

Quantitative Comparison of Planar Coded Aperture Imaging Reconstruction Methods

Meißner, T.^{a,b,1} Rozhkov, V.^c Hesser, J.^{a,d,e,f} Nahm, W.^b Loew, N.^a

^aMannheim Institute for Intelligent Systems in Medicine (MIISM), Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

^bInstitute of Biomedical Engineering (IBT), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

^cInternational Intergovernmental Organization Joint Institute for Nuclear Research, Dubna, Russia

^dInterdisciplinary Center for Scientific Computing (IWR), Heidelberg University, Heidelberg, Germany

^eCentral Institute for Computer Engineering (ZITI), Heidelberg University, Heidelberg, Germany

^fCZS Heidelberg Center for Model-Based AI, Heidelberg University, Mannheim, Germany

E-mail: tobias.meissner@medma.uni-heidelberg.de

ABSTRACT: Imaging distributions of radioactive sources plays a substantial role in nuclear medicine as well as in monitoring nuclear waste and its deposit. Coded Aperture Imaging has been proposed as an alternative to parallel or pinhole collimators, but requires image reconstruction as an extra step. Multiple reconstruction methods with varying run time and computational complexity have been proposed. Yet, no quantitative comparison between the different reconstruction methods has been carried out so far. This paper focuses on a comparison based on three sets of hot-rod phantom images captured with an experimental γ -camera consisting of a Tungsten-based MURA mask with a 2 mm thick 256×256 pixelated CdTe semiconductor detector coupled to a Timepix[®] readout circuit. Analytical reconstruction methods, MURA Decoding, Wiener Filter and a convolutional Maximum Likelihood Expectation Maximization (MLEM) algorithm were compared to data-driven Convolutional Encoder-Decoder (CED) approaches. The comparison is based on the contrast-to-noise ratio as it has been previously used to assess reconstruction quality. For the given set-up, MURA Decoding, the most commonly used CAI reconstruction method, provides robust reconstructions despite the assumption of a linear model. For single image reconstruction, however, MLEM performed best among analytical reconstruction methods, but took the longest with an average of 13 s run time. The fastest reconstruction method is the Wiener Filter with 67 ms and mediocre quality. The CED with a specifically tailored training set was able to succeed the most commonly used MURA decoding on average by a factor between 1.37 and 2.60 and a run time of around 300 ms.

KEYWORDS: Gamma camera, SPECT, PET PET/CT, coronary CT angiography (CTA); Image reconstruction in medical imaging; Medical-image reconstruction methods and algorithms, computer-aided diagnosis; Search for radioactive and fissile materials

¹Corresponding author.

Contents

1	Introduction	1
2	State of the art	3
2.1	Mathematical description	3
2.2	Reconstruction methods	4
2.2.1	MURA Decoding	4
2.2.2	Wiener Filter	4
2.2.3	Maximum Likelihood Expectation Maximization	5
2.2.4	Convolutional Encoder-Decoder	5
3	Material & methods	6
3.1	Experimental measurements	6
3.2	Generation of ground truth data	6
3.3	Analytical reconstruction methods	7
3.4	Data-driven reconstruction methods	8
3.4.1	Generation of training data	8
3.4.2	Architecture and training	10
3.5	Contrast-to-noise ratio	11
4	Results	12
4.1	Hot-rod phantom data	12
4.2	Single image reconstruction	12
4.3	Dual image reconstruction	15
5	Discussion	16
5.1	Hot-rod phantom data	16
5.2	Single image reconstruction	16
5.3	Dual image reconstruction	18
6	Conclusion & outlook	19

1 Introduction

Accurate localization and comprehensible visualization of radioactive source distributions constitutes the essence of nuclear medicine [18, 32, 37, 42, 44, 52, 53], nuclear waste monitoring [15, 19, 21, 50, 55] and x-ray fluorescence microscopy [9, 34]. When high-energy photons are involved, refractive or reflective camera systems are not realizable due to the negative refraction index or demand expensive manufacturing processes due to the high surface quality for x-ray mirrors.

Thus, modern γ -cameras are equipped with pinhole or parallel collimators to acquire the spatial information of incident rays. This kind of collimation represents a trade-off between resolution and photon efficiency: The smaller the openings or the higher the aspect ratio, the sharper the image, but the less photons pass through the collimator [35, 42, 52]. Due to this limitation, Coded Aperture Imaging (CAI) has been discussed, because it persuades a better compromise between resolution and photon efficiency. Proposed independently by Ables [2] and Dicke [16] for X- and γ -ray astronomy (see [12, 14] for further information), a mask between object and detector consisting of a radiopaque material with transparent elements in between, encodes direction information of incoming γ -rays. Since each transparent element of the mask acts as a pinhole and thus generates a projection of the object on the detector, the resulting image consists of a multitude of overlapping images and becomes incomprehensible. Therefore, image reconstruction is required to obtain an interpretable image.

CAI has been investigated as alternative collimator in single photon emission computed tomography (SPECT): Experiments for imaging the bio-distribution of radioactively labeled compounds in small animals, have been carried out [4, 6]. [18, 30, 47] analyzed the use of a coded aperture γ -camera for the localization of sentinel lymph nodes and [38, 39] examined the capability of imaging cold-spot lesions in cardiac imaging. Other research fields like x-ray fluorescence spectroscopy [34] or nuclear decommissioning [15, 21] analyzed the potential use of CAI as well.

When the object's longitudinal extension is small in comparison to the camera-object distance, one can assume that all radioactive sources lay within one fixed plane parallel to the detector. Subsequently, CAI can be regarded as an image-to-image mapping and is referred to as planar CAI. Several reconstruction methods for planar CAI have been proposed but yet, no comprehensive quantitative comparison has been carried out to assess their advantages and disadvantages.

The main contribution of this paper is a thorough and quantitative comparison of the most commonly used planar coded aperture reconstruction methods. Future research in various fields of nuclear imaging could benefit from this assessment in terms of run time and reconstruction quality. An additional contribution of this paper is a data set of high-resolution coded aperture images from three different hot-rod phantoms, acquired with an experimental γ -camera. Unlike before, this paper shows on experimental data that a Deep Learning approach is capable of delivering better reconstructions than state-of-the-art methods and proved worthy of further investigation. Furthermore, to the best of our knowledge, this paper includes the first re-implementation and application on experimental data of the convolution-based Maximum Likelihood Expectation Maximization algorithm proposed by [39].

The structure of this paper is as follows: First, section 2 gives an overview of the mathematical description of CAI and the most commonly used reconstruction methods. Then, section 3 describes the image acquisition and implementation of both *analytical* and *data-driven reconstruction methods*. Section 4 presents the reconstruction results in terms of reconstruction quality and run time, which are discussed in section 5. Finally, section 6 concludes this paper's main findings and provides an outlook on possible future research.

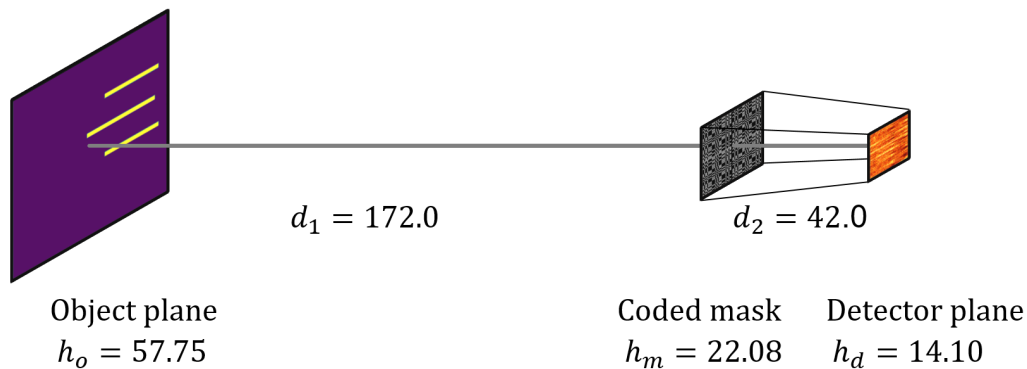


Figure 1: A sketch of the experimental γ -camera set-up, with all values given in millimeter. Planar Coded Aperture Imaging consists of three parallel planes: The object plane (left), the coded mask (middle), which encodes the object image and the position-sensitive detector (right). The gray line represent the central line. The coded mask and the detector form the γ -camera itself.

2 State of the art

2.1 Mathematical description

Let Φ be a two-dimensional radioactive source distribution on a plane parallel to the detector plane in a constant and defined distance. The detector image of each point source of Φ results in a shifted shadow of the coded mask pattern on the detector. As any distribution of sources can be described by a linear combination of point sources, planar CAI can be described as the convolution of an object image $f(x, y)$ with the single coded mask shadow $h(x, y)$ as long as the distance object - mask can be considered to be much larger than the diameter of the mask and as long as scatter at the mask can be neglected [16]. The resulting detector image is referred to as $p(x, y)$. In reality, this process is not deterministic, but especially in imaging of weakly radiating sources bound to the random Poisson process \mathfrak{P} . Additional detector noise $n(x, y)$ must be taken into account, resulting in the following system equation:

$$p(x, y) = \mathfrak{P}\{f(x, y) * h(x, y)\} + n(x, y) \quad (2.1)$$

Multiple families of coded aperture patterns can be found in literature (see [6] for an overview), but most commonly used are Modified Uniformly Redundant Arrays (MURA), based on the mathematical concept of quadratic residues [17, 23]. The pattern can be expressed by a binary matrix, where a “1” denotes a pinhole and “0” opaque material:

$$H_{ij} = \begin{cases} 1 & \text{if } i+j=0, \\ 1 & \text{if } j = 0, i \neq 0, \\ 1 & \text{if } C_i C_j = 1, \\ 0 & \text{otherwise} \end{cases} \quad \text{with} \quad C_k = \begin{cases} 1 & \text{if } k \text{ is a quadratic residue modulo } L, \\ -1 & \text{otherwise} \end{cases} \quad (2.2)$$

and zero-based indices i and j referring to rows and columns. L is called the rank and must be both prime and of the following form: $L = 4m + 1$ with m as natural number. MURA masks reach

a fraction of opaque to transparent cells of approximately 50% [23]. To yield a self-supporting mask, rows and columns of zeros can be inserted between all columns and rows. This so-called no-two-holes-touching (NTHT) mask is easier to fabricate but has the disadvantage of reducing the transparent fraction of this mask to around 12.5%. The MURA pattern used in this paper can be seen in figure 3. Commonly, a mosaic of four identical basic patterns is used to assure a complete mask shadow on the detector, while maintaining a wide field-of-view (FoV). The shadow is complete in the sense of circular convolution, which enables decoding only using the decoding pattern of the basic pattern and the central part of the detector [17]. With evolving computing power, a decoding using the complete mosaicked pattern and the whole detector image does not pose a problem anymore and reduces background noise [39].

It has been proposed to combine two single captures into a dual image acquisition, where one image is taken with the mask and a second image with the anti-mask to remove near-field artifacts [6, 39, 54]. The anti-mask is the inverse coded aperture mask, which has pinholes where the mask is opaque and vice versa. Certain MURA patterns inhibit the anti-mask in themselves and can be achieved by rotating the mask by 90°. This makes dual acquisition easy to implement as no additional mask needs to be manufactured or installed. The reconstructions of the dual images are referred to as *dual image reconstructions*.

2.2 Reconstruction methods

Mainly four methods for real-time capable image reconstruction have been proposed within the last few decades. MURA Decoding (also called inverse filtering, or cross-correlation analysis): [3, 4, 6, 18, 20, 21, 30, 31, 47, 48, 54], Wiener Filtering [24, 34], convolutional Maximum Likelihood Expectation Maximization (MLEM) reconstruction [38, 39] and data-driven Deep Learning approaches [34, 56, 57]. Although the standard MLEM algorithm has also been investigated [28, 36], it is not considered in this paper, because its reconstruction time of several hours makes it impractical for the application in mobile systems.

2.2.1 MURA Decoding

For each MURA mask pattern an inverse decoding pattern G can be determined [17]. Correlation of the decoding and encoding pattern results in a δ -distribution and can be regarded as an inverse filter [10]. The decoding pattern G in relation to the encoding pattern H equals

$$G_{ij} = \begin{cases} 1 & \text{if } i + j = 0, \\ 1 & \text{if } H_{ij} = 1, i + j \neq 0, \\ -1 & \text{if } H_{ij} = 0, i + j \neq 0. \end{cases} \quad (2.3)$$

In practical terms that means all 0 change to -1 and all 1 stay 1 except for the central pixel if a 2×2 mosaicked mask is chosen [15]. Convoluting the given detector image with the decoding pattern and central cropping yields the reconstructed object image.

2.2.2 Wiener Filter

The Wiener Filter executes reconstruction in an optimal balance between inverse filtering and noise smoothing with regards to the mean squared error. The underlying assumptions are a linear system

and Gaussian additive white noise [10]. It can be regarded as a weighted cross correlation with the mask pattern and has the design parameter signal-to-noise-ratio (SNR). In the Fourier domain with ν as frequency vector, the Wiener Filter reads as follows:

$$W(\nu) = \frac{1}{H(\nu)} \frac{|H(\nu)|^2}{|H(\nu)|^2 + \frac{1}{\text{SNR}(\nu)}} \quad (2.4)$$

Unlike MURA Decoding, the reconstruction turns into a multiplication in the Fourier space and thus offers fast reconstruction. However, both methods ignore the random nature of photon emission and detection in nuclear imaging [39].

2.2.3 Maximum Likelihood Expectation Maximization

The Maximum Likelihood Expectation Maximization (MLEM) algorithm is an iterative algorithm that estimates the object image with the highest likelihood for the measured detector image. The algorithm is derived from the assumption that the photon counts follow a Poisson process [13]. Originally, probabilistic system knowledge in the form of a transfer matrix A must be provided, where its entries a_{ij} denote the probability that a photon emitted from object pixel j is detected in image pixel i . For high-resolution imaging, A would contain billions of entries, rendering the standard MLEM computationally expensive. To overcome this major drawback, Mu & Hoang [39] proposed a convolutional-based version of the MLEM algorithm: Instead of using the system matrix A for forward- and back-projection, the mask pattern and the convolution operator is used.

$$\hat{f}^{k+1}(x, y) = \hat{f}^k(x, y) \odot \left[\frac{p(x, y)}{\hat{f}^k(x, y) * h(x, y)} \otimes h(x, y) \right] \quad (2.5)$$

with \odot and \otimes denoting element-wise multiplication and the correlation operator. The algorithm consists of two major steps: forward-projection and back-projection. The convolution in the denominator represents the forward-projection step using current estimate \hat{f}^k . Dividing the detector image p by the current forward-projection yields the relative difference. The correlation of the relative difference with the mask pattern h represents the back-projection, which is multiplied element-wise with the current estimate \hat{f}^k to obtain an updated estimation \hat{f}^{k+1} .

2.2.4 Convolutional Encoder-Decoder

A Convolutional Encoder-Decoder (CED) is a widely used architecture for Convolutional Neural Networks (CNN). A CED consists of trainable parameters in the form of multiple layers of convolutional kernels that successively transform an input into an output image. The transformation is not derived from a mathematical description of the imaging system. Instead, by providing a sufficient amount of training pairs an approximate transformation from the input domain into the output domain is deduced. Recent advances in Deep Learning in the field of image reconstruction [9, 25, 51, 59] underline the potential of CEDs for CAI reconstruction. By compressing the input image via strided convolution into a latent space representation from which the output image is generated by progressive up-sampling, CEDs effectively regularize the ill-posed inverse mapping by implicit denoising [9]. Thus, CNNs and CEDs in particular, have been deployed successfully on direct reconstruction of positron emission tomography scans from sonograms [25], Radon inversion [26], metal artifact reduction in computer tomography [58] or image deblurring [51], to mention a

few examples.

First experiments have been conducted on the application of CNNs to CAI reconstruction, but they either were on purely simulated and low-resolution images [56, 57] or compared only visually to other methods [34]. In the following, the above described reconstruction methods are referred to as *analytical reconstruction methods*, while trained CEDs are denoted by *data-driven reconstruction methods*.

3 Material & methods

3.1 Experimental measurements

Three different phantoms were designed and manufactured: A spatial resolution phantom (SRP), a linear resolution phantom (LRP) and a contrast phantom (CP). All three phantoms have the basic form of a cylinder with a height of 80 mm and 50 mm in diameter, where tubes along the vertical axis were filled with ^{99m}Tc , the most widely used radionuclide in nuclear medicine [42]. The total activity was 83 MBq, 50 MBq and 75 MBq for SRP, LRP and CP respectively. Phantoms were designed with a computer-aided-engineering software and afterwards milled out of Lucite. A depiction can be seen in figure 2.

The SRP has three tubes with a diameter of 1.1 mm, two of which are 15 mm and one 20 mm long. The tubes are arranged in parallel at three different positions. The LRP has eleven tubes in a straight line in radial direction. All tubes are 20 mm long and have a diameter of 1.1 mm, even though only every second tube was filled with ^{99m}Tc . The CP consists of two larger tubes, which are 25 mm long and 5 mm in diameter. They are centered around the vertical axis.

The experimental γ -camera for capturing the phantoms has already been described in [45] and only the main characteristics are repeated here and are depicted in figure 1: A 2×2 mosaicked, 1 mm thick Tungsten NTHM MURA mask of rank 31 with pinholes of 0.34 mm in diameter was placed 42 mm in front of a 2 mm thick 256×256 pixelated CdTe semiconductor detector coupled to a Timepix[®] readout circuit. The virtual object plane is 172 mm in front of the mask plane, resulting in a FoV of 57.75×57.75 mm. The mask is anti-symmetrical and can be rotated by 90° to form the anti-mask. Each phantom was placed on a rotational table in front of the camera. The detector was exposed for 2 min and afterwards the phantom was rotated along its vertical axis by 3° . One complete rotation per hot-rod phantom resulted in 120 images per phantom. Each phantom was captured once with mask and once with anti-mask and thus a total amount of 720 images were acquired. Detected photons outside the energy window of 10 keV centered at the photon peak of 140 keV for ^{99m}Tc were discarded.

To remove erroneous pixels, a global thresholding followed by pixel replacement was carried out. The average image of all 720 captures, was thresholded according to the 2nd and the 98th percentile. Pixels outside this range were considered outliers and replaced by the median of their 3×3 neighborhood.

3.2 Generation of ground truth data

In order to compare reconstruction methods, the perfect reconstruction (the so-called ground truth) is essential. Therefore, the phantom models were loaded into a visualization program (Paraview

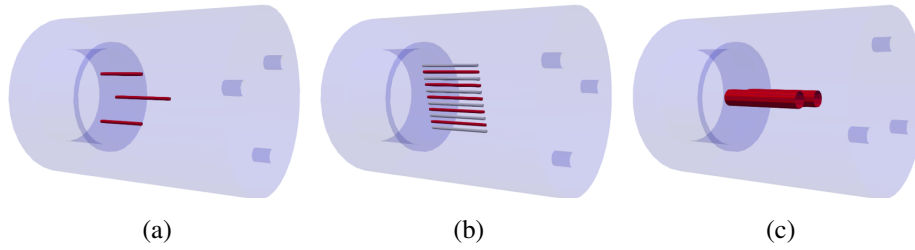


Figure 2: Three hot-rod phantoms: (a) spatial resolution phantom (SRP), (b) linear resolution phantom (LRP), (c) contrast phantom (CP). Marked in red are the tubes filled with ^{99m}Tc . The indentations on both sides were used to clamp the phantoms.

Version 5.9.1 [7]), where all surfaces except for the rods were set transparent. The point of view, i.e. camera position, focal point and view angle were set according to the experimental set-up. The background and the hot-rod color was set to black and white and the camera resolution to 256×256 . By using automated animation, the phantom was rotated along the horizontal axis and at each position the rendered image was stored as image file. To obtain fully binarized images and remove shadows, those images were cleaned by applying a simple thresholding. The results are binary images, which are 1 where γ -sources are expected and 0 elsewhere. The images were used as ground truth data for both the single and dual image reconstruction.

3.3 Analytical reconstruction methods

For all reconstruction methods, not just the central part, but as suggested by [39] the complete detector image and mask pattern was used for reconstruction. This allowed usage of the entire detector and FoV. The reconstructions of each method were finally normalized to the range of $[0, 1]$ to assure a fair comparison. Even though all reconstruction methods were intended for dual image reconstruction, the application to the single images was also analyzed.

The inference time per method was measured by averaging the elapsed time for the reconstruction of one image over 1,000 runs on the CPU of a computer with a 6-core Intel Core i7 processor (2.6 GHz) with six cores and 16 GB of RAM. TensorFlow allows the use of a GPU to speed up matrix and vector calculations, thus the inference time with GPU support was also evaluated with a Nvidia GeForce RTX 2070 and 8 GB dedicated RAM. All processing and implementation was carried out in TensorFlow 2.7 [1] and Python 3.9.7.

MURA Decoding: Two different decoding patterns were constructed. A two-holes-touching (THT) and a not-two-holes-touching (NTHT) pattern based on the basic MURA pattern of rank 31 from eq. 2.2. Both have been used as base pattern for reconstruction in literature [3, 24]. After converting the mosaicked pattern into the decoding pattern by eq. 2.3, it was resized to 512×512 pixels by nearest-neighbor interpolation to assure a binary final decoding pattern. For the reconstruction of single acquisition images, the decoding pattern can directly be convolved with the detector image to obtain the reconstructed image. For the reconstruction using the dual image, both images are reconstructed separately and summed up afterwards. However, due to the linear operation and the relationship from mask and anti-mask, it is equivalent to applying it once to the subtraction of both single images [54].

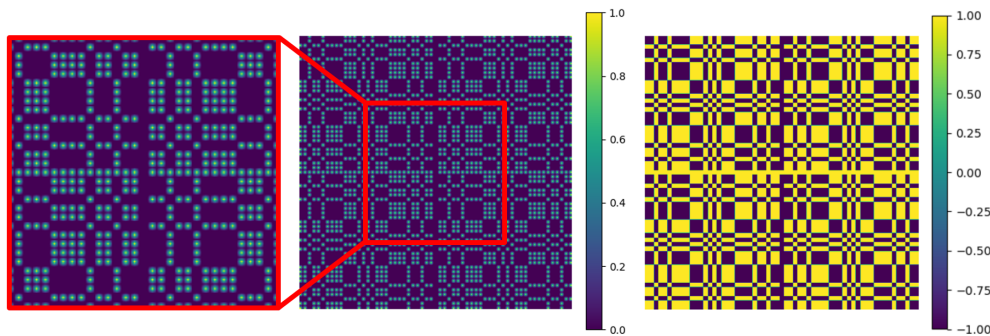


Figure 3: From left to right: The centralized basic MURA pattern (rank 31), its 2×2 arrangement with bilinear resizing for simulated forward-projection and the according THT decoding pattern for MURA Decoding.

Wiener Filter: The Wiener Filter was calculated according to eq. 2.4 with various SNR. Analogously to MURA Decoding, two types of mask pattern were used as base pattern for the Wiener Filter: The THT and the NTHT version. In order to avoid aliasing, the detector image was periodically padded to 512×512 pixels and transformed into Fourier domain, similar to [24]. After element-wise multiplication with the Wiener Filter, the result was transformed back into the spatial domain, where the central 256×256 pixels contained the final reconstructed image. A line search was carried out and $\text{SNR} = 10^{-7}$ was chosen and used for all succeeding analyses. Dual image reconstruction was carried out analogously to MURA Decoding.

MLEM: The algorithm according to eq. 2.5 was implemented with an additional small ε of 10^{-7} added to the denominator to avoid zero divisions. Both the THT and the NTHT version of the MURA pattern were tested for the MLEM algorithm. Instead of applying adaptive stopping criteria, a fixed number of 25 iterations was chosen heuristically, which confirmed to be a good compromise between reconstruction quality and noise amplification. The initial guess was set to a constant image of 0.5. Unlike MURA Decoding and the Wiener Filter, the MLEM algorithm is a non-linear operation. Thus, both single acquisition images were reconstructed independently from each other, with their corresponding mask and anti-mask. Afterwards, the final reconstructed image is obtained by averaging both reconstructions, as suggested by [39].

3.4 Data-driven reconstruction methods

3.4.1 Generation of training data

What separates the Machine Learning approach from analytical methods is the need for a large quantity of image pairs for training the network, consisting of an input image (the detector image) and a target image (the reconstructed image). A common way to solve the data problem is by using simulated training data [25, 34, 57]: In this paper, two different sets of training data were generated: One based on the large collection of natural photographs called *ImageNet* [46] and one based on horizontal lines on dark background: For the first set of training data, 50,000 images from the ImageNet’s validation set were cropped to 256×256 pixels and normalized to the range of 0 to 1. Those images posed as object images. For the second set of training data, 50,000 images with black background and different number of horizontal lines with varying thickness were created using the

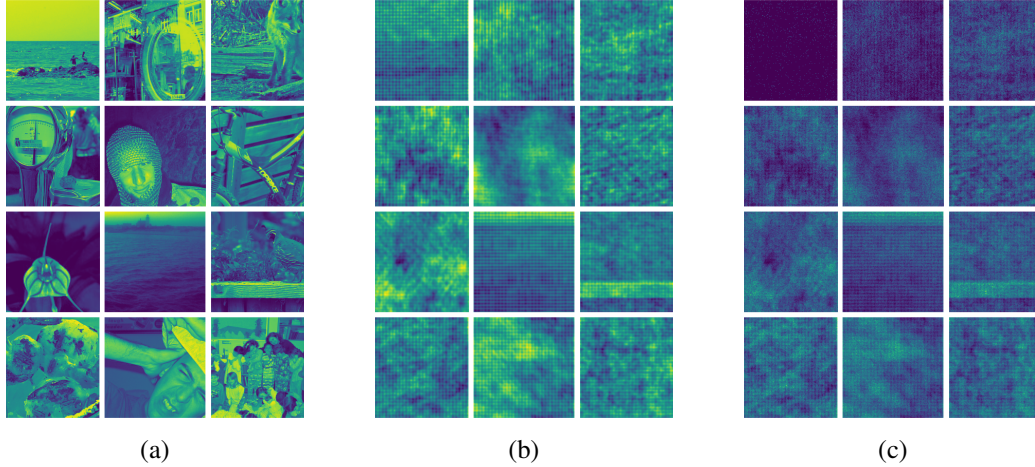


Figure 4: (a) shows exemplary target images from the ImageNet based training set, (b) their forward-projection and (c) the final training input image after Poisson randomization with an increasing photon count from top left to bottom right.

openCV library version 3.4.2 [11]. The number of lines varied between 2 and 6, with a height and width between 4 and 24 pixels and between 60 and 122 pixels, respectively. The center points were chosen randomly over the whole image. Each line has a uniform intensity between 0 and 1 and the intensities of overlapping lines were added, to ensure a non-binary intensity distribution. The final image was normalized.

The forward-projection, which results in the simulated detector image, was carried out according to eq. 2.1. The NTHT MURA mask pattern of the experimental γ -camera was generated. Similar to [34] the binary pattern needed to be rescaled with respect to the object-mask and mask-detector distances where two conditions needed to be fulfilled: First, the final mask pattern is required to be four times the size of the detector in pixels, so that the basic MURA pattern covers the entire detector. Second, the illuminated circle produced by one pinhole must match the magnified pinhole size: The illuminated circle in pixel was calculated according to the pinhole diameter (0.34 mm), the system’s magnification factor (1.244) and the detector pixel size (0.0551 mm/pixel). Bilinear resizing of the NTHT MURA mask to 512×512 pixels fulfilled both conditions and yielded the final mask array. The obtained matrix can be regarded as the complete point spread function (PSF) of the γ -camera and is depicted in the middle of figure 3. By convolving an object image with this complete PSF and afterwards cropping the central 256×256 pixels, one yields the detector image of the given object image. The additive noise $n(x, y)$ was set to be zero, assuming a ultra-low noise detector with single photon detection.

Different levels of Poisson noise were simulated: The sum of all pixel values, corresponding to the amount of photons the detector has detected, was scaled to reach a random number between the range of 1,000 and 1,000,000 photons. This scaled image represents the *expected average photon count per pixel* and formed the basis for a random draw from a Poisson distribution \mathfrak{P} .

For training the dual image reconstruction CED, the same forward-projection was carried out with the anti-mask as well and concatenated as a second color channel. Finally, the simulated detector image was normalized to the range of 0 to 1 for each channel. Exemplary images with their

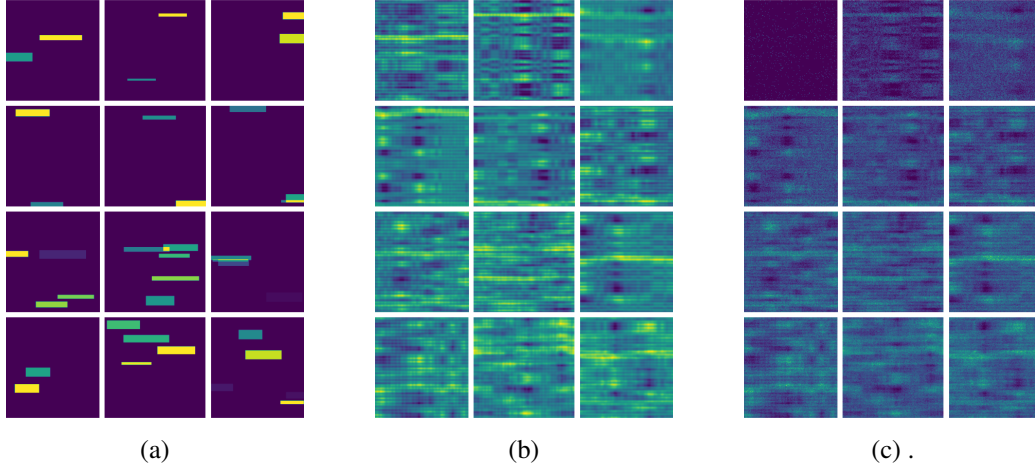


Figure 5: (a) shows exemplary target images from the horizontal line training set, (b) their forward-projection and (c) the final training input image after Poisson randomization with an increasing photon count from top left to bottom right.

simulated forward-projection with and without Poisson randomization from the ImageNet data set and for the horizontal line data set can be seen in figure 4 and figure 5. The generation of training data was done on-the-fly while training the CEDs.

3.4.2 Architecture and training

Figure 6 presents the Convolutional Encoder-Decoder (CED) architecture used in this paper. It is based on [26, 29, 43] and in particular on [25]. It was decided to discard the use of skip or additive connections, as there is no similarity on a pixel level between input and output domain. The main building block is a convolution block. It consists of a convolutional layer with different kernel and strides, batch normalization [27] and is concluded by a ReLU activation function [40]. A large kernel size of 7×7 pixels and 5×5 pixels in the encoding part ensures a large receptive field, which is especially important in CAI, as information is spread across the entire input image. The input image has either one or two channels for single and dual image reconstruction. During encoding, the spatial resolution is successively reduced to 32×32 pixels by strided convolution. Simultaneously, the number of channels increases from 1 or 2 to eventually 512 in the bottleneck. The bottleneck was intended for the actual transformation between detector domain and object domain and thus the majority of trainable parameters were located here. The decoder takes the encoded representation in the latent space and gradually increases the spatial size, while reducing the channels to reach a final output image with a single channel and 256×256 pixels. As described in [41], using transposed convolutions can lead to checkerboard artifacts in the output image, when performing image regression tasks. Therefore, bilinear resizing followed by unstrided convolutions was employed in this network’s decoder. The kernel size remains 3×3 for the complete decoder part, to produce high-resolution images. The final layer is a convolution followed by a linear activation function, because of its constant sensitivity. Overall, the CED architecture for single image reconstruction has a total of 16,414,817 trainable parameters. For dual image reconstruction this number increased to 16,416,385.

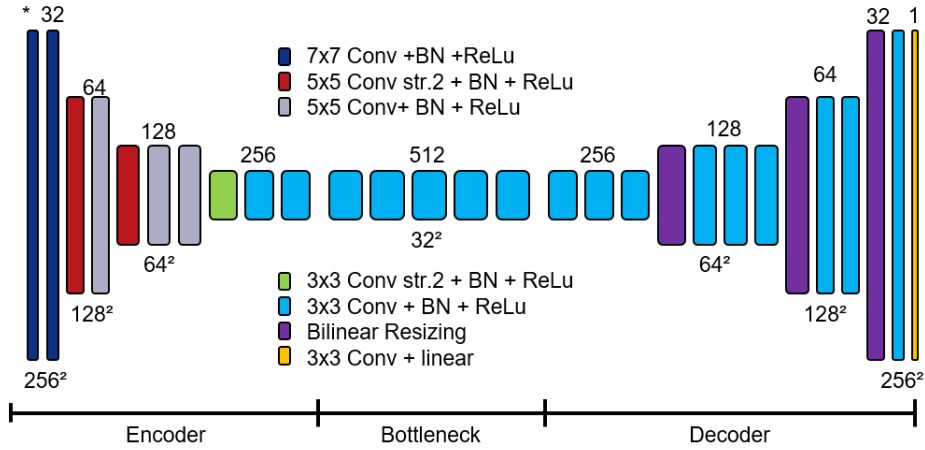


Figure 6: The architecture of the deployed CED network with the number of feature channels on top and the spatial size at the bottom. The asterisk indicates the input image channel size of 1 or 2 for single and dual image reconstruction respectively.

The CED, as well as all other reconstruction methods, were implemented, trained and evaluated in TensorFlow. The network intended for single image reconstruction and trained on the ImageNet data base is referred to as *CED-IN*, the network additionally fine-tuned on the horizontal lines data set as *CED-Lines* and the networks for dual image reconstruction as *DualCED-IN* and as *DualCED-Lines*. In the following both trained CEDs are considered as separate reconstruction methods.

The loss function used for training is the mean squared error of output and target image and was optimized by the Adam optimizer [33] with its standard parameters. A mini-batch size of 16 was chosen, as a compromise between short training time and a small generalization gap between training and validation data [22]. The network was trained on 35,000 training images and validated during training on separate 10,000 validation images for 20 epochs. If within three epochs the validation loss had not improved, the learning rate was automatically reduced by a factor of 0.4. Training was carried out for both training sets, whereas the network trained on the ImageNet data set served as pre-trained network for the horizontal line data set. Afterwards, the CEDs with the best performance on the validation set were chosen and applied to the hot-rod phantom data.

3.5 Contrast-to-noise ratio

The contrast-to-noise-ratio (CNR) takes the source visibility and the degradation from noise into consideration and is thus a suitable metric to evaluate the reconstruction quality. The following definition of CNR is employed [57]:

$$CNR = \frac{|\bar{S} - \bar{B}|}{\sigma}, \quad (3.1)$$

where \bar{S} denotes the mean intensity of the signal, \bar{B} the mean intensity and σ the standard deviation of the background. The binary ground truth images from Section 3.2 allowed the division of reconstructions into a signal part and a background part.

4 Results

4.1 Hot-rod phantom data

The first detector image for each of the three hot-rod phantoms is presented in figure 7a before and after post-processing together with the mask used for global outlier replacement (figure 7b). After capturing the data, pixels with high intensity corrupt the image and structure is barely visible as the top row shows. Replacing outliers by the median value of their neighborhood lead to a higher contrast and vertical stripes becoming visible.

The binary mask generated by thresholding the average detector image (figure 7b) shows multiple clusters of erroneous pixels in addition to pixels along the edge. Despite the post-processing, systematic pixel errors remain and are noticeable as black spots in the top right corner of all detector images of figure 9.

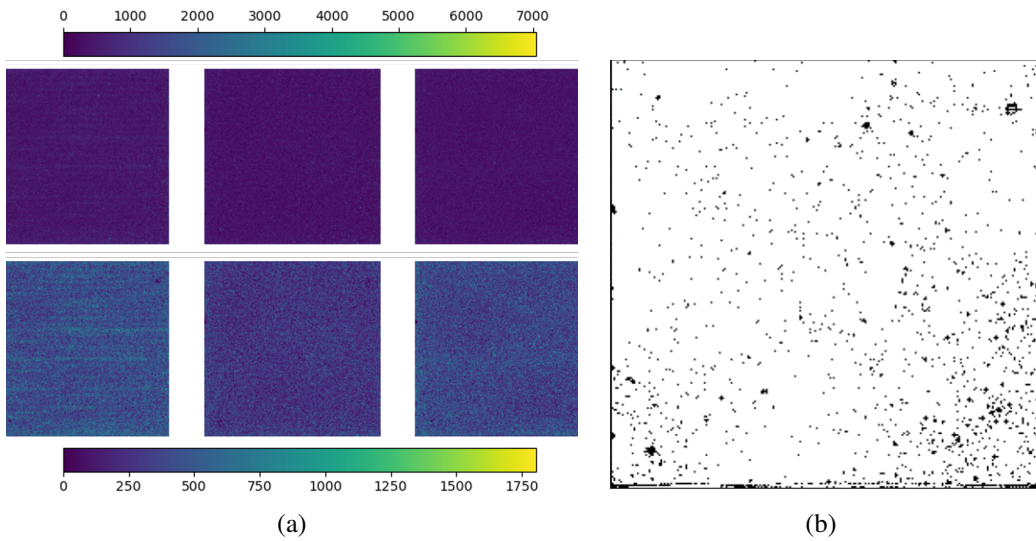


Figure 7: (a) shows the first detector images from each of the three captured hot-rod phantoms before (top) and after post-processing. (b) shows the binary mask which mark outliers in black according to the 2nd and 98th percentile of the average detector image.

4.2 Single image reconstruction

Training the CED-IN took roughly 11:55 h and additionally 3:45 h for the horizontal line data set. Generally, for all analytical reconstruction methods, the THT mask pattern for reconstruction obtained a higher contrast-to-noise ratio (CNR) than its NTH version and is presented here. The distribution of the 120 CNR values for the five different reconstruction methods are presented as grouped boxplots in figure 8. The median CNR values for each reconstruction method and phantom is listed on the left side of table 1.

In general, it is noticeable, that the SRP is reconstructed with a higher CNR than the other two phantoms: While the median CNR for the SRP varies between 1.72 and 5.22, the median CNR for the CP ranges between 0.57 and 1.42 and is even smaller for the LRP with median CNRs below 1.0.

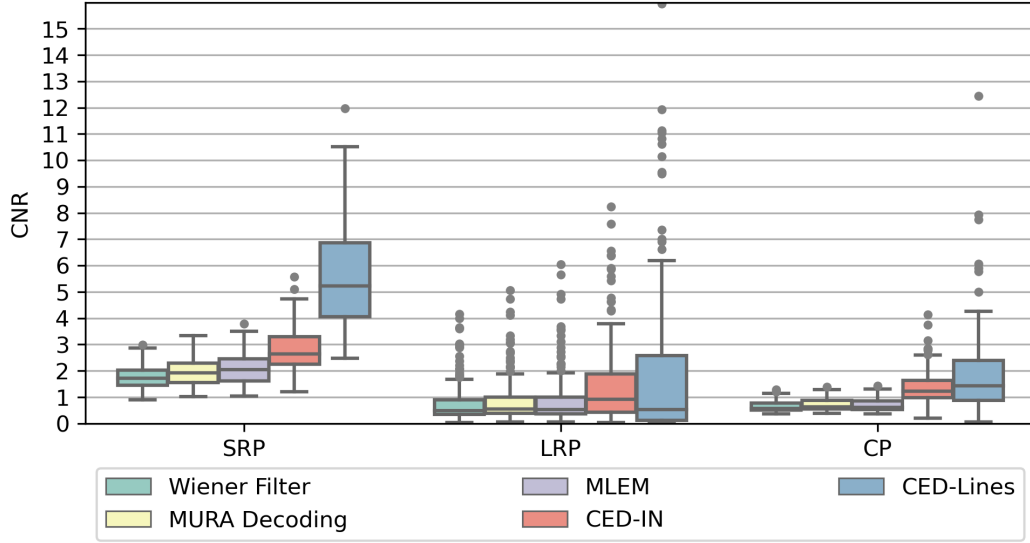


Figure 8: CNR distribution of the reconstructions of single acquisition images separated by the three hot-rod phantoms (SRP, LRP and CP). Three data points above a CNR of 16 are not shown.

The Wiener Filter and MURA Decoding show similar reconstruction results, with the latter producing a smoother background and hence reaching an overall higher median CNR of for all phantoms. Among the analytical reconstruction methods however, the MLEM algorithm yields the best reconstructions for the SRP and reaches a similar quality for the remaining phantoms. The MLEM algorithm produces reconstructions with a darker and more uniform background, but also slightly weaker tubes.

Both data-driven reconstruction methods reconstructed the phantoms with a higher CNR than all analytical reconstruction methods for two of the three phantoms, namely the SRP and CP. While the CED-IN is able to achieve a median CNR of 2.65 for the SRP, the CED-Lines achieves an even higher median of 5.22. Only for the LRP, the CEDs did not outperform the analytical reconstruction methods. The background generated by the data-driven reconstruction methods and especially for the CED-Lines is almost uniformly dark, except for few cloudy artifacts.

As figure 9 shows, the tubes of the SRP are clearly visible in all reconstructions, but are less visible for the LRP and CP, especially when the tubes are in an upright position. The second tube of the CP is missing from the reconstructions overall. Striking is a thin bright stripe on the top and in few cases also at the bottom of each image in most of the reconstructions, independent from the reconstruction method.

The average run times and their standard deviations are presented in table 2: 25 iterations of MLEM executed on the CPU takes an average of 12,995 ms, while MURA Decoding takes 288 ms and the Wiener Filter even less with 67 ms. The run time of both CEDs are almost identical with 285 ms and 286 ms on the CPU and 52 ms for both CEDs on the GPU. When deploying a GPU, the run times decrease, but are generally associated with a higher standard deviation.

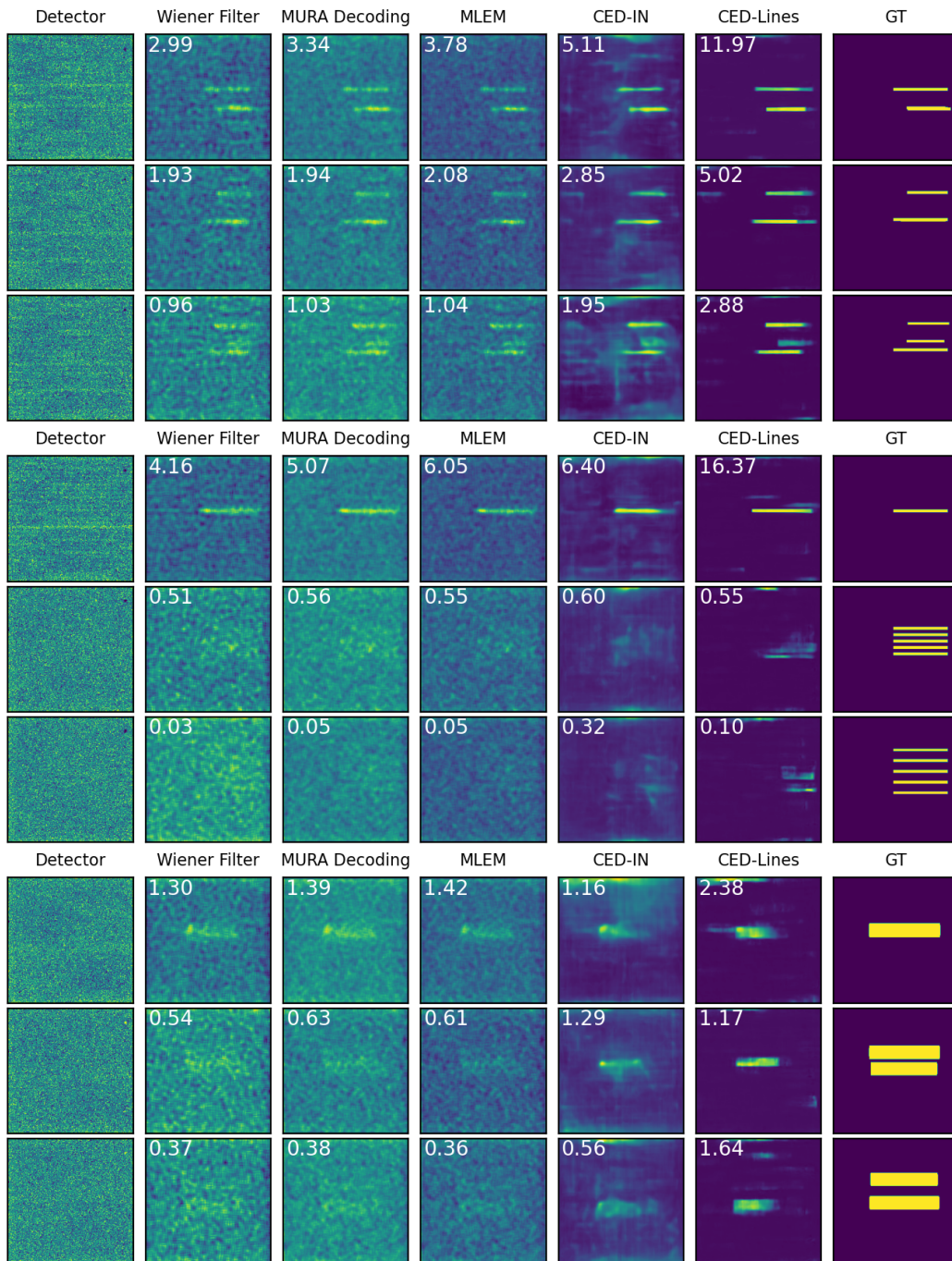


Figure 9: Exemplary reconstructions of single acquisition images from the SRP (top), LRP (middle) and CP (bottom). For each hot-rod phantom from top to bottom row: The best, median and worst reconstructions from MURA Decoding and the corresponding reconstructions from all other methods are displayed. The outer columns represents the detector and ground truth (GT) image respectively. Note, that the histogram of the detector images were equalized for better visualization.

Table 1: The median CNR separated in single or dual image reconstruction and hot-rod phantoms. The reconstruction method with the highest CNR is printed in bold.

Reconstruction method	single reconstruction			dual reconstruction		
	SRP	LRP	CP	SRP	LRP	CP
Wiener Filter	1.72	0.50	0.57	2.22	1.02	0.89
MURA Decoding	1.92	0.56	0.62	2.60	1.24	1.10
MLEM	2.04	0.53	0.61	1.98	0.92	0.81
Dual/CED-IN	2.65	0.92	1.23	4.00	1.54	2.37
Dual/CED-Lines	5.22	0.52	1.42	6.78	1.71	2.86

Table 2: Average run time of each algorithm in milliseconds on the computer specifications described in the text.

Reconstruction method	single reconstruction		dual reconstruction	
	CPU	GPU	CPU	GPU
Wiener Filter	67 ± 6	37 ± 44	66 ± 6	35 ± 35
MURA Decoding	288 ± 8	53 ± 56	311 ± 14	55 ± 106
MLEM	12,995 ± 345	823 ± 12	17,905 ± 299	1,097 ± 158
Dual/CED-IN	285 ± 10	52 ± 37	290 ± 9	51 ± 18
Dual/CED-Lines	286 ± 9	52 ± 17	300 ± 27	53 ± 37

4.3 Dual image reconstruction

The training of the DualCEDs took 18:42 h and additionally 8:45 h for the ImageNet based and horizontal line data set. Pairing the detector image with a second image taken with the anti-mask, resulting in dual image reconstruction, increased the reconstruction quality for almost all reconstruction methods and phantoms. Only the MLEM algorithm does not improve as much as the other methods and the CNR even decreased for the SRP by 0.12 as the boxplots in figure 10 and the right half of table 1 show.

The reconstruction method with the highest median CNRs for all three phantoms are the DualCED-Lines, followed by the DualCED-IN and MURA Decoding. The CNR difference between the Wiener Filter and MURA Decoding is higher for dual image reconstruction than for the single image reconstruction. The run times presented in right half of table 2 indicate, that only minor differences exist between the single image reconstruction methods. Mainly the MLEM algorithm run time increased by 4,910 ms to 17,905 ms on CPU and by 274 ms to 1,097 ms on the GPU.

As figure 11 shows, clearer tubes and a more uniform background is noticeable in comparison to the single image reconstructions for all reconstruction methods. With the LRP in an upright position, a bright block is visible, although, single tubes remain indistinguishable. The DualCED-Lines fails to reconstruct this structure and remains dark except for a single point and a small line. Additionally, parts of the second tube of the CP become visible for all reconstruction methods, however, less

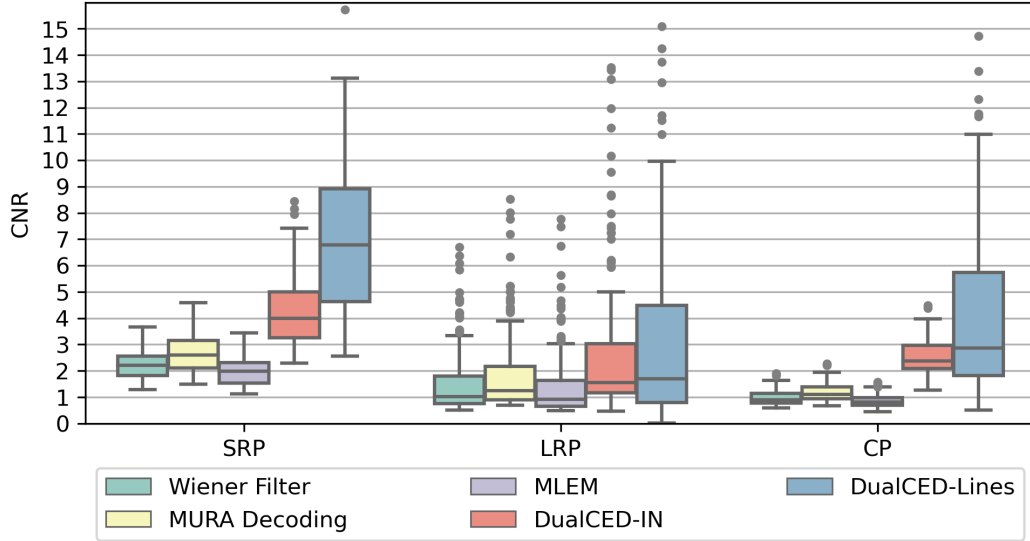


Figure 10: CNR distribution of the reconstruction of dual acquisition images separated by phantoms. 16 data points are above a CNR of 16.

prominent for the DualCED-Lines network.

5 Discussion

5.1 Hot-rod phantom data

Even after post-processing the acquired images, noise dominates all detector images and an image structure is more distinctive for the SRP and in lesser extend for the CP and LRP. An air bubble trapped in one tube of the CP, which was found after the measurements took place, posed an additional challenge. A major source of errors in planar reconstruction is the hot-rod phantoms' depth. Even though, the maximum depth of the radiating tubes is 17 mm and thus small in comparison to the object to camera distance of 172 mm, one can assume that reconstruction artifacts are partially caused by this *out-of-focus* effect [54].

Despite the global outlier replacement, the acquired detector images still partially suffered from systematic pixel errors. Capturing a flood-field image would lead to an improved outlier detection and an overall higher image uniformity and is already planned for the future. Nonetheless, all reconstruction methods are equally affected by noise, the *out-of-focus* effect and pixel errors, which assures a fair comparison and assessment of each method.

5.2 Single image reconstruction

The most commonly used MURA Decoding proved to be good compromise between reconstruction quality and run time, despite its negligence of Poisson noise and near-field artifacts, which remain uncompensated in single image reconstruction. The Wiener Filter yields reconstructions of comparable quality but has the advantage of a more than four times faster run time. The MLEM algorithm offers higher reconstruction quality but comes with higher computational cost as it takes more than

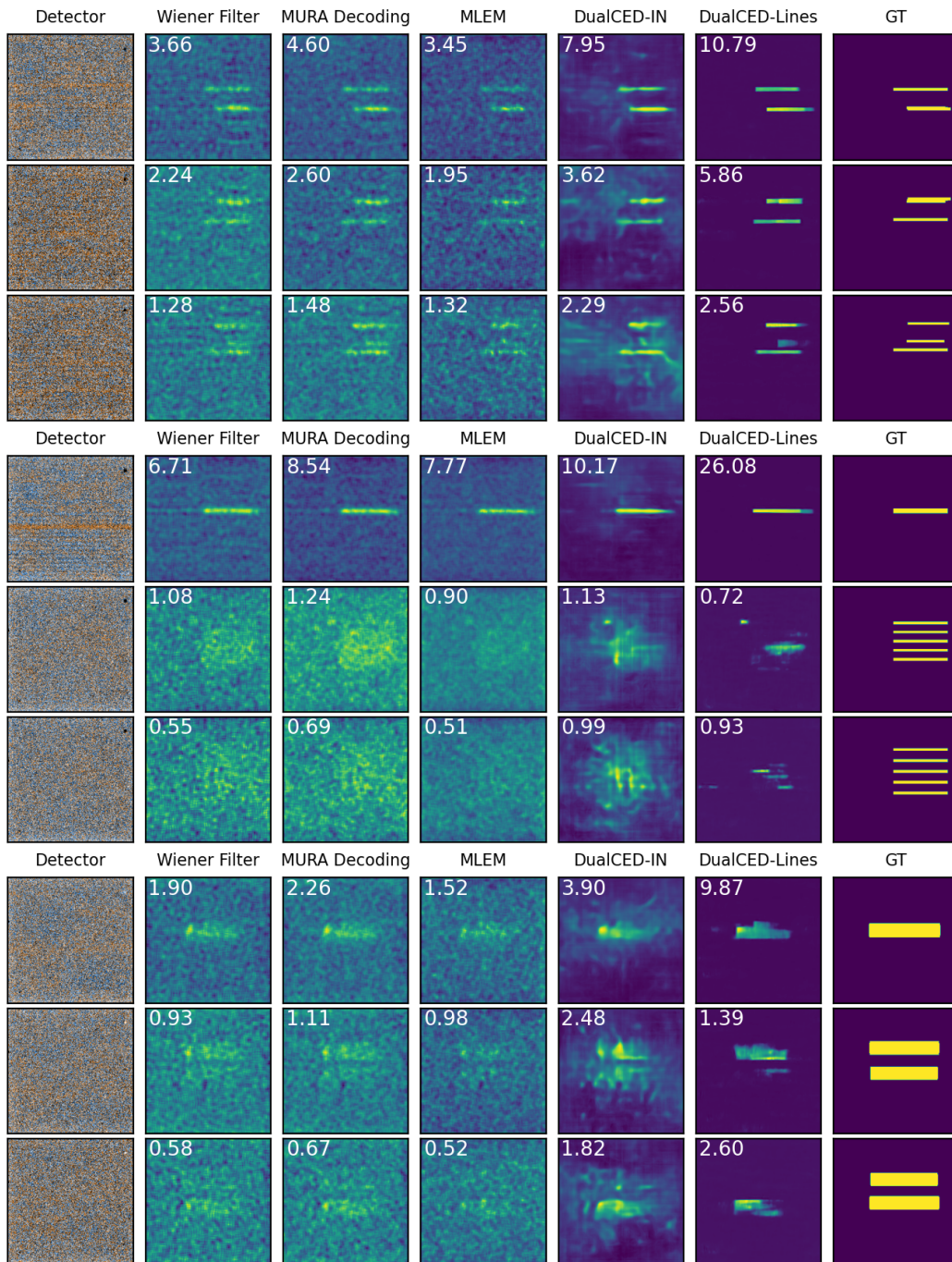


Figure 11: Exemplary images of dual image reconstructions taken from the SRP, LRP and CP with corresponding detector image and ground truth. The detector image taken with mask and with anti-mask are presented in false colors and were also histogram equalized.

45 times as long for a single reconstruction compared to MURA Decoding. It is conceivable to accelerate the run time by reducing the number of iterations. This could be achieved, for example, by combining the MLEM algorithm with a Wiener Filter for the initial guess.

Overall, both data-driven reconstruction methods reconstructed the phantoms better than all analytical reconstruction methods, even though the simulated training data (see Section 3.4.1) did not account for near-field artifacts: Most of the reconstructions, independent from the reconstruction method, have a thin bright stripe on the top and few also at the bottom of each image. This is a hint that collimation appears caused by the small pinholes as is described in detail in [39]. Especially the SRP is reconstructed 1.38 times better by CED-IN compared to the most commonly used MURA Decoding and 2.71 times better by the CED-Lines. Only in the case of particularly noisy detector images, which is the case for many of the LRP images, as explained below, the CEDs were not superior to the analytical reconstruction methods. The run time of both CEDs are almost identical and circa 286 ms thus equally as time consuming as MURA Decoding.

Not only the CEDs, but all reconstruction methods have difficulties in reconstructing the phantoms when the tubes are in an upright position. In these cases, the radiating sources are spread over a large portion of the FoV. This leads to a more uniform distribution of the photons hitting the detector, which reduces the photon count per pixel. Hence, Poisson noise dominates these detector images. The almost blank reconstructions of the LRP show that the data-driven reconstruction methods cannot perform miracles and rely on a certain quality of the detector image.

5.3 Dual image reconstruction

All reconstruction methods benefit enormously from the use of the dual image approach. As has been stated before [5, 39, 54], combining two captures with mask and anti-mask improves contrast and diminishes artifacts emerging from the near-field application. For example, the bright stripe at the top and bottom of the reconstructions from the single image acquisition disappeared for the dual acquisition. The MLEM algorithm did not benefit as much from the dual image approach as the other methods, suggesting that iterative forward- and backward-projection weakens the compensation effect.

With all methods, the background becomes more uniform and with the LRP in the upright position a bright block is even visible, although, single tubes remain indistinguishable. The DualCED-Lines fails to reconstruct this structure and only shows singular bright spots. Additionally, parts of the second tube of the CP become visible for all reconstruction methods, except for the DualCED-Lines network. Overall, however, DualCED-Lines increased its superiority over MURA Decoding with a factor of 1.45 (SRP), 1.37 (LRP) and 2.6 (CP). However, as mentioned above and depicted in figure 10, certain images are reconstructed worse than by MURA Decoding.

The fact that the data-driven reconstruction methods produce reconstructions with similar characteristic (e. g. the small bright spot in the middle row of the LRP, or the highlighted tip in the top image of the CP in figure 11) as all analytical reconstruction methods supports the hypothesis, that the CEDs "learnt" an actual image-to-image mapping and not just searched for known patterns.

Unlike the data-driven CEDs, analytical reconstruction methods are based on a formal description of the imaging system. Two problems arise from this: First, it has been shown, that Convolutional Neural Networks can inherently possess instabilities [8, 49]: For CAI, instabilities mean that perturbations of the detector image, that are caused by small unnoticeable changes of the object

image, lead to nonsensical reconstructions, for example large uniform areas or negative pixel values. However, instabilities were only noticed during training and none of the final deployed networks showed instable behavior. Second, the reconstruction results are not independent from the type of training data. The different reconstruction results between DualCED-IN and DualCED-Lines can be attributed solely to the different image domains used for training. This data dependency has also been reported by [34] and indicates, that Deep Learning reconstruction methods are not capable yet, to learn a domain-independent image-to-image mapping. This can be used as an advantage as the following paragraph explains:

The two training sets used in this paper represent different approaches to CAI reconstruction: First, using natural photographs as object images forms a general approach aiming at the question "What do we have in front of our γ -camera?". Second, when the approximate amount, shape and size of radioactive sources is known, a training set can be tailored to that specific application. This approach contains a-priori knowledge and seeks to answer the question "Where is the source located?". For instance, when a coded aperture γ -camera is applied to the task of sentinel lymph node mapping, like several groups intend [18, 30, 31, 47], the radiating sources can be considered to be of spherical shape with maximally a few centimeters in diameter. A set of training data could be generated for this specific task.

In this paper, synthetic training data was generated (see Section 3.4.1), which does not consider near-field artifacts, thus, causing a domain gap between training data and measured phantom data. Neglecting transmission or scattering of γ -photons through the mask or collimation by the small pinholes are major simplifications and inherently limit the CEDs reconstruction ability. It can be expected, that a more realistic generation of training data has the potential to improve CAI reconstruction.

All in all, the quantitative comparison of this paper is limited to the described experimental set-up consisting of one object-to-camera distance and a single mask with fixed thickness. To what extent the described results apply to extreme near-field application remains an open question: The more a radiation source approaches the γ -camera, the more the imaging deviates from the mathematical description as a convolution. Thus, near-field effects, mainly the angular and the collimation effect, deteriorate the detector image. Specific compensation techniques have been proposed by [6, 39] and might be considered.

6 Conclusion & outlook

This paper compared different reconstruction methods for planar coded aperture imaging (CAI) on data acquired by an experimental γ -camera and three hot-rod phantoms. For the given set-up, MURA Decoding, the most commonly used CAI reconstruction method, provides robust reconstructions despite the assumption of a linear system model. The Wiener Filter produces results of almost similar quality, with the advantage of being more than four times faster. For single image reconstruction, however, the MLEM algorithm reaches the highest reconstruction quality among the analytical reconstruction methods, but comes with a computing time of roughly 13 s.

All in all, both Convolutional Encoder-Decoders (CED) were able to outperform analytical reconstruction methods for two out of three phantoms with single and for all three phantoms with dual image reconstruction using contrast-to-noise ratio for quality measure. The superior reconstruction

quality was obtained, despite a simple low-fidelity simulation of training data. The Deep Learning approach is especially successful, when a-priori knowledge about the expected radiation sources in the form of a tailored training set is used. Even though, none of the CED reconstructions exhibited instabilities, its possibility must be kept in mind. Therefore, a more in depth analysis of the CED regarding its trustworthiness and the influence of training data is required.

The CED used in this paper is only one architecture among many. It would be of interest to find out how other Deep Learning architectures perform, especially with regards to the computational costs and trustworthiness.

For mobile applications, a short exposure time and easy handling is of importance, and thus single image acquisition is preferred. In order to keep the exposure time as short as possible, the relationship between the number of photons detected and the reconstruction quality could be investigated. By incorporating near-field artifacts and mask transmission into the generation of training data, the domain gap for single image reconstruction could be decreased and make data-driven reconstruction methods even more attractive for mobile γ -cameras.

The acquired hot-rod phantom data are available upon request to the co-author Vladislav Rozhkov: rozhkov@jlnr.ru.

Acknowledgments

The authors gratefully acknowledge the data storage service SDS@hd supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) and the German Research Foundation (DFG) through grant INST 35/1314-1 FUGG and INST 35/1503-1 FUGG. This paper was partially funded by the Zentrales Innovationsprogramm Mittelstand (ZIM) under grant KK5044701BS0.

References

- [1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mane, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viegas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang et al. Zheng. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv preprint*, mar 2016.
- [2] J. G. Ables. Fourier Transform Photography: A New Method for X-Ray Astronomy. *Publications of the Astronomical Society of Australia*, 1(4):172–173, dec 1968.
- [3] Roberto Accorsi, Luigi Celentano, Paolo Laccetti, Richard C. Lanza, Marcello Marotta, Giovanni Mettivier, Maria Cristina Montesi, Giuseppe Roberti, and Paolo Russo. High-resolution 125I small animal imaging with a coded aperture and a hybrid pixel detector. *IEEE Transactions on Nuclear Science*, 55(1):481–490, 2008.
- [4] Roberto Accorsi, Assunta Simona Curion, Paola Frallicciardi, Richard C. Lanza, Adele Lauria, Giovanni Mettivier, Maria Cristina Montesi, and Paolo Russo. Preliminary evaluation of the tomographic performance of the mediSPECT small animal imaging system. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 571(1-2 SPEC. ISS.):415–418, 2007.

- [5] Roberto Accorsi, Francesca Gasparini, and Richard C. Lanza. Optimal coded aperture patterns for improved SNR in nuclear medicine imaging. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 474(3):273–284, 2001.
- [6] Roberto Accorsi and Richard C. Lanza. Near-field artifact reduction in planar coded aperture imaging. *Applied Optics*, 40(26):4697, sep 2001.
- [7] James Ahrens, Berk Geveci, and Charles Law. Paraview: An end-user tool for large data visualization. *The visualization handbook*, 717(8), 2005.
- [8] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, and Anders C. Hansen. On instabilities of deep learning in image reconstruction and the potential costs of AI. *Proceedings of the National Academy of Sciences*, 117(48):30088–30095, dec 2020.
- [9] Chinmay Belthangady and Loic A Royer. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature Methods*, 16(12):1215–1225, dec 2019.
- [10] Jürgen Beyerer, Fernando Puente León, and Christian Frese. *Machine vision: Automated visual inspection: Theory, practice and applications*. Springer, 2015.
- [11] Gary Bradski. The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*, 25(11):120–123, 2000.
- [12] João Braga. Coded Aperture Imaging in High-energy Astrophysics. *Publications of the Astronomical Society of the Pacific*, 132(1007):12001, 2020.
- [13] Thorsten Buzug. *Computed tomography: From photon statistics to modern cone-beam CT*. Springer Berlin Heidelberg, 2008.
- [14] E. Caroli, J. B. Stephen, G. Di Cocco, L. Natalucci, and A. Spizzichino. Coded aperture imaging in X-and gamma-ray astronomy. *Space Science Reviews*, 45(3-4):349–403, 1987.
- [15] Michał J. Cieślak, Kelum A.A. Gamage, and Robert Glover. Coded-aperture imaging systems: Past, present and future development – A review. *Radiation Measurements*, 92:59–71, sep 2016.
- [16] RH Dicke. Scatter-hole cameras for x-rays and gamma rays. *The astrophysical journal*, 153:L101, 1968.
- [17] E. E. Fenimore. Coded aperture imaging: predicted performance of uniformly redundant arrays. *Applied Optics*, 17(22):3562, 1978.
- [18] Hirofumi Fujii, John D. Idoine, Sylvain Gioux, Roberto Accorsi, David R. Slochower, Richard C. Lanza, and John V. Frangioni. Optimization of coded aperture radiosciintigraphy for sentinel lymph node mapping. *Molecular Imaging and Biology*, 14(2):173–182, 2012.
- [19] Olivier Gal, Mehdi Gmar, Oleg P. Ivanov, Frédéric Lainé, Fabrice Lamadie, Christophe Le Goaller, Charly Mahé, Erwan Manach, and Vyacheslav E. et al. Stepanov. Development of a portable gamma camera with coded aperture. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 563(1):233–237, 2006.
- [20] F. Garibaldi, Roberto Accorsi, M. N. Cinti, E. Cisbani, S. Colilli, F. Cusanno, G. De Vincentis, A. Fortuna, R. Fratoni, B. Girolami, F. Ghio, F. Giuliani, M. Gricia, Richard C. Lanza, A. Loizzo, S. Loizzo, M. Lucentini, S. Majewski, F. Santavenere, R. Pani, R. Pellegrini, A. Signore, F. Scopinaro, and P. et al. Veneroni. Small animal imaging by single photon emission using pinhole and coded aperture collimation. *IEEE Transactions on Nuclear Science*, 52(3 I):573–579, 2005.
- [21] M. Gmar, M. Agelou, F. Carrel, and V. Schoepff. GAMPIX: A new generation of gamma camera.

Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 652(1):638–640, 2011.

- [22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. The MIT Press, Cambridge, Massachusetts ; London, England, 2016.
- [23] Stephen R Gottesman and E. E. Fenimore. New family of binary arrays for coded aperture imaging. *Applied Optics*, 28(20):4344, 1989.
- [24] A Haboub, A A Macdowell, S Marchesini, and D Y Parkinson. Coded aperture imaging for fluorescent x-rays. *Review of Scientific Instruments*, 85(6):63704, 2014.
- [25] Ida Häggström, C. Ross Schmidlein, Gabriele Campanella, and Thomas J. Fuchs. DeepPET: A deep encoder–decoder network for directly solving the PET image reconstruction inverse problem. *Medical Image Analysis*, 54:253–262, 2019.
- [26] Ji He, Yongbo Wang, and Jianhua Ma. Radon Inversion via Deep Learning. *IEEE Transactions on Medical Imaging*, 39(6):2076–2087, jun 2020.
- [27] Sergey Ioffe and Christian Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 448–456, Lille, France, 2015. PMLR.
- [28] Manhee Jeong and Mark D. Hammig. Comparison of gamma ray localization using system matrixes obtained by either MCNP simulations or ray-driven calculations for a coded-aperture imaging system. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 954(July 2018):161353, 2020.
- [29] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael Unser. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017.
- [30] Ioannis Kaissas, Christos Papadimitropoulos, K. Karafasoulis, C. Potiriadis, and C. P. Lambropoulos. 3-D localization of gamma ray sources with coded apertures for medical applications. *Journal of Physics: Conference Series*, 637(1), 2015.
- [31] Ioannis Kaissas, Christos Papadimitropoulos, C. Potiriadis, K. Karafasoulis, D. Loukas, and C. P. Lambropoulos. Imaging of spatially extended hot spots with coded apertures for intra-operative nuclear medicine applications. *Journal of Instrumentation*, 12(1):1–9, 2017.
- [32] John A. Kennedy, Ora Israel, and Alex Frenkel. 3D iteratively reconstructed spatial resolution map and sensitivity characterization of a dedicated cardiac SPECT camera. *Journal of Nuclear Cardiology*, 21(3):443–452, 2014.
- [33] Diederik P. Kingma and Jimmy Lei Ba. Adam: A method for stochastic optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, pages 1–15, 2015.
- [34] Anicó Kulow, Ana Guilherme Buzanich, Uwe Reinholz, Franziska Emmerling, Sven Hampel, Ursula Elisabeth Adriane Fittschen, Christina Strelí, and Martin Radtke. Comparison of three reconstruction methods based on deconvolution, iterative algorithm, and neural network for X-ray fluorescence spectroscopy with coded apertures. *Journal of Analytical Atomic Spectrometry*, 2020.
- [35] Mark T. Madsen. Recent advances in SPECT imaging. *Journal of Nuclear Medicine*, 48(4):661–673, 2007.

- [36] A. Martineau, J. M. Rocchisani, and J. L. Moretti. Coded aperture optimization using Monte Carlo simulations. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 616(1):75–80, 2010.
- [37] Philipp Matthies, Kanishka Sharma, Asli Okur, José Gardiazabal, Jakob Vogel, Tobias Lasser, and Nassir Navab. First use of mini gamma cameras for intra-operative robotic SPECT reconstruction. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, pages 163–70, 2013.
- [38] Zhiping Mu, Lawrence W. Dobrucki, and Yi Hwa Liu. SPECT Imaging of 2-D and 3-D Distributed Sources with Near-Field Coded Aperture Collimation: Computer Simulation and Real Data Validation. *Journal of Medical and Biological Engineering*, 36(1):32–43, feb 2016.
- [39] Zhiping Mu and Yi-Hwa Liu. Aperture collimation correction and maximum-likelihood image reconstruction for near-field coded aperture imaging of single photon emission computerized tomography. *IEEE Transactions on Medical Imaging*, 25(6):701–711, jun 2006.
- [40] Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve Restricted Boltzmann machines. In *ICML 2010 - Proceedings, 27th International Conference on Machine Learning*, pages 807–814, 2010.
- [41] Augustus Odena, Vincent Dumoulin, and Chris Olah. Deconvolution and Checkerboard Artifacts. *Distill*, 1(10):1–4, 2017.
- [42] Todd E. Peterson and Lars R. Furenlid. SPECT detectors: The Anger Camera and beyond. *Physics in Medicine and Biology*, 56(17), 2011.
- [43] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241, 2015.
- [44] Daniel Roth, Erik Larsson, Anna Sundlöv, and Katarina Sjögreen Gleisner. Characterisation of a hand-held CZT-based gamma camera for ¹⁷⁷Lu imaging. *EJNMMI Physics*, 7(1), 2020.
- [45] V. Rozhkov, G. Chelkov, I. Hernández, O. Ivanov, D. Kozhevnikov, A. Leyva, A. Perera, D. Rastorguev, P. Smolyanskiy, L. Torres, and A. Zhemchugov. Visualization of radiotracers for SPECT imaging using a Timepix detector with a coded aperture. *Journal of Instrumentation*, 15(06):P06028–P06028, 2020.
- [46] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3):211–252, 2015.
- [47] Paolo Russo, Francesca Di Lillo, Veronica Corvino, Paola Maria Frallicciardi, Antonio Sarno, and Giovanni Mettivier. CdTe compact gamma camera for coded aperture imaging in radioguided surgery. *Physica Medica*, 69(June 2019):223–232, 2020.
- [48] Dawid Schellingerhout, Roberto Accorsi, Umar Mahmood, John Idoine, Richard C. Lanza, and Ralph Weissleder. Coded aperture nuclear scintigraphy: A novel small animal imaging technique. *Molecular Imaging*, 1(4):344–353, 2002.
- [49] Jiawei Su, Danilo Vasconcellos Vargas, and Kouichi Sakurai. One Pixel Attack for Fooling Deep Neural Networks. *IEEE Transactions on Evolutionary Computation*, 23(5):828–841, 2019.
- [50] Shifeng Sun, Yang Liu, and Xiaoping Ouyang. Near-field high-resolution coded aperture gamma-ray imaging with separable masks. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 951(October 2019):163001, 2020.

- [51] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-Recurrent Network for Deep Image Deblurring. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 8174–8182, 2018.
- [52] Makoto Tsuchimochi and Kazuhide Hayama. Intraoperative gamma cameras for radioguided surgery: Technical characteristics, performance parameters, and clinical applications. *Physica Medica*, 29(2):126–138, 2013.
- [53] Karen Van Audenhaege, Roel Van Hoken, Stefaan Vandenberghe, Christian Vanhove, Scott D. Metzler, and Stephen C. Moore. Review of SPECT collimator selection, optimization, and fabrication for clinical and preclinical imaging. *Medical Physics*, 42(8):4796–4813, 2015.
- [54] Olga I. Vassilieva and Roy C. Chaney. Method for reducing background artifacts from images in single-photon emission computed tomography with a uniformly redundant array coded aperture. *Applied Optics*, 41(7):1454, 2002.
- [55] Kai Vetter, Ross Barnowski, Andrew Haefner, Tenzing HY Joshi, Ryan Pavlovsky, and Brian J Quiter. Gamma-ray imaging for nuclear security and safety: Towards 3-d gamma-ray vision. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 878:159–168, 2018.
- [56] Rui Zhang, Pin Gong, Xiaobin Tang, Peng Wang, Cheng Zhou, Xiaoxiang Zhu, Le Gao, Dajian Liang, and Zeyu Wang. Reconstruction method for gamma-ray coded-aperture imaging based on convolutional neural network. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 934:41–51, aug 2019.
- [57] Rui Zhang, Xiaobin Tang, Pin Gong, Peng Wang, Cheng Zhou, Xiaoxiang Zhu, Dajian Liang, and Zeyu Wang. Low-noise reconstruction method for coded-aperture gamma camera based on multi-layer perceptron. *Nuclear Engineering and Technology*, 52(10):2250–2261, oct 2020.
- [58] Yanbo Zhang and Hengyong Yu. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE transactions on medical imaging*, 37(6):1370–1381, 2018.
- [59] Bo Zhu, Jeremiah Z. Liu, Stephen F. Cauley, Bruce R. Rosen, and Matthew S. Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, mar 2018.