

Article

Simulation Study: Data-Driven Material Decomposition in Industrial X-ray Computed Tomography

Moritz Weiss ^{1,2,*}, Nick Brierley ², Mirko von Schmid ² and Tobias Meisen ¹

¹ Institute for Technologies and Management of Digital Transformation, University of Wuppertal, 42119 Wuppertal, Germany; meisen@uni-wuppertal.de

² diondo GmbH, 45525 Hattingen, Germany; nick.brierley@diondo.com (N.B.); mirko.vonschmid@diondo.com (M.v.S.)

* Correspondence: moritz.weiss@uni-wuppertal.de

[†] Current address: Institute for Technologies and Management of Digital Transformation, University of Wuppertal, Gausstr. 20, 42119 Wuppertal, Germany.

Abstract: Material-resolving computed tomography is a powerful and well-proven tool for various clinical applications. For industrial scan setups and materials, several problems, such as K-edge absence and beam hardening, prevent the direct transfer of these methods. This work applies dual-energy computed tomography methods for material decomposition to simulated phantoms composed of industry-relevant materials such as magnesium, aluminium and iron, as well as some commonly used alloys like Al-Si and Ti64. Challenges and limitations for multi-material decomposition are discussed in the context of X-ray absorption physics, which provides spectral information that can be ambiguous. A deep learning model, derived from a clinical use case and based on the popular U-Net, was utilised in this study. For various reasons outlined below, the training dataset was simulated, whereby phantom shapes and material properties were sampled arbitrarily. The detector signal is computed by a forward projector followed by Beer–Lambert law integration. Our trained model could predict two-material systems with different elements, achieving a relative error of approximately 1% through simulated data. For the discrimination of the element titanium and its alloy Ti64, which were also simulated, the relative error increased to 5% due to their similar X-ray absorption coefficients. To access authentic CT data, the model underwent testing using a 10c euro coin composed of an alloy known as Nordic gold. The model detected copper as the main constituent correctly, but the relative fraction, which should be 89%, was predicted to be $\approx 70\%$.

Keywords: computer vision; deep learning; X-ray imaging; computed tomography; dual-energy computed tomography



Citation: Weiss, M.; Brierley, N.; von Schmid, M.; Meisen, T. Simulation Study: Data-Driven Material Decomposition in Industrial X-ray Computed Tomography. *NDT* **2024**, *2*, 1–15. <https://doi.org/10.3390/ndt2010001>

Academic Editor: Fabio Tosti

Received: 26 October 2023

Revised: 14 December 2023

Accepted: 18 December 2023

Published: 5 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

X-ray computed tomography (CT) is a non-destructive, three-dimensional imaging technique which is used for a broad spectrum of applications. In clinical practice, CT is utilised as an *in vivo* diagnostic method in relation to a wide range of diseases and injuries, as well as a planning tool for interventions. Industrially, CT is used as an advanced method of quality control and failure analysis and applied to a wide range of samples, including materials such as metals, plastic or advanced composites.

For clinical and industrial use cases, dual-energy CT (DECT) offers the possibility of measuring additional spectral information about an object. DECT projections or tomograms can be used, for example, to discriminate between different materials (material decomposition) [1].

Incorporating more data, like additional energy channels from a DECT scan, into the processing pipeline, the engineering complexity of analysis algorithms increases rapidly. In fact, the problem consists of a large number of variables on the input side with hidden or

even unknown correlations between the variables. Especially for this type of problem, data-driven methods offer the benefit of learning the mapping function from a data pool without the knowledge of the underlying mechanism. Complex multi-material compositions and structures are nowadays needed for numerous industrial, medical and consumer products. The demand for material-resolving CT is expected to increase, especially given the new manufacturing possibilities that are enabled by additive manufacturing techniques. Safety-critical parts, in particular, necessitate a high level of quality assurance, which may include confirmation of locally varying composition.

This work is a general simulation study that describes a fundamentally new concept for discrete material decomposition using an industrial CT scan setup with a data-driven method. For this purpose, established data-driven methods from clinical CT applications will be transferred and adapted. Our contributions are as follows:

- A fast CT simulation pipeline which is capable of generating thousands of tomograms for real-time training of the data-driven model.
- Quantitative end-to-end material decomposition results of simulated alloys without relying on K-edge absorption and leveraging spatial information.

The last point is of high interest because the lack of K-edge absorption forces the data-driven model to exploit other features, such as cupping, a common CT artefact, and object or particle size, which cannot be extracted by classical pixel-wise algorithms. The data-driven model learns this highly complex mapping without further engineering effort. On top of this, the model may resolve material mixture ambiguities by extracting additional but hidden information given by the dual-energy tomograms.

The remainder of this paper is divided into an overview of the CT background, related work, the description of the data simulation, an overview of the data-driven model, selected results showing representative or established material systems from a simulation, results of a real scan of a 10c euro coin and an outlook. Additionally, we discuss the reasons for certain boundaries given by the model itself, as well as the limitations given by physics and how to exploit them in future work.

2. Computed Tomography Background

Following the discovery of X-rays in the late 19th century, it took more than 75 years to reach the next technological milestone necessary for three-dimensional X-ray imaging—the first commercial computed tomography scanner. X-ray images are projections of an object's interior onto a detector plane, where spatial depth information about the scanned object is lost. Using reconstruction algorithms, depth information is retrieved by a mathematical combination of all projections gathered from different angles of the same object. In detail, the volume, which is the output of the reconstruction, is discretised on a grid of three-dimensional volume pixels (voxels). Johann Radon introduced the underlying equation, which was named after him as the Radon transform [2]. It is used to calculate projections of an object on a detector plane from different angles. Therefore, the inverse Radon transform is a reconstruction algorithm. In addition to the Radon transform, which is only formulated for parallel-beam X-ray geometries, Feldkamp, Davis and Kress proposed a cone-beam reconstruction algorithm (FDK algorithm), which allows the use of a point source for X-rays and a two-dimensional detector in a practical setup at a reasonable cost of computation [3].

The properties of the X-ray source are typically tuned to the individual use case. To penetrate objects with a high X-ray attenuation coefficient, the X-ray spectrum can be hardened by increasing the acceleration voltage. Especially for strongly attenuating objects, the use of pre-filtration can improve the overall image contrast measured by the detector. The X-ray spectra produced by laboratory sources are initially polychromatic due to the Bremsstrahlung effect by which X-rays are generated in the source's target. More information about the X-ray properties is provided in Section 4.2.

On the detector side, two basic technologies can be distinguished: photon-counting detectors (PCD) and energy-integrating detectors (EID). As the name implies, EIDs integrate the energy of arriving photons, which causes a loss of spectral information. Characteristic

K-edge absorption is usually not observable using an EID without prior knowledge and a careful setup of the X-ray source. In contrast, photon counting detectors measure counts resolved by multiple energy bins across a certain range. Therefore, the spectral information carried by the incoming photons is squeezed into the energy bins but is not completely lost.

Industrial and clinical CT systems differ in their machine setup. While clinical CT scans attempt to provide sufficient image quality for diagnosis using the lowest amount of X-ray dose possible, industrial CT scans are typically applied to objects that do not suffer damage from high radiation exposures. It is also practical for the clinical application to allow the patient to lie down and rotate the scanner around the body, which is generally called gantry CT. For dual-energy measurements, different acquisition strategies can be used to collect multiple sets of projections at different energy levels. In case the overall measurement time is not capped and the object is fixed, the easiest strategy is to run two scans (dual-scan) sequentially with the same parameters but with different X-ray spectra. For DECT measurements in the presence of patient/sample motion, alternative strategies are used, including either two independent CT imaging lines simultaneously (dual-source) or using an X-ray tube, which supports fast kVp-switching. A further approach uses the independent energy channels of a PCD or dual-layer detector to collect spectral information. All DECT acquisition strategies lead to two tomograms that are not perfectly registered. Mechanical uncertainties in the CT geometry and the spatial shift of the focal spot or movement of the sample or patient have to be taken into account when processing the DECT tomograms.

Taking a closer look at the reconstruction algorithm, it returns linear attenuation coefficients $\mu_L^{(x,y,z)}$ across three spatial dimensions (x, y, z) , an arrangement also known as a tomogram. A material has a specific X-ray mass attenuation coefficient $\mu_m(E)$, which can be multiplied by the material's density ρ to obtain the linear attenuation coefficient

$$\mu_L(E) = \mu_m(E) \cdot \rho. \quad (1)$$

Looking at a single volume element, different materials with different volume fractions can be present, so the total effective $\mu_L(E)$ is a linear combination of the individual $\mu_{L,i}(E)$ where i enumerates the materials. Especially with more than two materials present, multiple combinations of $\mu_{m,i}(E)$ and ρ_i can yield the same $\mu_L(E)$ measured for a given spectrum.

3. Related Work

Deep-learning approaches for material decomposition are rarely seen in industrial applications. Fang et al. proposed a neural network for material decomposition using CT scans of cargo containers in the MeV range [4]. Their method discriminates discrete materials without allowing mixtures, but we are looking for a method to calculate material fractions in multi-material systems. For this problem, several methods were studied for clinical applications, which will be discussed in the following.

Heismann et al. published a fundamental method for the atomic number and density decomposition of dual-energy CT measurements [1]. Their algorithm involves knowledge about the source spectrum as well as the spectral response of the detector, which can be quite challenging to measure exactly for a real CT scanner. Simply put, the method maps two linear attenuation coefficients given by reconstructed dual-energy pixels to two values that can be interpreted as atomic number and density. For material mixtures, the calculated atomic number and density are averages of the underlying components, as the linear attenuation coefficients result from combining those of the components. Extending Equation (1) for mixtures containing N materials, the linear attenuation coefficient μ_L is given by

$$\mu_L(E) = \sum_{i=0}^N \mu_{m,i}(E) \cdot \rho_i. \quad (2)$$

Therefore, it is only defined for reconstructed volumes. This approach does not use any spatial context knowledge, which means that the proposed algorithm is rather

vulnerable to a low signal-noise ratio and CT artefacts like ring and beam hardening artefacts, to name just a few. Additionally, due to the lack of spatial context, the dual-energy tomograms must be perfectly registered. Depending on the DECT acquisition strategy (dual-scan, dual-source and fast kVp-switching), the registration of the dual-energy tomograms can be quite challenging. To our advantage, unlike living objects, industrially scanned objects rarely move, so the dual-scan strategy is applicable if the measurement is not time-critical. In contrast to Heismann's approach, which incorporates expert domain knowledge about X-ray CT, a data-driven approach has been pursued by several authors, which exclusively operates on a large amount of exemplary data showing inputs and outputs of the problem. Badea et al. proposed an end-to-end method to directly calculate material maps from provided DECT tomograms [5]. They simulate two-dimensional phantoms using a Delaunay algorithm [6], which samples triangular regions. Each region consists of random volume fractions of the base materials. Subsequently, they use a forward projector with different source spectra to calculate X-ray projections and an FDK reconstruction to calculate the corresponding tomograms for each spectrum. The multi-energy tomograms are then used as the input, and the initially sampled phantoms are used as the ground truth output for their model. As described in Section 4, our approach follows this procedure for generating the training data. In an interesting extension, Abascal et al. used tomograms generated from the forward projections of experimental data [7], which narrows the sim-to-real gap discussed later but also restricts the amount of training data available. In a nutshell, there is currently no benchmark available as we have not come across an approach that suits our intended application.

4. Methods

This section describes the basic physics behind material decomposition with DECT and the methods used for data collection, model architecture and training. Since the performance of a neural network relies on the quality and amount of data used for training, it is very important to use data distributions matching the observed, case-specific distribution. If the training data originates from a simulation and is not gathered and labelled from the target distribution directly, the trained model might face a gap, which is termed the sim-to-real gap. Nevertheless, this sim-to-real gap may be small and surmountable by methods that will not be described in the scope of this paper but will be studied in future work in order to use our algorithm on real-world CT scans.

Being a niche use case, we did not find any pre-trained models fitting into our environment, so we decided to train the model from scratch with data collected from a simulation pipeline, which will be described in Section 4.1.

4.1. CT Simulation

Broadly, a CT scan consists of the acquisition of a sequence of projections followed by a reconstruction. The projections p_θ collected by the detector are formally the result of a projection operator \mathcal{P}_θ applied to the scanned object A under the CT angle θ resulting in

$$p_\theta = \mathcal{P}_\theta A. \quad (3)$$

As described in the introduction, the recombination of all projections p_θ by the reconstruction operator \mathcal{P}_θ^{-1} results in the tomogram A'

$$A' = \int_0^{2\pi} d\theta \mathcal{P}_\theta^{-1} p_\theta. \quad (4)$$

This is, regardless of the artefacts, a representation of the original object A . Equation (4) integrates over a full-circle trajectory in the interval $[0; 2\pi]$ since this is the scan mode used in the scope of this paper. It is also possible to reconstruct with projections gathered in the interval $[0; \pi + \delta]$ where δ is the fan-beam opening angle. For our model, we used tomograms as inputs and ground truth phantoms as outputs. An

overview of the data generation pipeline is provided in Figure 1. The phantoms with areas of different material compositions were generated by the phantom sampler, which uses a Delaunay triangulation to create areas within certain boundaries randomly on a given grid. For each triangular area, the material volume fractions were polled randomly from a uniform distribution regarding overall volume conservation. The outer bounding box of the phantoms was a two-dimensional square, which had an edge length between 10 mm and 50 mm, polled randomly. In the following, LE and HE will refer to the low- and high-energy channels of the DECT measurements. The phantoms were projected and reconstructed using the LE proj + reco and HE proj + reco methods, which wrapped the ASTRA Toolbox [8,9]. To maintain proximity to the industrial application, a fan-beam geometry was used together with an FDK reconstruction. The projection operator is the critical part of this pipeline. Looking more closely at the projection operator, this can be divided into a raytracer, which measures the intersection length of a specific ray through the object, and a photon pipeline, which calculates the signal in a pixel resulting from the X-ray spectrum and the attenuation physics in the object as well as the detector. For this work, we used the forward projector implemented in the ASTRA toolbox, which integrates densities along the intersection lengths of a ray through the object. The intersection lengths were fed into the photon pipeline, which applied the Beer–Lambert law pixel-wise for some initial spectrum $I_0(E)$, the mass attenuation coefficient $\mu_{m,i}$ of the phantom’s materials and the phantom’s intersection lengths x_i^ρ weighted by the density ρ , while summing over i as the material index:

$$I_{\text{Phantom}}(E) = I_0(E) \cdot e^{-\sum_i \mu_{m,i}(E) \cdot x_i^\rho}. \quad (5)$$

Remember that the forward projector of the ASTRA Toolbox measures intersection lengths x_i^ρ through voxel volumes weighted by the voxel values. Due to this handy mechanism, the density ρ is already multiplied by the intersection length and vanishes from the exponent in Equation (5). Equation (5) yields the photon counts behind the phantom for each energy bin. Subsequently, the same equation is applied again to calculate the photon counts after traversing the scintillator in the detector, which yields $I_{\text{Scintillator}}(E)$ and the number of photons absorbed in the scintillator for each energy bin.

$$\text{Signal} = \int_{\min(E_\gamma)}^{\max(E_\gamma)} dE \zeta(I_{\text{Phantom}}(E) - I_{\text{Scintillator}}(E)) \quad (6)$$

Integration over all energy bins with the according photon counts yields the detector signal in Equation (6). Prior to the energy integration, the photon count within a specific energy bin, denoted as E , undergoes adjustment through the application of the photon statistics operator, represented as ζ . This operator extracts a random number from a Gaussian distribution, with the actual photon count serving as the mean value and the square root of the photon count acting as the standard deviation. The signal from Equation (6) was normalised to the interval of the 16-bit unsigned integer datatype to maintain proximity to real detectors in the experiment. Some constant electronic dark field noise was added to the signal after normalisation to mimic the basic noise properties of a real detector. Caesium iodide (CsI) was chosen as a scintillator material. Currently, we do not simulate higher-order effects like, for example, the conversion efficiency of the photodiodes, pixel cross-talk or scattering inside the detector. For best performance, the photon pipeline was implemented using Nvidia’s CUDA API. The photon pipeline is likely to critically affect the magnitude of the sim-to-real gap, but it also has significant execution time implications. Therefore, a balance between performance and accuracy is required, and we decided to neglect photon scattering for the scope of this work. The simulation pipeline, as shown in Figure 1, is able to generate an image tuple in about 0.25 s using eight threads on an Intel Xeon E5-1650 v4 hexacore CPU and a Nvidia 1080Ti GPU for reconstruction at a resolution of $2 \times 256 \times 256$ (channels, width, height). To increase the model’s robustness, the CT geometry was randomly changed in a certain range. The magnification factor was

polled between 2 and 15. For two-dimensional phantoms, the line detector pixel pitch was fixed at 0.2 mm and the pixel count was fixed at 1000. Consequently, 1000 projections were calculated uniformly over a full-circle trajectory.

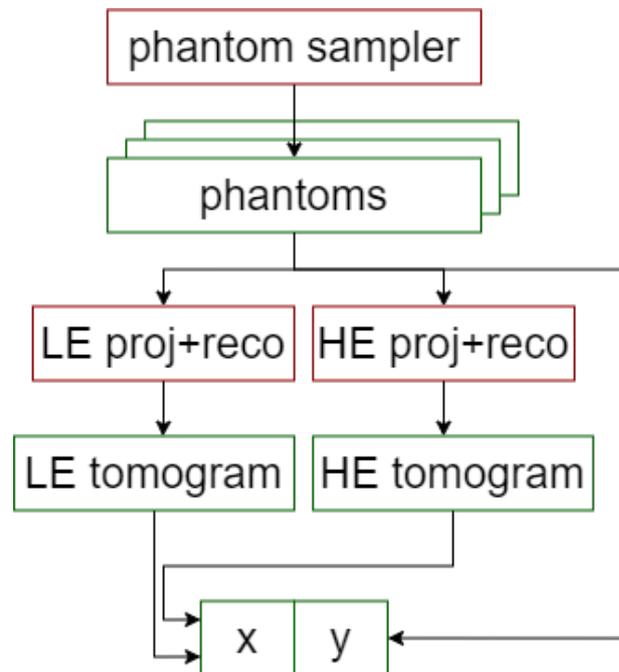


Figure 1. Pipeline overview for training data generation with computing modules in red and data objects in green. A tuple (x, y) consists of a tensor of tomograms x and a tensor of material phantoms y . The spatial dimensionality of x and y is identical, and the tomograms and phantoms are fully registered.

4.2. Data

We implemented a simulation pipeline that samples data tuples similar to the data used by Badea [5], but using more industry-relevant material combinations, such as magnesium, aluminium and iron, as well as harder X-ray spectra, i.e., above 100 kVp. Due to the heavily attenuating materials involved in the phantom and the physical dimensions of several centimetres in diameter, this minimum peak voltage is required to have sufficient photon penetration.

Figure 2 shows exemplary spectra of 100 kVp and 300 kVp together with the attenuation coefficients of magnesium and aluminium. The spectrum of 100 kVp was pre-filtered with 1 mm aluminium and the spectrum of 300 kVp was pre-filtered with 1 mm copper and 1 mm tin. We used an energy-integrating detector model, which is commonly used for industrial CT. The incident photons deposit a fraction of their total energy into the scintillator, with a probability that depends on their intersection length through the scintillator and energy. Therefore, the shaded areas indicate the photon energies with the highest contribution (top 95%) to the measured signal. Compared to material decomposition in clinical CT, where the K-edge absorption of different materials can be exploited, even the lower spectrum of 100 kVp is completely unaffected by the highest available K-edge of aluminium at around 1.56 keV. Photon energies below 50 keV are largely inconsequential for industrial CT scans of metal samples. Thereby, the measured and observable differences in attenuation coefficients between the materials are of fairly small amplitude. Using lower energy spectra, the tomogram is increasingly affected by artefacts originating from a rapidly decreasing signal-noise ratio. On the other side, higher energy spectra supply enough high-energy photons to penetrate the object, but at the cost of decreasing contrast due to the decreasing difference in linear attenuation coefficients with rising photon energy. In order to train the model successfully involving two different materials, we needed carefully balanced spectra to generate a sufficient contrast in low- and high-energy tomograms.

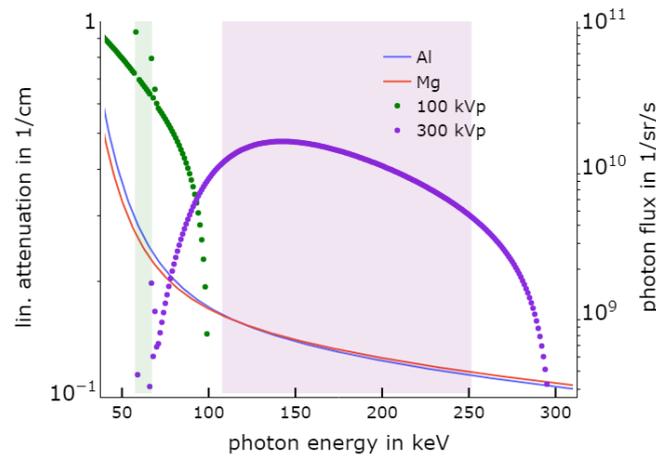


Figure 2. Comparison of spectra and absorption coefficients: 95% of the signal, measured with an energy-integrating detector, originates from the photons in the energy bins shaded below the curve assuming a 0.4 mm CsI scintillator.

Figure 3 shows a training tuple, which has been simulated with magnesium and aluminium at 80 kVp and 150 kVp. The training tuple consisted of two DECT tomograms as inputs and two material maps as outputs, which were called ground truth in the context of the training described in Section 4.3. As expected regarding Figure 2, higher attenuation coefficients occurred in the low-energy tomogram. As mentioned before, we studied two-material systems exclusively because the studied energy regime is dominated by Compton scattering, where the attenuation behaviour of a third material can be expressed as a linear combination of the other materials, resulting in a highly ill-posed prediction.

All phantoms displayed were sampled in a square and two-dimensional frame of $50 \times 50 \text{ mm}^2$. The resulting beam hardening for this configuration shown in Figure 3 was almost negligible. Nevertheless, it can be observed as a brightening around exterior edges in scans with low acceleration voltages and high atomic number constituents in the phantom. One dataset describing a copper-iron system is available for download via Zenodo [10].

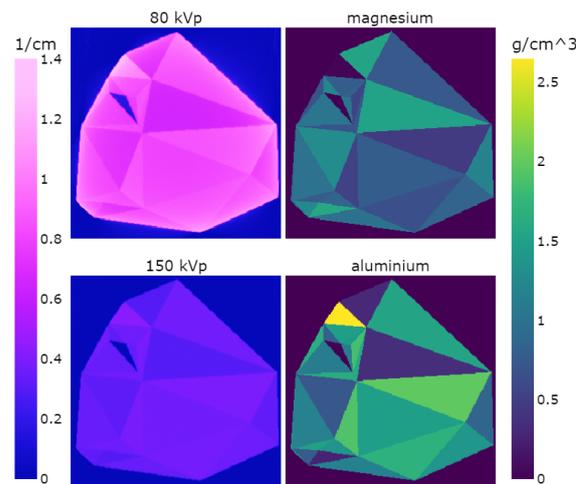


Figure 3. Data tuple used for training with a 80 kVp tomogram and a 150 kVp tomogram as inputs on the left and ground truth material distributions of magnesium and aluminium on the right. The tomograms share the colour scale on the left side. The material maps share the colour scale on the right side.

4.3. Model and Training

Regarding the model, we mainly followed the implementation of a U-Net proposed by Ronneberger [11]. Figure 4 gives an overview of the model used. In detail, the input tensor has a shape of $2 \times 256 \times 256$ (channels, rows, columns), where the first dimension represents the high- and low-energy channels, and the remaining describes the spatial dimensions. The model can be split into two main parts, the encoder and the decoder. This specific encoder uses convolutional and pooling layers to reduce the spatial dimensionality while increasing the features. The convolution is conducted with a kernel size of 3, a stride of 1 and a zero-padding of 1. In the deepest layer, the tensor consists of 128 features and has a remaining spatial dimensionality of 60×60 pixels. The decoder uses bilinear interpolation to increase the spatial dimensions, replenishes the tensor with the corresponding tensor from the encoder and convolves them in order to reduce the total feature count. We used linear interpolation operations for the upsampling layers instead of upconvolutions to reduce the number of trainable parameters and upsampling artefacts in general.

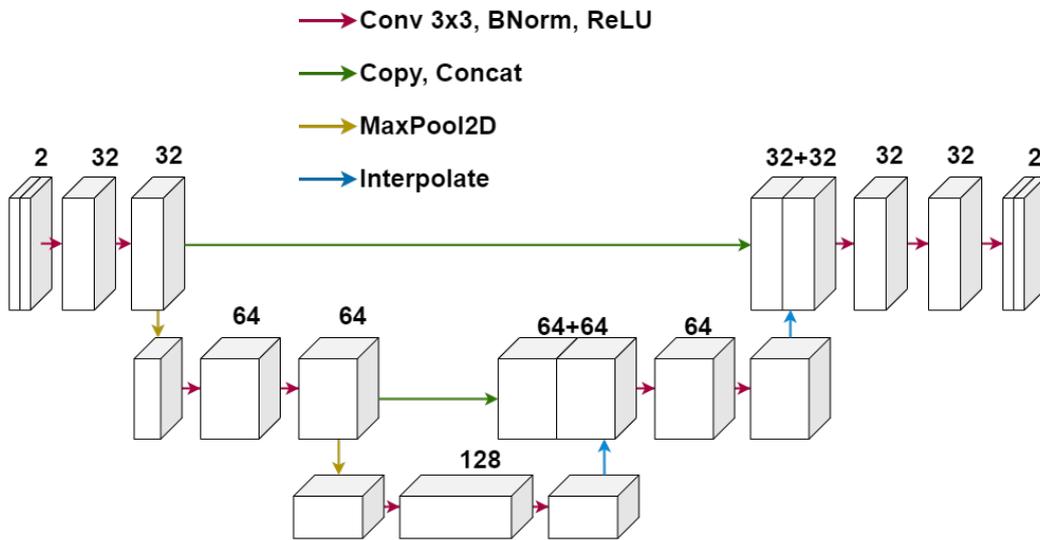


Figure 4. Model architecture mainly used for this work related to a shallow U-Net. Numbers above tensors indicate the number of features.

For training, we used the Adam optimiser together with weight decay for regularization [12]. The learning rate was initialized at $1 \cdot 10^{-4}$ and weight decay was set to $1 \cdot 10^{-3}$. The model was trained on 8000 tomogram-material-tuples for 300 epochs in about two hours using one Nvidia RTX 4090. Training with more data leads to a faster convergence, but the results are similar. Since the model does not overfit in the classical regime (see Deep Double Descent [13]), we did not have to apply a certain stopping criterion. Training was stopped when observing convergence or simply after 300 epochs, which worked in most cases.

5. Results

This section describes the material decomposition tests of simulated samples in Section 5.1 and the tests on real data in Section 5.2.

5.1. Test on Simulated CT Data

Figure 5 shows the results for aluminium (Al) and magnesium (Mg) using a low-energy spectrum at 100 kVp and a high-energy spectrum at 300 kVp. The tomograms are in the first column. Notice that in contrast to Figure 3, the high-energy tomogram appears to have higher values than the low-energy tomogram. This change is due to the normalisation of the tomograms across the whole dataset. With increasing peak acceleration voltage,

and hence photon energy, the contrast seen in the tomograms decreases. This results in a narrow distribution of values seen in the dataset compared to the low-energy dataset.

The ground truths are in the second column of the figure, the model predictions are in the third, and the absolute differences between ground truth and model predictions are in the last column.

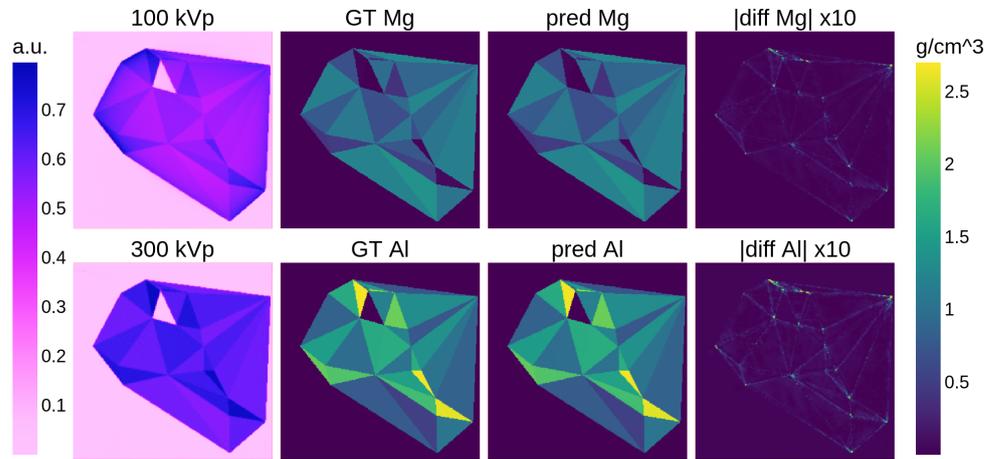


Figure 5. Qualitative comparison of the model’s prediction (pred) and the ground truth (GT). The input DECT tomograms are in the first column and normalised separately, but the colour scale is shared next to them. The absolute differences between ground truths and predictions are shown in the final column and are in units of density described by the colour scale on the right, which also relates to ground truths and predictions. The amplitude of the difference images is magnified by 10.

The highest errors appear at edges or more precise discontinuities, where the model’s upsampling linear layers are less accurate due to interpolation. Larger areas are predicted with high accuracy. Even areas with a low density of magnesium are predicted correctly, while the model does not assume mass or volume conservation at this point. A quantitative assessment of performance is provided by Figure 6. The plot was calculated using 8000 test tuples. Pixels containing neither magnesium nor aluminium have been excluded because the prediction is trivial. As indicated by the colour scale, the clear majority of all points sit close to the perfect fit where prediction equals ground truth.

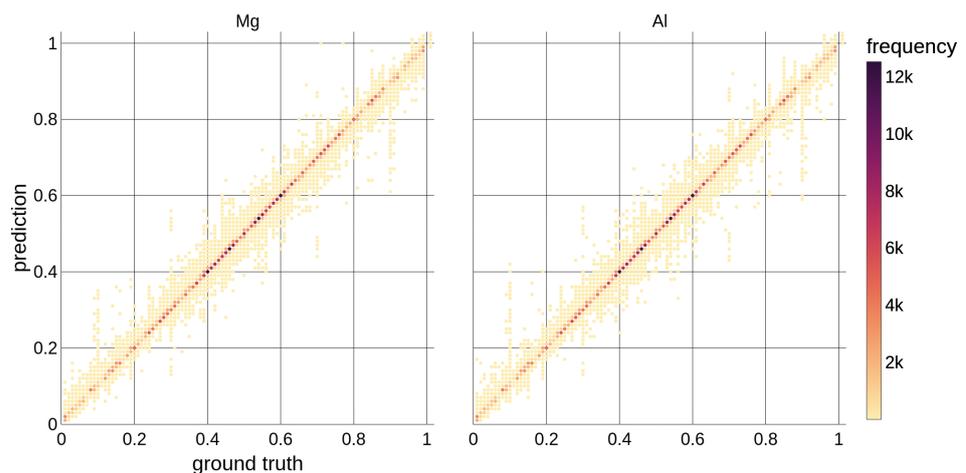


Figure 6. Pixel-by-pixel comparison of ground truth and prediction for 8000 randomly polled tuples from the test dataset. The colour scale indicates the number of points that fall in a certain bin. The prediction and ground truth are expressed in relative fraction units for a specific pixel and material. For instance, if a pixel consists of 30% magnesium, the corresponding ground truth value will be 0.3.

The mean ν and standard deviation σ of the residuals for this material system are given by Table 1 along with other material systems that have been studied. The mean residuals

for all material systems are close to zero. Mg–Al, Al–Fe and Ti64–AlSi are discriminated successfully with the DECT setup using 100 kVp and 300 kVp as indicated by the low standard deviation. For Fe–Cu and Ti64–Ti, the peak voltages were increased to 250 kVp and 450 kVp in order to penetrate the materials. A plot showing photon flux and linear attenuation for iron and copper is given by Figure A1 in the Appendix A. In the case of Fe–Cu, the material decomposition works with almost the same precision as for the other material systems with lighter elements. The decomposition of Ti64 versus Ti did not work as well since the materials are extremely similar regarding the X-ray attenuation coefficients (see Figure A2 in the Appendix A). A possibility to enhance the signal difference between titanium and its alloy Ti64 is to lower the peak voltages of the spectra used, but lowering the peak voltage is only possible when using smaller objects.

Table 1. Mean ν and standard deviation σ of the model’s residuals for different material combinations. The superscripted dagger [†] indicates the usage of higher peak voltages for the X-ray spectra to penetrate the samples sufficiently.

	Mg–Al	Al–Fe	Fe–Cu [†]	Ti64–AlSi	Ti64–Ti [†]
$\nu/10^{-4}$	−6.2	9.00	9.56	−5.5	4.8
$\sigma/10^{-3}$	9.17	9.56	10.26	9.3	52.4

Figure 7 shows the quantitative results for Ti64–Ti.

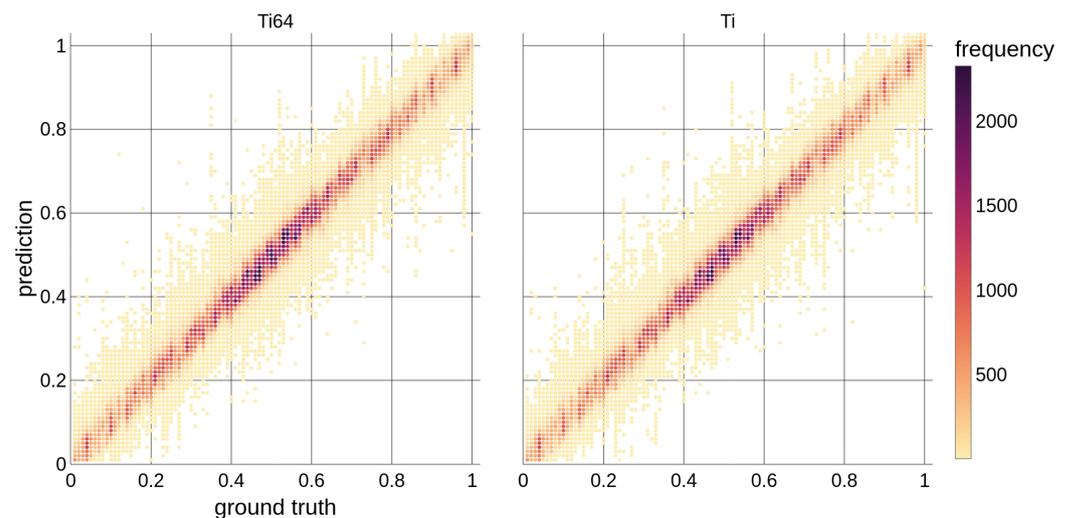


Figure 7. Pixel-by-pixel comparison of ground truth and prediction for a material consisting of Ti and Ti64. The colour scale indicates the number of points that fall in a certain bin. Note the different range described by the colour scale in comparison to Figure 6. The prediction and ground truth are expressed in relative fraction units for a specific pixel and material.

Besides the five times higher standard deviation compared to the other material systems studied above, the decomposition works qualitatively, as shown in Figure 8. The algorithm identifies the main contributing element of each triangular area correctly. As mentioned before, we did not find any comparable study to benchmark against. The dataset [10] describing a copper-iron mixture, which has been uploaded to Zenodo, will be used in future studies as a benchmark.

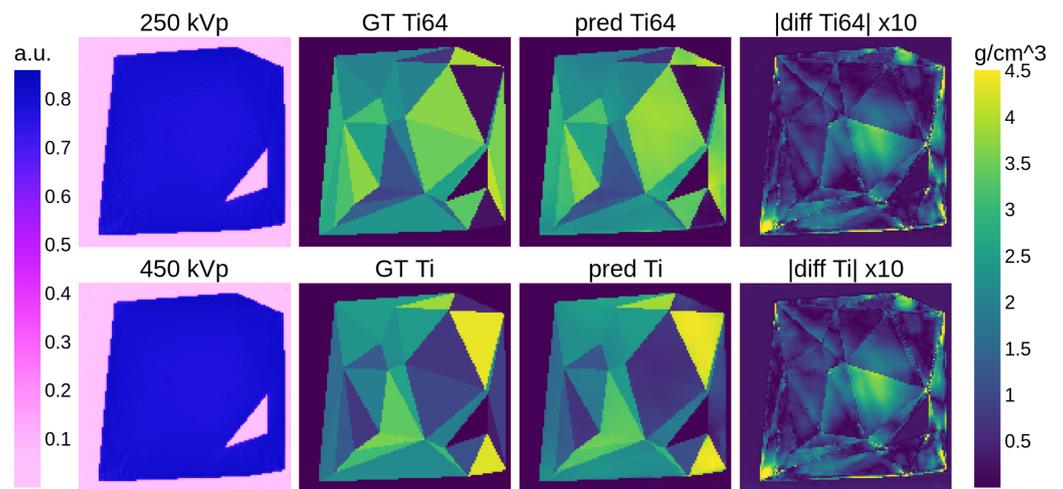


Figure 8. Qualitative comparison of the model’s prediction (pred) and the ground truth (GT). The input DECT tomograms are in the first column and normalised separately but share the colour scale next to them. The absolute differences between ground truths and predictions are shown in the final column and are in units of density described by the colour scale on the right, which also relates to ground truths and predictions. The amplitude of the difference images is magnified by 10. Note the lower image resolution compared to Figure 5 due to the high cost of computation and generally slower convergence for this material system.

5.2. Test on Real CT Data

Despite being a simulation study, it is still highly valuable to assess the model’s performance with real CT data. However, it should be noted that the simulation was not adjusted to the actual geometry of a CT scanner, its detector, or the characteristics of the source. We scanned a 10c euro coin using the CT scan setup in Table 2.

Table 2. CT scan setup and system specifications for the scan of a 10c euro coin.

CT Geometry		Source: low-energy/high-energy	
Magnification	3.5	Tube voltage	300 kVp/450 kVp
Projections	1500 full-circle	Tube Current	2 mA/1 mA
Scan time	8 min	Filter	1 mm Cu + 1 mm Sn
Detector			
Type	energy-integrating		
Scintillator	400 μm caesium iodide		
Pixel Size	278 μm		
Number of Channels	(1500, 1500)		

A 10c euro coin is made out of an alloy called Nordic gold which consists of 89% copper, 5% aluminium, 5% zinc and 1% tin. We trained a model to separate copper from a 5 + 5 + 1 aluminium, zinc and tin mixture, which is called residuals in the following, and air. Therefore, the model predicts three channels: copper, residuals and air. Figure 9 shows the low-energy, 300 kVp, tomogram, both energy tomograms along the cut line and the model’s prediction.

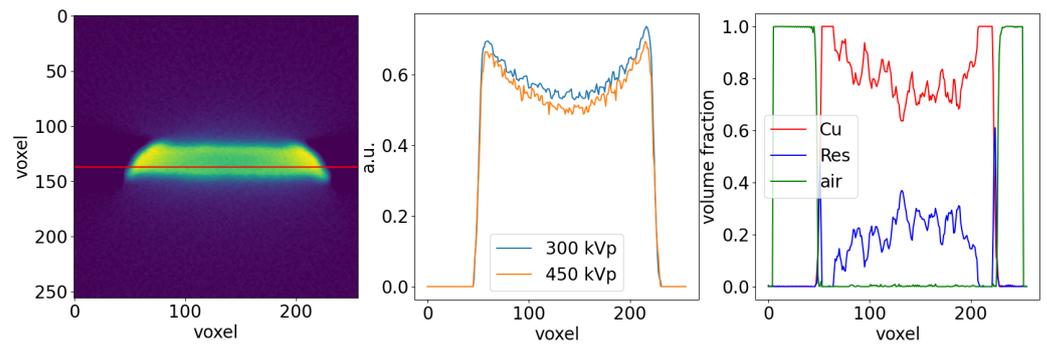


Figure 9. Overview of the 10c euro coin scan. **(Left)** low-energy tomogram with the cut line. **(Middle)** low-energy and high-energy tomograms along the cut line. **(Right)** model’s prediction on copper, residuals and air. The tomograms are normalised against the min-max range from the underlying training dataset. The model output is in units of relative volume fraction. Ideally, 89% copper and 11% residuals are expected homogenically inside the coin.

The model recognises copper as the major constituent in the mixture as well as the absence of air in the coin (see Figure 9 and Table 3). From a more quantitative point of view, the copper fraction, which should be somewhere around 89%, does fluctuate between 65% and 95% (see Figure 9), which is quite a large margin. Table 3 shows the quantitative results of the coin’s region.

Table 3. Mean ν and standard deviation σ of the model’s prediction for each base material in a real CT scan of Nordic gold calculated in the coin. The real copper fraction in Nordic gold is 89%, which is around 20% higher than our model predicted. The mean fractions in the table do not sum to 100% since the values shown are rounded.

	Copper	Residuals	Air
$\nu/10^{-2}$	69.8	30.2	0.1
$\sigma/10^{-2}$	21.7	21.7	0.3

The relative fraction between copper and the residuals predicted by the model seems to depend on the intensity of the cupping artefact, which might be the outcome of inaccurate modelling of the underlying X-ray physics. This problem might be solved by improving the simulation with a more accurate detector model and scatter radiation, which contradicts our idea behind this approach. Decreasing the sample size or removing heavily attenuating materials from the sample may increase the model’s performance. Nevertheless, samples in industrial CT scanning scenarios are manufactured using certain materials, and the exclusion of some materials or samples above certain dimensions does not seem feasible. The approach presented in this paper is limited to small samples and light materials without further investigations. One solution for developing a more flexible system might be the use of a linear accelerator with acceleration voltages up to several MeV. While decreasing the contrast between materials significantly, the increased X-ray penetration may solve some of the limitations of our current approach.

6. Outlook

As discussed, our work describes a model that has been trained on simulated datasets exclusively; hence, the next step will be the transfer to real-world CT scans. In Section 5.2, the model is tested on a real CT scan of a 10c euro coin. The results are promising from a qualitative point of view: the model identifies copper as the main constituent of the alloy called Nordic gold. Taking a closer look at the quantitative copper fractions predicted, the model is not able to handle heavy cupping artefacts, which makes it hard to use on samples comparable to a 10c euro coin in terms of material and dimensions. The use of a linear accelerator could help due to the increased X-ray penetration, but it

will lead to other problems, such as reduced material contrast and even more complex simulations due to the increasing influence of scatter radiation, which is not modelled in the current simulation. We will investigate the noise characteristics of a real CT scan compared to the currently implemented noise model in the simulation and evaluate as well as improve the simulation of beam hardening effects. Obviously, when doing the transfer to a real CT scanner, machine uncertainties, for example, mechanical displacements and fluctuating properties of the X-ray spectra, have to be taken into account. In order to compare all these effects quantitatively, a known sample is needed to conduct CT scans and simulations simultaneously. Knowing the different behaviours of the simulation and the real CT scan, two different approaches are possible: The obvious way is to increase the quality of the simulation until it becomes indistinguishable from real CT scans, which is not economical. Monte-Carlo-based X-ray simulations can produce outstanding results, but they are extremely expensive to run. An alternative to overcome the sim-to-real gap can be an adapted training strategy, which has been studied intensively in the past in other contexts. Pre-training the model followed by a subsequent fine-tuning on real-world data is referred to as transfer learning and has been studied for a segmentation task using a U-Net on clinical ultrasonic and X-ray datasets by Amiri et al. [14].

Another parameter which will be crucial for industrial applications is the physical size of the sample to be scanned. As discussed in Section 5, larger samples are more difficult to process since the X-ray attenuation coefficients of different materials narrow in spread with increasing photon energy. Additionally, different physical samples will be scanned and published to have a benchmark for future works with respect to sample size, material and machine effects.

Regarding the model architecture itself, the next step could be the implementation of a three-dimensional U-Net to process volumetric data from tomograms. Additionally, the material decomposition for more than two materials is highly ill-posed. For this reason, the model should state some kind of measure for the uncertainty of the prediction. Bayesian models could handle these uncertainties by using probability distributions instead of explicit weights inside the model.

Author Contributions: Conceptualization, M.W., N.B., M.v.S. and T.M.; methodology, M.W., M.v.S. and T.M.; software, M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used for training and evaluation can be downloaded via Zenodo [10].

Conflicts of Interest: Authors Moritz Weiss, Nick Brierley and Mirko von Schmid were employed by the company diondo GmbH. Moritz Weiss is conducting his PhD research at diondo GmbH as a cooperation with the University of Wuppertal. The remaining author, Tobias Meisen, declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Appendix A

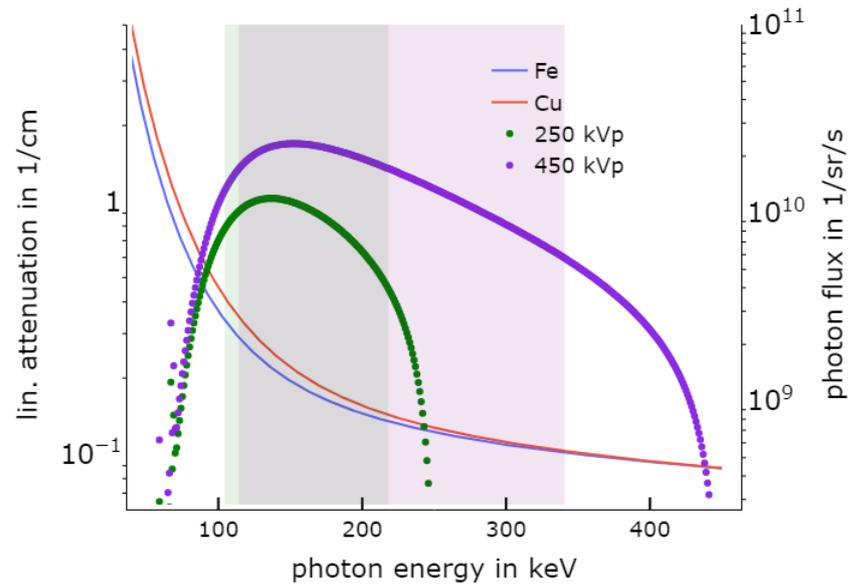


Figure A1. Comparison of spectra and absorption coefficients. 95% of the signal, measured with an energy-integrating detector, originates from the photons in the energy bins shaded below the curve assuming a 0.4 mm CsI scintillator.

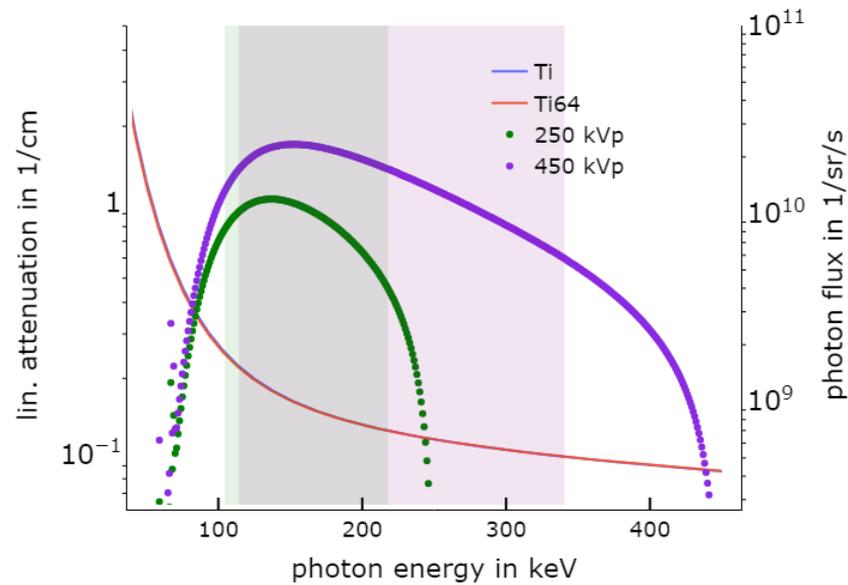


Figure A2. Comparison of spectra and absorption coefficients. 95% of the signal, measured with an energy-integrating detector, originates from the photons in the energy bins shaded below the curve assuming a 0.4 mm CsI scintillator.

References

1. Heismann, B.J.; Leppert, J.; Stierstorfer, K. Density and atomic number measurements with spectral x-ray attenuation method. *J. Appl. Phys.* **2003**, *94*, 2073–2079. . . [CrossRef]
2. Radon, J. Über die Bestimmung von Funktionen längs gewisser Mannigfaltigkeiten. *Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig* **1917**, *69*, 262–277.
3. Feldkamp, L.A.; Davis, L.C.; Kress, J.W. Practical cone-beam algorithm. *J. Opt. Soc. Am. A* **1984**, *1*, 612–619. . [CrossRef]
4. Fang, W.; Li, L. Attenuation image referenced (AIR) effective atom number image calculation for MeV dual-energy container CT using image-domain deep learning framework. *Results Phys.* **2022**, *35*, 105406. . [CrossRef]

5. Clark, D.P.; Holbrook, M.; Badea, C.T. Multi-energy CT decomposition using convolutional neural networks. In *Medical Imaging 2018: Physics of Medical Imaging*; Chen, G.-H., Lo, J.Y., Gilat Schmidt, T., Eds.; SPIE: Houston, TX, USA, 2018. Available online: <https://www.spiedigitallibrary.org/conference-proceedings-of-spie/10573/2293728/Multi-energy-CT-decomposition-using-convolutional-neural-networks/10.1117/12.2293728.full> (accessed on 21 August 2023).
6. MathWorks. Delaunay Triangulation. Available online: <https://www.mathworks.com/help/matlab/ref/delaunay.html> (accessed on 21 August 2023).
7. Abascal, J.F.; Ducros, N.; Pronina, V.; Rit, S.; Rodesch, P.A.; Broussaud, T.; Bussod, S.; Douek, P.C.; Hauptmann, A.; Arridge, S.; et al. Material Decomposition in Spectral CT using deep learning: A Sim2Real transfer approach. *IEEE Access* **2021**, *9*, 25632–25647. . [[CrossRef](#)]
8. Van Aarle, W.; Palenstijn, W.J.; Cant, J.; Janssens, E.; Bleichrodt, F.; Dabrovolski, A.; De Beenhouwer, J.; Batenburg, K.J.; Sijbers, J. Fast and flexible X-ray tomography using the astra toolbox. *Opt. Express* **2016**, *24*, 25129–25147. . [[CrossRef](#)] [[PubMed](#)]
9. Van Aarle, W.; Palenstijn, W.J.; De Beenhouwer, J.; Altantzis, T.; Bals, S.; Batenburg, K.J.; Sijbers, J. The astra toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy* **2015**, *157*, 35–47. . [[CrossRef](#)] [[PubMed](#)]
10. Weiß, M. Simulated industrial CT dataset for deep learning with dual-energy tomograms and ground truth material maps for copper and iron. *Zenodo* **2022**. . [[CrossRef](#)]
11. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. *arXiv* **2015**, arXiv:1505.04597.
12. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2017**, arXiv:1412.6980.
13. Nakkiran, P.; Kaplun, G.; Bansal, Y.; Yang, T.; Barak, B.; Sutskever, I. Deep Double Descent: Where Bigger Models and More Data Hurt. *arXiv* **2021**, arXiv:1912.02292.
14. Amiri, M.; Brooks, R.; Rivaz, H. Fine-Tuning U-Net for Ultrasound Image Segmentation: Different Layers, Different Outcomes. *IEEE Trans. Ultrason. Ferroelectr. Freq. Control* **2020**, *67*, 2510–2518. . [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.