstruction and segmentation steps are skipped for the generation of ground truth for objects in the test set.

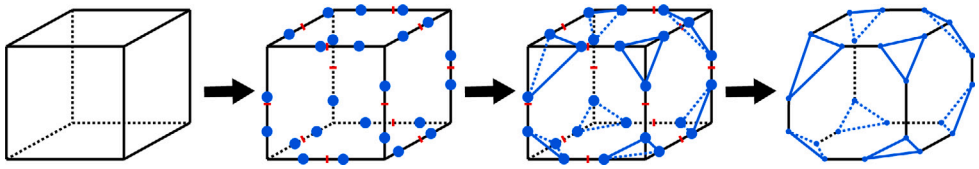


Fig. 12. The process of cutting off corners with planes from cubes for creating the simulated base objects. The red stripes indicate edge midpoints and the blue dots are the randomly chosen points between those midpoints and the corners.

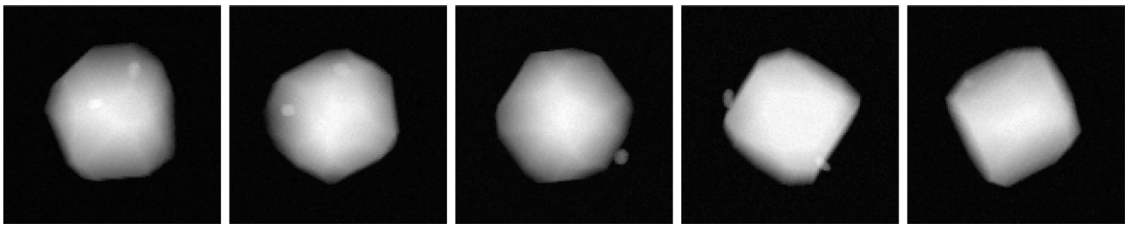
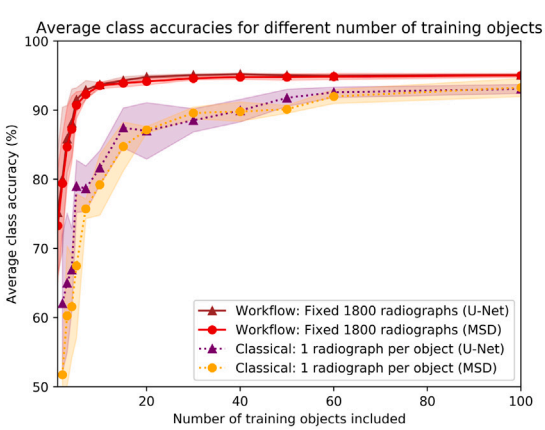
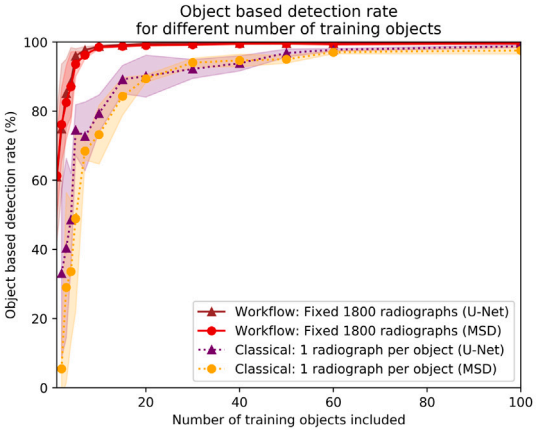


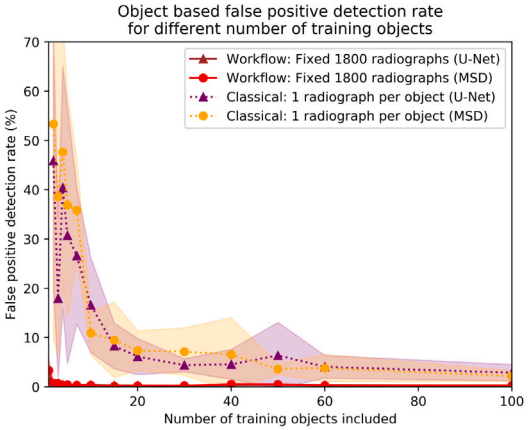
Fig. 13. Example radiographs of five simulated objects. Foreign objects are located at various positions in or on the border of the base object in the radiographs, and there can be one or two of these present.



(a) Average class accuracy



(b) Object based detection rate



(c) Object based false positive detection rate

Fig. 14. The average class accuracy, the object based detection rate, and the object based false positive detection rate of segmentations with trained U-Net and MSD networks on simulated data for different number of training objects and different training strategies. The results are averaged over 5 trained networks, with a different training object order for each run. The shaded regions indicate the standard deviations.

decreases, the detection rate will decrease. Similarly, if the δ threshold in condition (3) is increased, there will be more false positives (and fewer when the threshold is decreased). Changing the thresholds η and δ will therefore change Figs. 8, 9 and 14 as well, but we expect the overall shape of the curves and their relative distances to each other will remain similar.

When the thresholds are fixed, the maximum detection accuracy that can be achieved depends on the nature of the foreign detection

problem. For instance, if the X-ray flux is low and the noise is high, foreign objects are more difficult to detect from the radiographs. In the case of the laboratory experiments, foreign objects are difficult to detect when the cylindrical shape is located with the long edge on the ground and oriented orthogonal to the detector. The radiographs should contain sufficient discriminatory information such that foreign object detection with deep learning is possible. Additionally, for the dataset to be suitable for supervised machine learning, the ground truth should also be of sufficient quality, although this seemed to be less of an issue in our experiments as we observed no negative effects from occasional noise in the ground truth on the training and detection accuracy.

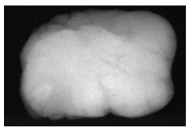
With the above considerations in mind, the workflow is designed to be modular. Every stage of the proposed workflow can be designed according to the available data-acquisition equipment, the intended detection accuracy, the type of base objects and foreign objects, and the available computer memory, among other things. We highlight some possible considerations for every stage:

- **Objects** (Fig. 3a): The set of objects can be enlarged or diversified when the accuracy of the trained neural network is not satisfactory. Also, more objects can be added to obtain a more diverse representation of objects when a more diverse array of objects or orientations are considered to be subjected to X-rays in the industrial application, such as on a conveyor belt. When a completely new type of objects is considered, these objects should be added to the workflow as well.
- **Scanning routine** (Fig. 3b): In our experimental setting we have used data resulting from low exposure times as input for both the neural networks and the reconstruction algorithm. If the foreign objects turn out to be too difficult to separate in the reconstructions, more scanning angles may be considered. Additionally, if the factory settings are allowed to be altered, higher fluxes, different tube voltages or longer exposure times can be used to obtain radiographs of higher quality, as long as the processing times remain acceptable. Also, more discrimination can be achieved by applying spectral imaging (dual-energy Rebuffel & Dinten, 2007 or multi-energy imaging Einarsson et al., 2017; Si-Mohamed et al., 2017; Taguchi et al., 2020) such that the neural network can distinguish the foreign objects from the base objects. If changing the quality of the radiographs is not possible, a separate high-quality scan of the same object can be made under the same angles, to achieve more contrast of the foreign object in the reconstructions. The scanning routine can be carried out in any lab, as long as it done under similar conditions as in the intended industrial X-ray imaging setting.
- **Reconstruction algorithm** (Fig. 3e): Depending on the type of data, different reconstruction algorithms may be considered (Buzug, 2008; Hansen, Jørgensen, & Lionheart, 2021). In this work, we have used the SIRT algorithm to account for the noise in the data, but other reconstruction algorithms such as Feldkamp–Davis–Kress (FDK) algorithm (Feldkamp, Davis, & Kress, 1984) or the Conjugate Gradient method for Least Squares (CGLS) (Hestenes & Stiefel, 1952) can be considered as well. Also, when dealing with spectral or generic multi-channel data, multi-channel reconstruction methods (Kazantsev et al., 2018; Rigie & La Rivière, 2015; Sawatzky, Xu, Schirra, & Anastasio, 2014; Semerci, Hao, Kilmer, & Miller, 2014; Zeegers, Lucka, & Batenburg, 2018) can be used to increase the reconstruction accuracy even further. When dealing with objects that may change in time, dynamic reconstruction methods can be considered (Djurabekova et al., 2019; Hauptmann, Öktem, & Schönlieb, 2020; Nikitin, Carlsson, Andersson, & Mokso, 2019).
- **Segmentation algorithm** (Fig. 3f): In this work we have used a simple global thresholding scheme, but many more segmentation methods are available, as well as approaches to reduce possible noise (Diwakar & Kumar, 2018), or bounding boxes when the location of the foreign object is known (Kern & Mastmeyer, 2021). In case of multi-channel data, a multi-dimensional thresholding scheme can be used, as well as clustering methods. Discrete reconstructions algorithms that combine reconstruction and segmentation are also available (Batenburg & Sijbers, 2011; Herman & Kuba, 1999).
- **Virtual projection** (Fig. 3g): When creating the virtual projection, post-processing on the generated ground truth projections can be applied to increase the training target quality, for instance by denoising the obtained ground truth projections.
- **Supervised learning** (Fig. 3c and h): To validate the workflow, we have used the U-Net architecture with ADAM optimization on cross entropy loss and dice loss, as well as the MSD network with ADAM optimization (Kingma & Ba, 2015) on the cross-entropy loss. Other neural network architectures (see Section 2.2) can also be considered, as well as different optimization strategies and loss functions. Note that the foreign object detection problem considered in this work may be ambiguous, since for a base object containing a foreign object another base object can theoretically be constructed (without foreign object) that results in the same radiograph. This constructed base object may have an unnatural shape when compared with other base objects, but if it happens, it may lead to inconsistent training data for the network. However, this possible problem is independent of the workflow and can be resolved by multi-spectral imaging or multi-angle imaging, and training the networks with multiple images from the same object resulting from these imaging methods. However, creating reconstructions with data from these advanced imaging methods would not be necessary.

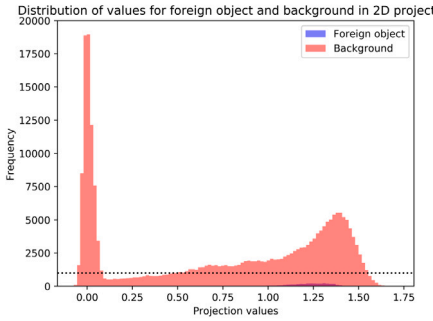
In Section 4, we have compared the segmentation results with workflow-generated data from many angles for each object to results with workflow-generated data with only one radiograph for each object. According to the results in Figs. 8 and 14, segmentation and detection accuracy can be improved by using multiple annotated radiographs for each training object. For a comparative classical approach, ideally manual annotation of the data should be carried out, taking into account the variation that may occur in different annotation for the same set of radiographs. In addition to the tremendous effort required to manually annotate a large number of radiographs, there are a number of issues that arise in the food industry. To carry out segmentation on 2D radiographs, highly specific knowledge is required for which it is difficult to find experts, as opposed to trained radiologists in medical imaging. This is also reflected in the lack of suitable publicly available datasets with annotated radiographs of food products. Also, the manual annotations may vary depending on the detection goal by a manufacturer (such as object sizes, positions or number of foreign objects). For these reasons, we eventually chose not to compare the proposed data generation with manual annotation. Manual annotation could still be used in conjunction with this workflow to replace the segmentation step. Whether or not this is feasible and will yield better results will ultimately depend on the specific application at hand.

6. Conclusions

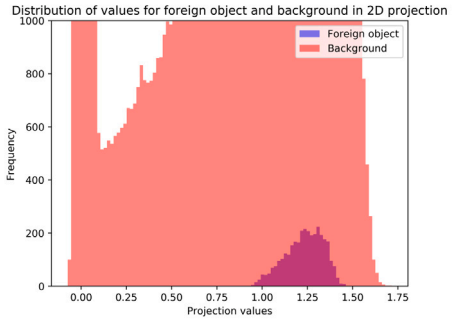
In this research, a new workflow is proposed for generating training data for supervised deep learning for foreign object detection in an industrial setting. In this workflow, a number of representative objects are scanned using X-ray imaging, reconstructed using computed tomography, segmented and virtually projected in an objective and reproducible manner to obtain the true foreign object locations in a large set of radiographs, after which supervised machine learning can be applied to detect foreign objects with high accuracy depending on the number representative objects included. We demonstrate this workflow on both laboratory and simulated data using neural networks for the deep learning task. Through laboratory experiments, we have verified



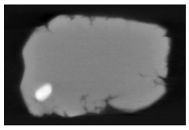
(a) 3D radiograph with foreign object on bottom left



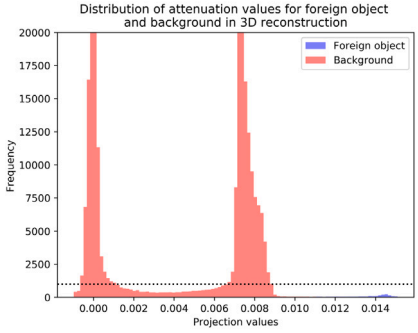
(b) Intensity value distribution for the 2D radiograph



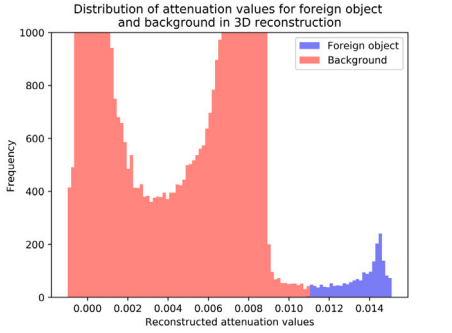
(c) Intensity value distribution for the 2D radiograph (zoomed)



(d) Slice of the reconstructed 3D volume with foreign object on bottom left



(e) Attenuation value distribution for the slice of the reconstructed 3D object



(f) Attenuation value distribution for the slice of the reconstructed 3D object (zoomed)

Fig. A.1. Radiograph of an object containing a foreign object (**a**) and a slice of the corresponding 3D reconstruction showing its attenuation values (**d**), indicating the difference in contrast. Additionally, histograms of intensity value distribution of the radiograph (**b–d**) and the attenuation value distribution of the slice of the reconstructed 3D object (**e–f**). In both cases, the histograms of the voxels or pixels of the foreign object are plotted separately from the other voxels or pixels. In the 3D volume, the foreign object is much easier to distinguish based on intensity values.

that the workflow produces adequate target images. The introduced measures assess the quality of foreign object detection with networks trained using datasets generated with this workflow. All experiments show a consistent result in which the accuracy increases significantly with a few number of training objects, and less significantly for every additional training object. In the laboratory experiment, we consistently obtain high accuracies for detecting gravel in modeling clay with low exposure times using this workflow, demonstrating its application potential in an industrial setting.

CRediT authorship contribution statement

Mathé T. Zeegers: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Tristan van Leeuwen:** Conceptualization, Writing – review & editing, Supervision, Project administration. **Daniël M. Pelt:** Conceptualization, Methodology, Software, Writing – review & editing, Project administration. **Sophia Bethany Coban:** Conceptualization, Writing – review & editing. **Robert van Liere:** Conceptualization, Writing – review & editing, Funding acquisition. **Kees Joost Batenburg:** Conceptualization, Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets generated for this paper are available at Zenodo. Separate submissions are made for the processed data resulting in radiographs with ground truth for object detection (Zeegers, 2022b), (Zeegers, 2022b), as well as the unprocessed CT scan data for complete reproduction of the results in this paper (Zeegers, 2022a) (Zeegers, 2022a).

Acknowledgments

The authors acknowledge financial support from the Netherlands Organisation for Scientific Research (NWO), on the project number 639.073.506. D. M. Pelt is supported by The Netherlands Organisation for Scientific Research (NWO), on the project number 016.Veni.192.235. The authors also acknowledge TESCAN-XRE NV for their collaboration and support of the FleX-ray laboratory.

Appendix. Intensity value histograms

We compare the intensity distributions for radiographs and for a CT reconstruction of an object in Fig. A.1, which shows a number of statistics about the pixel and voxel intensities for object 3 (Fig. 6). For both approaches, the intensity value distributions are plotted and separated into values of pixel or voxels that have been marked as foreign object by the thresholding method. The 3D case has a clear separation between foreign object and the base object based on attenuation, such that a simple global threshold based on Otsu's method (Otsu, 1979)

is sufficient to segment the foreign object. On the other hand, in the 2D radiograph case, the intensity values corresponding to the foreign object locations are similar to values of the base object.

References

Akcay, S., & Breckon, T. (2022). Towards automatic threat detection: A survey of advances of deep learning within X-ray security imaging. *Pattern Recognition*, *122*, Article 108245. http://dx.doi.org/10.1016/j.patcog.2021.108245.

Al-Sarayreh, M., Reis, M. M., Yan, W. Q., & Klette, R. (2019). A sequential CNN approach for foreign object detection in hyperspectral images. In M. Vento, & G. Percannella (Eds.), *International conference on computer analysis of images and patterns* (pp. 271–283). Springer.

Andriiashen, V., van Liere, R., van Leeuwen, T., & Batenburg, K. J. (2021). Unsupervised foreign object detection based on dual-energy absorptiometry in the food industry. *Journal of Imaging*, *7*(7), 10. http://dx.doi.org/10.3390/jimaging7070104.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(12), 2481–2495. http://dx.doi.org/10.1109/TPAMI.2016.2644615.

Batenburg, K. J., & Sijbers, J. (2011). DART: A practical reconstruction algorithm for discrete tomography. *IEEE Transactions on Image Processing*, *20*(9), 2542–2553. http://dx.doi.org/10.1109/TIP.2011.2131661.

Buzug, T. M. (2008). Computed tomography: From photon statistics to modern cone-beam CT. In *Springer handbook of medical technology* (1st ed.). (pp. 311–342). Springer.

Chartrand, G., Cheng, P. M., Vorontsov, E., Drozdzal, M., Turcotte, S., Pal, C. J., et al. (2017). Deep learning: a primer for radiologists. *Radiographics*, *37*(7), 2113–2131. http://dx.doi.org/10.1148/rg.2017170077.

Coban, S. B., Lucka, F., Palenstijn, W. J., Van Loo, D., & Batenburg, K. J. (2020). Explorative imaging and its implementation at the fleX-ray laboratory. *Journal of Imaging*, *6*(4), 18. http://dx.doi.org/10.3390/jimaging6040018.

Deshpande, A. M., Minai, A. A., & Kumar, M. (2020). One-shot recognition of manufacturing defects in steel surfaces. *Procedia Manufacturing*, *48*, 1064–1071. http://dx.doi.org/10.1016/j.promfg.2020.05.146.

Diwakar, M., & Kumar, M. (2018). A review on CT image noise and its denoising. *Biomedical Signal Processing and Control*, *42*, 73–88. http://dx.doi.org/10.1016/j.bspc.2018.01.010.

Djurabekova, N., Goldberg, A., Hauptmann, A., Hawkes, D., Long, G., Lucka, F., et al. (2019). Application of proximal alternating linearized minimization (PALM) and inertial PALM to dynamic 3D CT. In S. Matej, S. D. Metzler (Eds.), *15th international meeting on fully three-dimensional image reconstruction in radiology and nuclear medicine, Vol. 11072* (pp. 30–34). SPIE, International Society for Optics and Photonics.

Einarsdóttir, H., Emerson, M. J., Clemmensen, L. H., Scherer, K., Willer, K., Bech, M., et al. (2016). Novelty detection of foreign objects in food using multi-modal X-ray imaging. *Food Control*, *67*, 39–47. http://dx.doi.org/10.1016/j.foodcont.2016.02.023.

Einarsson, G., Jensen, J. N., Paulsen, R. R., Einarsdottir, H., Ersbøll, B. K., Dahl, A. B., et al. (2017). Foreign object detection in multispectral X-ray images of food items using sparse discriminant analysis. In P. Sharma, F. Bianchi (Eds.), *Scandinavian conference on image analysis* (pp. 350–361). Springer.

Feldkamp, L. A., Davis, L. C., & Kress, J. W. (1984). Practical cone-beam algorithm. *Journal of the Optical Society of America A*, *1*(6), 612–619. http://dx.doi.org/10.1364/JOSAA.1.000612.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., & Garcia-Rodriguez, J. (2017). A review on deep learning techniques applied to semantic segmentation. arXiv preprint arXiv:1704.06857.

Gjesteby, L., De Man, B., Jin, Y., Paganetti, H., Verburg, J., Giantsoudi, D., et al. (2016). Metal artifact reduction in CT: where are we after four decades? *Ieee Access*, *4*, 5826–5849.

Grandini, M., Bagli, E., & Visani, G. (2020). Metrics for multi-class classification: an overview. arXiv preprint arXiv:2008.05756.

Guo, Y., Liu, Y., Georgiou, T., & Lew, M. S. (2018). A review of semantic segmentation using deep neural networks. *International Journal of Multimedia Information Retrieval*, *7*(2), 87–93. http://dx.doi.org/10.1007/s13735-017-0141-z.

Haff, R. P., & Toyofuku, N. (2008). X-ray detection of defects and contaminants in the food industry. *Sensing and Instrumentation for Food Quality and Safety*, *2*(4), 262–273. http://dx.doi.org/10.1007/s11694-008-9059-8.

Hansen, P. C., Jørgensen, J. S., & Lionheart, W. R. B. (2021). *Computed tomography: algorithms, insight, and just enough theory* (1st ed.). SIAM.

Hauptmann, A., Öktem, O., & Schönlieb, C. (2020). Image reconstruction in dynamic inverse problems with temporal models. arXiv preprint arXiv:2007.10238.

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask R-CNN. In *Proceedings of the ieee international conference on computer vision* (pp. 2961–2969). IEEE.

He, Y., Xiao, Q., Bai, X., Zhou, L., Liu, F., & Zhang, C. (2021). Recent progress of nondestructive techniques for fruits damage inspection: a review. *Critical Reviews in Food Science and Nutrition*, 1–19. http://dx.doi.org/10.1080/10408398.2021.1885342.

Hendriksen, A. A., Pelt, D. M., & Batenburg, K. J. (2020). Noise2inverse: Self-supervised deep convolutional denoising for tomography. *IEEE Transactions on Computational Imaging*, *6*, 1320–1335. http://dx.doi.org/10.1109/TCI.2020.3019647.

Herman, G. T., & Kuba, A. (1999). *Discrete tomography: Foundations, algorithms, and applications* (1st ed.). Springer.

Hestenes, M. R., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, *49*, 409–436. http://dx.doi.org/10.6028/JRES.049.044.

Hubbell, J. H., & Seltzer, S. M. (1995). Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 KeV to 20 MeV for elements Z=1 to 92 and 48 additional substances of dosimetric interest. *Tech. Rep.*, National Institute of Standards and Technology-PL, Gaithersburg, MD, USA. Ionizing Radiation Div..

Jadon, S. (2020). A survey of loss functions for semantic segmentation. arXiv preprint arXiv:2006.14822.

Kak, A. C., Slaney, M., & Wang, G. (2002). *Principles of computerized tomographic imaging* (1st ed.). Wiley Online Library.

Kazantsev, D., Jørgensen, J. S., Andersen, M. S., Lionheart, W. R. B., Lee, P. D., & Withers, P. J. (2018). Joint image reconstruction method with correlative multi-channel prior for X-ray spectral computed tomography. *Inverse Problems*, *34*(6), Article 064001. http://dx.doi.org/10.1088/1361-6420/aaba86.

Kern, D., & Mastmeyer, A. (2021). 3D bounding box detection in volumetric medical image data: A systematic literature review. In *2021 IEEE 8th international conference on industrial engineering and applications* (pp. 509–516). IEEE.

Kingma, D. P., & Ba, L. (2015). ADAM: A method for stochastic optimization. In Y. Bengio, & Y. LeCun (Eds.), *In proceedings of the international conference on learning representations*.

Kwon, J., Lee, J., & Kim, W. (2008). Real-time detection of foreign objects using X-ray imaging for dry food manufacturing line. In *2008 IEEE international symposium on consumer electronics* (pp. 1–4). IEEE.

Lagerwerf, M. J., Pelt, D. M., Palenstijn, W. J., & Batenburg, K. J. (2020). A computationally efficient reconstruction algorithm for circular cone-beam computed tomography using shallow neural networks. *Journal of Imaging*, *6*(12), 135. http://dx.doi.org/10.3390/jimaging6120135.

Lenchik, L., Heacock, L., Weaver, A. A., Boutin, R. D., Cook, T. S., Itri, J., et al. (2019). Automated segmentation of tissues using CT and MRI: a systematic review. *Academic Radiology*, *26*(12), 1695–1706. http://dx.doi.org/10.1016/j.acra.2019.07.006.

Li, S., Luo, H., Hu, M., Zhang, M., Feng, J., Liu, Y., et al. (2019). Optical non-destructive techniques for small berry fruits: A review. *Artificial Intelligence in Agriculture*, *2*, 85–98. http://dx.doi.org/10.1016/j.aiia.2019.07.002.

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1925–1934). IEEE.

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440). IEEE.

Mathanker, S. K., Weckler, P. R., & Bowser, T. J. (2013). X-ray applications in food and agriculture: A review. *Transactions of the ASABE*, *56*(3), 1227–1239. http://dx.doi.org/10.13031/trans.56.9785.

Mery, D., Lillo, I., Loebel, H., Riffo, V., Soto, A., Cipriano, A., et al. (2011). Automated fish bone detection using X-ray imaging. *Journal of Food Engineering*, *105*(3), 485–492. http://dx.doi.org/10.1016/j.jfoodeng.2011.03.007.

Mohd Khairi, M. T., Ibrahim, S., Md Yunus, M. A., & Faramarzi, M. (2018). Noninvasive techniques for detection of foreign bodies in food: A review. *Journal of Food Process Engineering*, *41*(6), Article e12808. http://dx.doi.org/10.1111/jfpe.12808.

Narsaiah, K., Biswas, A. K., & Mandal, P. K. (2020). Nondestructive methods for carcass and meat quality evaluation. In A. K. Biswas, & P. K. Mandal (Eds.), *Meat quality analysis* (pp. 37–49). Academic Press: http://dx.doi.org/10.1016/B978-0-12-819233-7.00003-3.

Nicolaï, B. M., Defraeye, T., De Ketelaere, B., Herremans, E., Hertog, M. L. A. T. M., Saeys, W., et al. (2014). Nondestructive measurement of fruit and vegetable quality. *Annual Review of Food Science and Technology*, *5*, 285–312. http://dx.doi.org/10.1146/annurev-food-030713-092410.

Nikitin, V. V., Carlsson, M., Andersson, F., & Mokso, R. (2019). Four-dimensional tomographic reconstruction by time domain decomposition. *IEEE Transactions on Computational Imaging*, *5*(3), 409–419. http://dx.doi.org/10.1109/TCI.2019.2898088.

Noh, H., Hong, S., & Han, B. (2015). Learning deconvolution network for semantic segmentation. In *Proceedings of the IEEE international conference on computer vision* (pp. 1520–1528). IEEE.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. http://dx.doi.org/10.1109/TSMC.1979.4310076.

Pan, H., Zhou, C., Zhu, Q., & Zheng, D. (2018). A fast registration from 3D CT images to 2D X-ray images. In *2018 IEEE 3rd international conference on big data analysis* (pp. 351–355). IEEE.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., et al. (2017). Automatic differentiation in PyTorch. In *NIPS-W*.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, R. Garnett (Eds.), *Advances in neural information processing systems* (pp. 8026–8037).

Pelt, D. M. (2019). GitHub - dmpelt/msdnet: Python implementation of the mixed-scale dense convolutional neural network. https://github.com/dmpelt/msdnet (Accessed on 24 November 2020).

Pelt, D. M., Batenburg, K. J., & Sethian, J. A. (2018). Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks. *Journal of Imaging*, *4*(11), 128. http://dx.doi.org/10.3390/jimaging4110128.

Pelt, D. M., & Sethian, J. A. (2018). A mixed-scale dense convolutional neural network for image analysis. *Proceedings of the National Academy of Sciences*, *115*(2), 254–259. http://dx.doi.org/10.1073/pnas.1715832114.

Rebuffel, V., & Dinten, J. (2007). Dual-energy X-ray imaging: benefits and limits. *Insight - Non-Destructive Testing and Condition Monitoring*, *49*(10), 589–594. http://dx.doi.org/10.1784/insi.2007.49.10.589.

Rigie, D. S., & La Rivière, P. J. (2015). Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization. *Physics in Medicine and Biology*, *60*(5), 1741. http://dx.doi.org/10.1088/0031-9155/60/5/1741.

Rong, D., Xie, L., & Ying, Y. (2019). Computer vision detection of foreign objects in walnuts using deep learning. *Computers and Electronics in Agriculture*, *162*, 1001–1010. http://dx.doi.org/10.1016/j.compag.2019.05.019.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *International conference on medical image computing and computer-assisted intervention* (pp. 234–241). Springer.

Russo, P. (2017). *Handbook of X-ray imaging: physics and technology* (1st ed.). CRC Press.

Sawatzky, A., Xu, Q., Schirra, C. O., & Anastasio, M. A. (2014). Proximal ADMM for multi-channel image reconstruction in spectral X-ray CT. *IEEE Transactions on Medical Imaging*, *33*(8), 1657–1668. http://dx.doi.org/10.1109/TMI.2014.2321098.

Semerci, O., Hao, N., Kilmer, M. E., & Miller, E. L. (2014). Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. *IEEE Transactions on Image Processing*, *23*(4), 1678–1693. http://dx.doi.org/10.1109/TIP.2014.2305840.

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging*, *13*(1), 146–165. http://dx.doi.org/10.1117/1.1631315.

Si-Mohamed, S., Bar-Ness, D., Sigovan, M., Cormode, D. P., Coulon, P., Coche, E., et al. (2017). Review of an initial experience with an experimental spectral photon-counting computed tomography system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *873*, 27–35. http://dx.doi.org/10.1016/j.nima.2017.04.014.

Silva, G., Oliveira, L., & Pithon, M. (2018). Automatic segmenting teeth in X-ray images: Trends, a novel data set, benchmarking and future perspectives. *Expert Systems with Applications*, *107*, 15–31. http://dx.doi.org/10.1016/j.eswa.2018.04.001.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., & Cardoso, M. J. (2017). Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In M. J. Cardoso, T. Arbel, G. Carneiro, T. Syeda-Mahmood, J. M. R. S. Tavares, M. Moradi, A. Bradley, H. Greenspan, J. P. Papa, A. Madabhushi, J. C. Nascimento, J. S. Cardoso, V. Belagiannis, & Z. Lu (Eds.), *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 240–248). Springer.

Taguchi, K., Blevis, I., & Iniewski, K. (2020). *Spectral, photon counting computed tomography: technology and applications* (1st ed.). CRC Press.

Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, Article 101693. http://dx.doi.org/10.1016/j.media.2020.101693.

Van Aarle, W., Palenstijn, W. J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., et al. (2016). Fast and flexible X-ray tomography using the ASTRA toolbox. *Optics Express*, *24*(22), 25129–25147. http://dx.doi.org/10.1364/OE.24.025129.

Van Aarle, W., Palenstijn, W. J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K. J., et al. (2015). The ASTRA toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy*, *157*, 35–47. http://dx.doi.org/10.1016/j.ultramic.2015.05.002.

Van De Looverbosch, T., Raeymaekers, E., Verboven, P., Sijbers, J., & Nicolaï, B. (2021). Non-destructive internal disorder detection of conference pears by semantic segmentation of X-ray CT scans using deep learning. *Expert Systems with Applications*, *176*, Article 114925. http://dx.doi.org/10.1016/j.eswa.2021.114925.

Van der Sluis, A., & Van der Vorst, H. A. (1990). SIRT-and CG-type methods for the iterative solution of sparse linear least-squares problems. *Linear Algebra and its Applications*, *130*, 257–303. http://dx.doi.org/10.1016/0024-3795(90)90215-X.

Wilm, K. H. (2012). Foreign object detection: Integration in food production. *Food Safety Magazine*, *18*, 14–17.

Wu, H., Liu, Q., & Liu, X. (2019). A review on deep learning approaches to image classification and object segmentation. *Computers, Materials & Continua*, *60*(2), 575–597. http://dx.doi.org/10.32604/cmc.2019.03595.

Xiong, Z., Sun, D., Pu, H., Gao, W., & Dai, Q. (2017). Applications of emerging imaging techniques for meat quality and safety detection and evaluation: A review. *Critical Reviews in Food Science and Nutrition*, *57*(4), 755–768. http://dx.doi.org/10.1080/10408398.2014.954282.

Zeegers, M. T. (2022a). *A collection of 131 CT datasets of pieces of modeling clay containing stones*. Zenodo, http://dx.doi.org/10.5281/zenodo.5866228.

Zeegers, M. T. (2022b). *A collection of X-ray projections of 131 pieces of modeling clay containing stones for machine learning-driven object detection*. Zenodo, http://dx.doi.org/10.5281/zenodo.5681008.

Zeegers, M. T., Lucka, F., & Batenburg, K. J. (2018). A multi-channel DART algorithm. In R. P. Barneva, V. E. Brimkov, & J. M. R. S. Tavares (Eds.), *International workshop on combinatorial image analysis* (pp. 164–178). Springer.

Zeegers, M. T., Pelt, D. M., van Leeuwen, T., van Liere, R., & Batenburg, K. J. (2020). Task-driven learned hyperspectral data reduction using end-to-end supervised deep learning. *Journal of Imaging*, *6*(12), 132. http://dx.doi.org/10.3390/jimaging6120132.

Zhao, H., Shi, J., Qi, X., Wang, X., & Jia, J. (2017). Pyramid scene parsing network. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 2881–2890). IEEE.

Zhao, Z., Zheng, P., Xu, S., & Wu, X. (2019). Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *30*(11), 3212–3232. http://dx.doi.org/10.1109/tnnls.2018.2876865.

Zhong, J., Zhang, F., Lu, Z., Liu, Y., & Wang, X. (2019). High-speed display-delayed planar X-ray inspection system for the fast detection of small fishbones. *Journal of Food Process Engineering*, *42*(3), Article e13010. http://dx.doi.org/10.1111/jfpe.13010.

Zhu, L., Spachos, P., Pensini, E., & Plataniotis, K. N. (2021). Deep learning and machine vision for food processing: A survey. *Current Research in Food Science*, *4*, 233–249. http://dx.doi.org/10.1016/j.crfs.2021.03.009.