ktu
1922

Kaunas University of Technology

Faculty of Mathematics and Natural Sciences

The Identification and Removal of Artefacts in Medical Images

Masters's Final Degree Project

Dovydas Lugas

Project author

Assoc. Prof. Dr. Kristina Šutienė

Supervisor

Kaunas, 2023

Kaunas University of Technology

Faculty of Mathematics and Natural Sciences

The Identification and Removal of Artefacts in Medical Images

Masters's Final Degree Project

Applied Mathematics (6211AX006)

Dovydas Lugas

Project author

Assoc. Prof. Dr. Kristina Šutienė

Supervisor

Assoc. Prof. Dr. Liepa Bikulčienė

Reviewer

Kaunas, 2023

Kaunas University of Technology

Faculty of Mathematics and Natural Sciences

Dovydas Lugas

The Identification and Removal of Artefacts in Medical Images

Declaration of Academic Integrity

I confirm the following:

1. I have prepared the final degree project independently and honestly without any violations of the copyrights or other rights of others, following the provisions of the Law on Copyrights and Related Rights of the Republic of Lithuania, the Regulations on the Management and Transfer of Intellectual Property of Kaunas University of Technology (hereinafter – University) and the ethical requirements stipulated by the Code of Academic Ethics of the University;

2. All the data and research results provided in the final degree project are correct and obtained legally; none of the parts of this project are plagiarised from any printed or electronic sources; all the quotations and references provided in the text of the final degree project are indicated in the list of references;

3. I have not paid anyone any monetary funds for the final degree project or the parts thereof unless required by the law;

4. I understand that in the case of any discovery of the fact of dishonesty or violation of any rights of others, the academic penalties will be imposed on me under the procedure applied at the University; I will be expelled from the University and my final degree project can be submitted to the Office of the Ombudsperson for Academic Ethics and Procedures in the examination of a possible violation of academic ethics.

<div align="right">

Dovydas Lugas

Confirmed electronically

</div>

Summary

Artefacts in computed tomography images are a relatively common occurrence. These include beam hardening, motion, ring, noise, metallic and other types of artefacts. Among these, some of the most common are metallic and metal-induced artefacts like beam hardening and Poisson noise. In this paper, several models for metal artefact reduction in computed tomography images are examined. Three models were tested: the artefact disentanglement network, the reused convolutional network, and a reused convolutional network enhanced with sinogram information. The study showed that the models performed similarly quantitatively, as indicated by comparable values in metrics such as peak signal-to-noise ratio and structural similarity index measure. However, in qualitative assessments, it was determined that the reused convolutional network yielded higher-quality images compared to the other models.

Santrauka

Artefaktai kompiuterinėje tomografijoje yra gana dažnas reiškinys. Jų būna įvairių – spindulio deformacijos, judėjimo, žiedo, triukšmo, metalo ir kiti. Vieni iš dažniausiai pasitaikančių yra metaliniai ir metalinių objektų sukelti artefaktai. Šiame darbe nagrinėjami trys modeliai skirti metalo artefaktų šalinimui kompiuterinės tomografijos vaizduose: artefaktų atskyrimo tinklas, pakartotinis konvoliucinis tinklas ir pakartotinis konvoliucinis tinklas, sustiprintas informacija gauta iš sinogramų. Tyrimas parodė, kad modeliai kiekybiškai – pagal rodiklius kaip maksimalus signalo triukšmo santykis ir struktūrinio panašumo indeksas, veikia panašiai. Vis dėlto, kokybiniu vertinimu buvo nustatyta, kad pakartotinis konvoliucinis tinklas duoda aukštesnės kokybės rezultatus lyginant su kitais modeliais.

Table of contents

List of Figures

Introduction

Rapid advancements in deep learning, especially computer vision, have encouraged the adoption of AI techniques in medical image analysis. The problems solved in this field are varied from pathology analysis to lesion segmentation to image quality enhancement. Image quality enhancement, specifically image denoising, is especially important because it influences the quality of the solutions to the other problems. One of the most common types of noise caused in medical images in general, and in CT images in particular, is noise induced by metallic implants. This noise tends to obscure important information in the image, which hinders interpretability and diagnosis, and complicates other types of AI tasks.

Naturally, the AI techniques applied to these problems primarily fall under supervised learning, i.e., they require a large amount of paired data. However, acquiring such datasets in medical contexts can be difficult. This is especially apparent in medical image denoising, where training with labeled data would require having two versions of the same image - noisy and clean. Clearly, this is impossible, so synthetically generated noise-affected images are used as a substitute, which can have a negative effect on performance when applying models trained on synthetic data to natural image datasets. This is where unsupervised approaches offer an alternative, as they do not require labels.

Several unsupervised methods have been proposed in recent years specifically for medical image denoising; however, most of them operate only in the image domain. Advances in supervised approaches have demonstrated that dual domain methods that combine information present in both sinogram and image domains offer superior performance. This indicates that the application of dual-domain information for unsupervised AI methods may be worth studying.

Objective – devise an unsupervised AI architecture that utilizes dual-domain information to remove artefacts from CT images caused by metallic implants.

Tasks:
- To conduct a review of the relevant literature.
- To generate a CT dataset with synthetic artefacts for model evaluation.
- To study the effects of parallel sinogram information integration technique on model performance.
- To compare the results of the final model to other state-of-the-art unsupervised models.

1. Literature Review

1.1. Noise in Medical Imaging

As mentioned in the introduction, artefacts are very common in clinical CT imaging [1]. There are many different types of CT artefacts, including noise, beam hardening, scatter, motion, ring, and metal artefacts. Artefacts appear when conditions for an ideal CT image are not met. This could be due to malfunctioning equipment, patient motion, low radiation dosage, and other factors. For example, ring artefacts are caused by malfunctioning detectors, while Poisson noise generally appears due to low photon counts. Examples of both can be seen in figure 1.
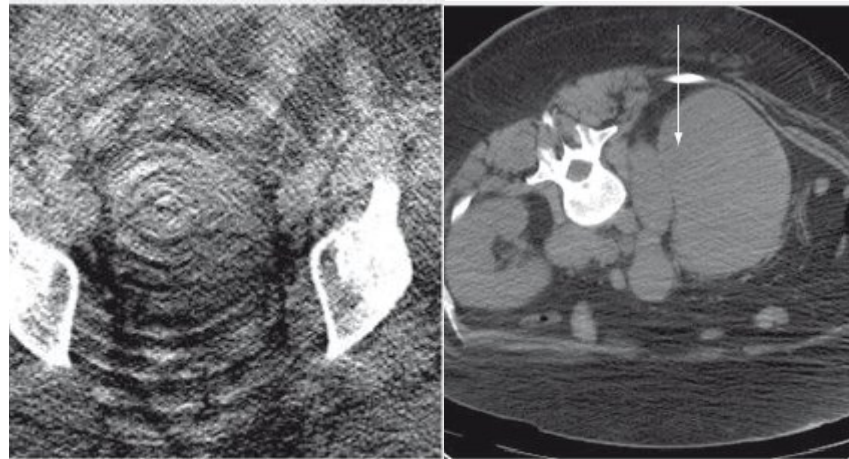
Fig. 1. An example of ring artefact and Poisson noise [1]

Metallic implants are the primary source of beam hardening and scatter, which produce dark streaks with adjacent white streaks, while patient motion is responsible for blurring and double images. An example can be seen in figure 2.
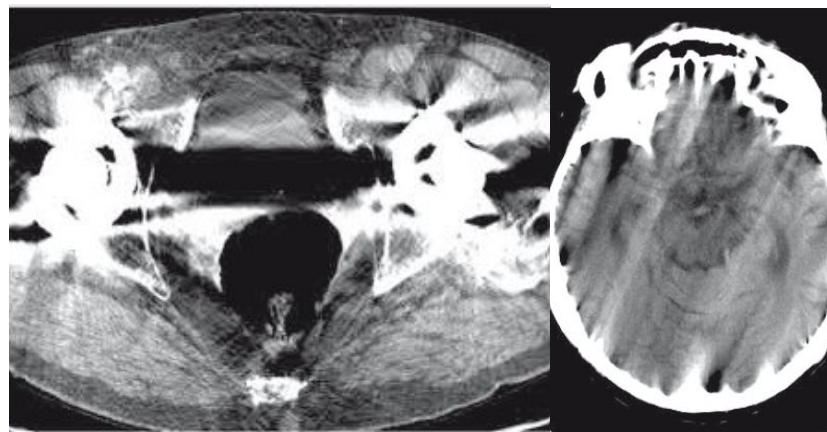
Fig. 2. An example of metal-induced and motion artefacts [1]

Replicating ideal conditions for a CT scan may not be possible due to factors like metallic implants inside the patient's body or the usage of low-dose radiation due to concerns for the health of the patient. This makes artefact reduction a necessary area of study to improve diagnosis and patient outcomes.

1.2. Conventional Metal Artefact Reduction

There are a couple of conventional metal artefact reduction techniques that are worth mentioning in this discussion since they are often used for benchmarking or preliminary preprocessing of model inputs:

– Linear interpolation (LI).
– Normalized artefact reduction (NMAR).

Both can be classified as projection completion techniques, which involve the interpolation of missing or corrupt values of images in the projection domain. The simpler approach is linear interpolation, which is fast and has no parameters but tends to introduce secondary artefacts. That is why it is seldom used on its own. Instead, it often provides a simple way to obtain an image prior that can be input into more complex models.

In general, pure interpolation models are not very popular, because they tend to introduce additional artefacts after image reconstruction [2]. These include blurring around the metal edges and the formation of streak artefacts due to insufficiently smooth transitions between the interpolated and real regions. The normalized artefact reduction model (NMAR) [2] attempts to solve these problems by first normalizing the projection to produce a more homogenous sinogram and only then applying the interpolation algorithm. A diagram is presented in figure 3.

Fig. 3. Diagram of the NMAR method [2]

As you can see in figure 3, the result of the interpolation is denormalized and combined with the sinogram prior obtained through thresholding. This procedure provides for smoother transitions between interpolated and real data reducing the incidence of secondary artefacts.

1.3. Supervised Metal Artefact Reduction

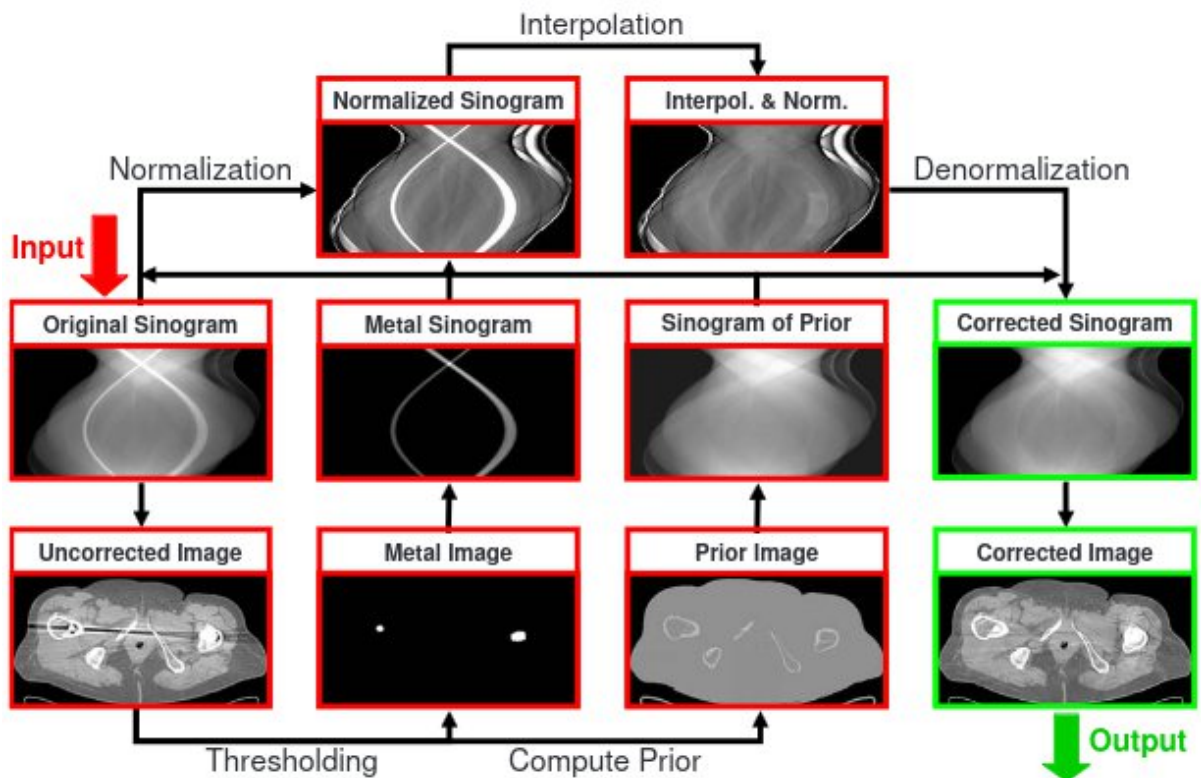Several supervised deep learning models have been proposed in recent years specifically tailored to the task of metal artefact reduction. One of the most well-known is the convolutional neural network based metal artefact reduction (CNNMAR) model [3], which is visualized in figure 4.
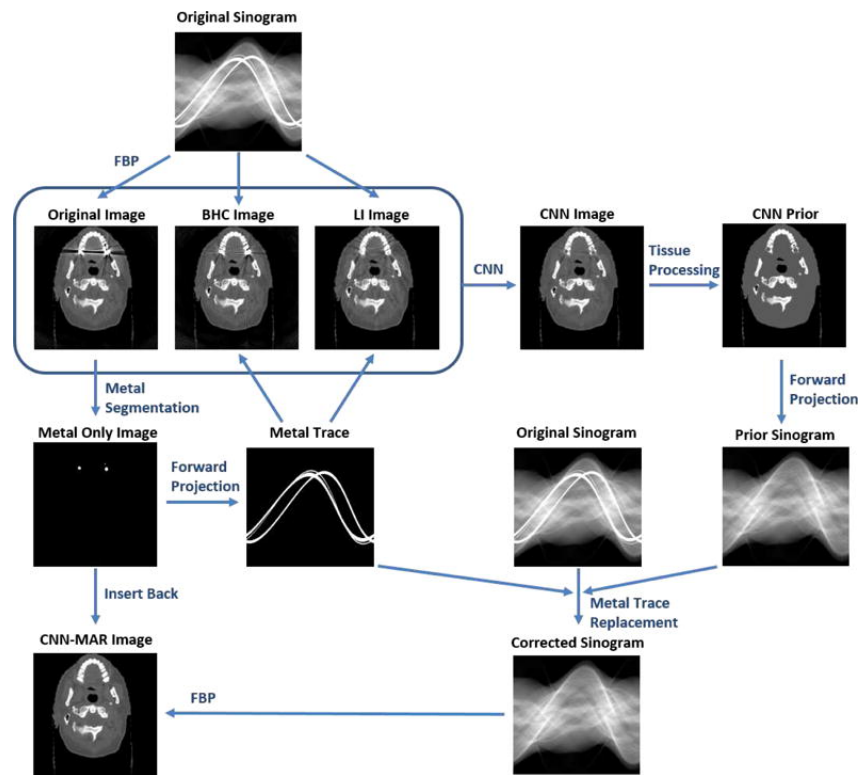
Fig. 4. Architecture of the CNNMAR method [3]

Based on Figure 4, you can see that CNNMAR utilizes a triplet of images as inputs into the CNN:
– The original uncorrected image.
– Image corrected using beam hardening correction (BHC).
– Image corrected using linear interpolation.

The inputs are treated as three-channel images. Small patches extracted from the images are input into the model, with corresponding patches from artefact-free images acting as targets. After training, the image is segmented into bone, water and air. The result is forward projected (FP) into the sinogram domain, combined with the original sinogram, and the metal trace to produce the final corrected sinogram. Then filtered back-projection (FBP) is used to obtain the final image.

Another method that relies on image prior generation and sinogram completion was proposed in [4]. Here the network inputs are fewer – they consist of an LI corrected image and the corresponding metal-affected image. A prior is generated using the U-Net architecture. Then a residual sinogram is obtained by subtracting the prior from the LI corrected sinogram. This, together with the metal trace, acts as an input into the sinogram correction net, which is a pyramid U-Net.

Wang et al. [5] have proposed a dual-domain MAR method that uses two residual U-Nets to process sinograms and images reconstructed from them. The sinogram processing network takes the residual obtained from subtracting an LI corrected sinogram from the original, as well as the metal trace that

is retrieved from the image through thresholding. The output is added to the LI corrected sinogram and projected into the image domain, where it is further processed by the image processing network.

Another example of a sequentially connected dual-domain approach was presented in [6]. Here the sinogram preprocessing model is a generative adversarial network (GAN) with inputs consisting of LI corrected sinograms. The outputs of the network are projected into the image domain and then further refined with the help of a partial convolution U-Net that uses binary metal masks to exclude pixels belonging to the metal regions from the convolution operation.

The paradigm of sequentially connected dual-domain networks is well-established at this point. However, questions arise on whether all of the information available in both domains is exploited with this approach. That is why Wang et al. [7] have proposed a dual-domain parallel network (IDOL-net). A visual is displayed in figure 5.
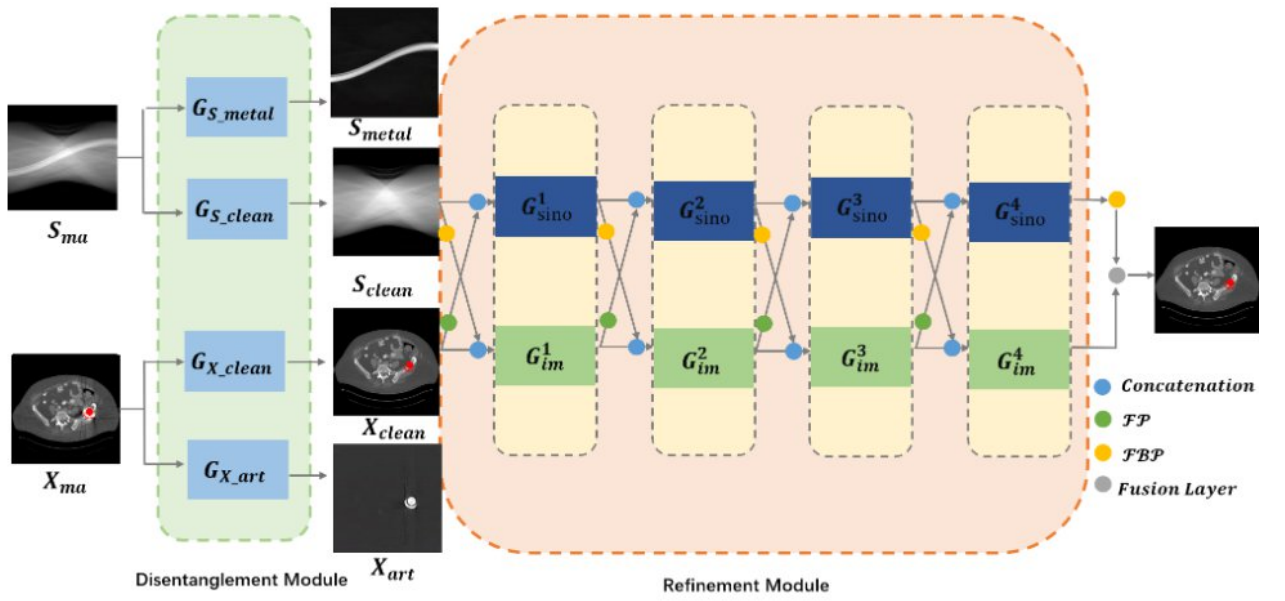
Fig. 5. Architecture of the IDOL-net [7]

As you can see in figure 5, the model consists of a disentanglement and a refinement module. The disentanglement module separates the image from the metal artefacts in both domains, producing two outputs that are further processed in the refinement module. The refinement module consists of two sets of four residual U-Nets that run in parallel. The output of each network is projected into the other domain and concatenated. The final outputs are fused to produce the restored image.

Although most of the recently proposed supervised architectures leverage dual-domain information to obtain gains in performance, this does not imply that single-domain models are obsolete. A pure sinogram domain correction model has been proposed by Zhu et al. [8] as recently as this year. Here, the input sinogram is input into a U-Net that separates the metal trace. It is then preprocessed with LI and subtracted from the original sinogram to obtain a residual. The residual is enhanced with the help of another U-Net and added to the metal trace. The result is projected into the image domain with FBP and fused with the LI-corrected image.

1.4. Unsupervised Metal Artefact Reduction

While the primary area of research for metal artefact reduction involves supervised models, there has been some progress with unsupervised models as well. One of the first such attempts was the artefact disentanglement network (ADN) proposed by Liu et al. [9]. It uses two pairs of generators and discriminators to achieve unsupervised learning for unpaired sets of images. An illustration is presented in figure 6.
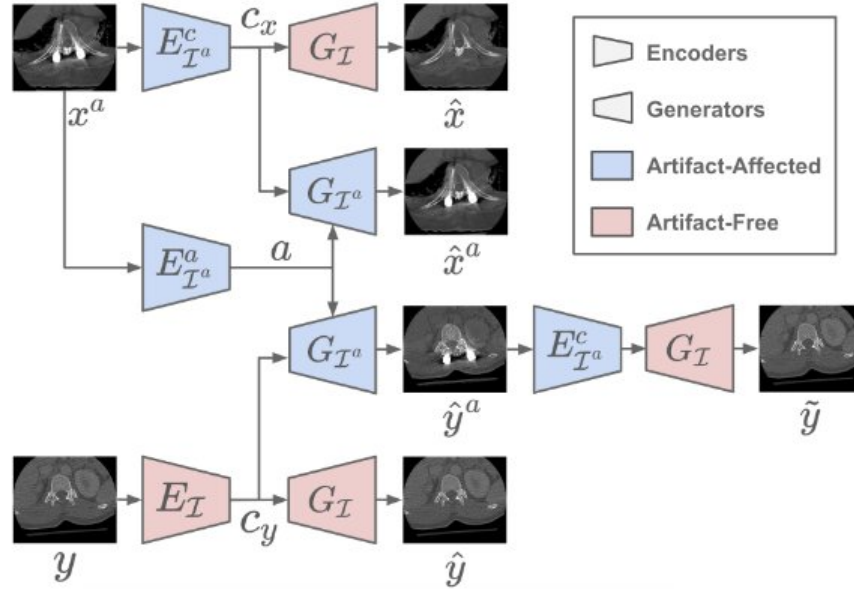
Fig. 6. Architecture of the ADN [8]

As you can see in figure 6, the model produces five outputs. Two of them – $\hat{x}^a$ and $\hat{y}^a$ are used to compute the adversarial losses. One discriminator is trained to differentiate images from sets $\hat{x}$ and y and another from sets x^a and $\hat{y}^a$. The rest of the outputs are used to compute the other components of the loss i.e. the artefact consistency loss, reconstruction loss, and self-reduction loss, which are all necessary to ensure that the model preserves correct image information. The total loss is determined by the weighted sum of the components, which means that optimal weights have to be determined during training as well. The model exhibits some sensitivity to these parameters, with different datasets requiring different values. Typically, the weight assigned to the adversarial losses is 1, while the other losses share a single weight parameter.

The performance of the ADN rivals that of other state-of-the-art supervised and unsupervised methods; however, its architecture is complex, and the resulting model is difficult to train. That is why Zhao et al. [10] have introduced a simplified architecture that does well to preserve model performance. The model structure is presented in figure 7.

Fig. 7. Architecture of the reused convolutional network [10]

As you can see in figure 7, the model consists of a single generator and two discriminators while producing only two outputs. The structure of the loss remains the same as ADN, though. In general, any supervised setup could be transformed into an unsupervised one with this approach. However, as mentioned before, a substantial overhead is incurred due to additional losses that are required to regularize the outputs.

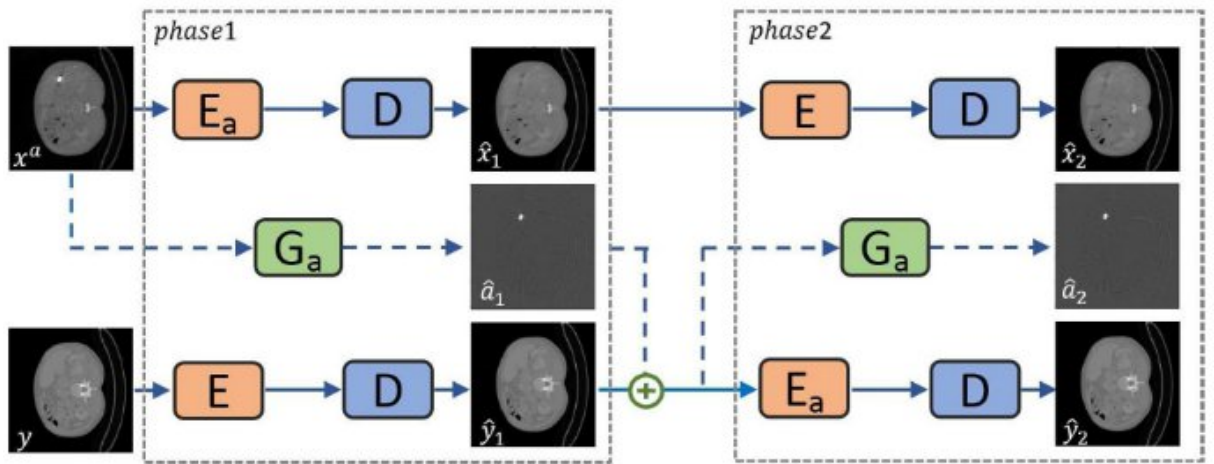Another mechanism that is sometimes used to boost the performance of generative approaches is attention. Attention mechanisms allow the model to focus on specific parts of the input data. They accomplish this by assigning varying weights to different input elements based on their relevance to the task. Two instances of attention utilization are demonstrated in recent studies. In [11], a U-net architecture incorporated attention modules within skip connections to serve as a generator. In [12], a transformer was employed to encode the content information of sinograms. Both approaches exhibit comparable performance to other state-of-the-art methods.

Although most of the unsupervised approaches are single domain, there are a few exceptions. For example the unpaired dual-domain network (U-DuDoNET) [13] utilizes sinogram completion to segment metal artefacts within the image. The obtained result is then projected into the image domain and subtracted from the original noisy image. Further post-processing is applied to achieve the final denoised output.

Although most of the focus in recent years has shifted to deep learning, Zhu et al. [14] have eschewed this and devised a completely analytical physics-based model that does not rely on deep learning. The model inputs consist of a single uncorrected sinogram, which is processed in three ways:
  – The sinogram is projected into the image domain, where the metal trace is extracted and then projected back into the sinogram domain.
  – The sinogram is corrected via the NMAR method.
  – The sinogram is corrected using the physics model devised by the authors.

All three results are then fused to produce the final output, which is later projected into the image domain and post-processed to obtain the restored image. The performance of this model is comparable to that of other state-of-the-art methods.

1.5. Unsupervised Artefact Reduction

Numerous unsupervised denoising techniques exist that, while not explicitly developed for reducing metal artefacts, effectively handle certain forms of unstructured noise, such as Poisson and Gaussian. Consequently, these methods can be incorporated into pipelines for reducing metal artefacts. Some of them are discussed in this chapter.

The Deep Image Prior (DIP) [15] architecture uses untrained deep convolutional neural networks (CNN) to recover clean images from noisy ones. In this approach, a random noise image acts as the model input, while the image to be denoised is set as the target for training. The original image without noise is reconstructed in the course of optimizing model weights. Additionally, the quality of the image is improved by applying smoothing on the outcomes of the final iterations as well as averaging the solutions over multiple different random inputs. This method does quite well in removing Gaussian noise but is otherwise ineffective. However, it is noteworthy because it serves as a basis for several other more advanced models.

The DeepRED [16] method is a modification of the DIP that merges DIP with Regularization by Denoising (RED). The DeepRED refines the initial estimates provided by either Block Matching 3D (BM3D) [17] or Non-local Means NLM [18] with DIP. This fusion of models has been shown to outperform both RED and DIP individually. Nevertheless, its area of application remains primarily limited to dealing with unstructured noise.

The Deep Decoder (DD) [19] is another extension of the DIP that incorporates several simplifications of the underlying CNN structure to improve performance. A visualization is presented in figure 8.
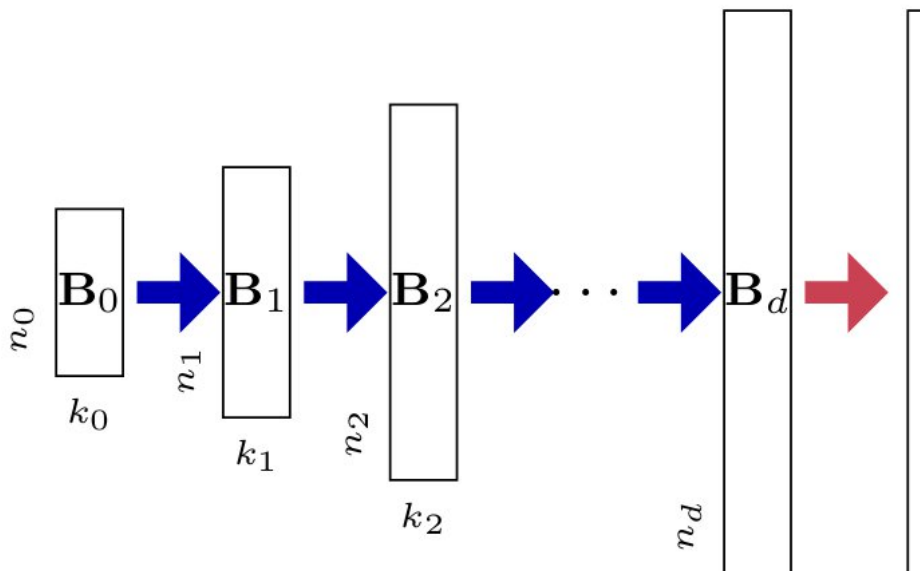
Fig. 8. Architecture of the Deep Decoder [19]

As you can see in figure 8, a random input is upscaled several times through the network. This is achieved by applying pixel-wise linear combinations across channels (sometimes referred to as 1x1 convolutions), channel-wise normalization, upsampling, and ReLU in each layer. It has been shown to make the Deep Decoder more robust to overfitting and choice of hyperparameters compared to the Deep Image Prior.

Noise2Noise (N2N) [20] is another denoising technique that does not rely on clean targets for training. It leverages the fact that different noisy versions of the same clean image share common underlying information. This technique is not tied to any specific model architecture, rather, it is an approach where the inputs and the outputs of the model are two corrupted versions of the same image. The authors conducted tests with BM3D as well as DIP and demonstrated that the performance of these models with N2N is comparable to conventional methods. However, this approach has limitations as obtaining pairs of noisy images with similar content in practice may be difficult.

A modification of the N2N called Noise2Void (N2V) [21] attempts to mitigate the limitations of N2N. This technique, in contrast to N2N, can be applied to problems where neither noisy image pairs nor clean target data are available. This is accomplished by deriving both input and target data from a single noisy image. The process is visualized in figure 9.
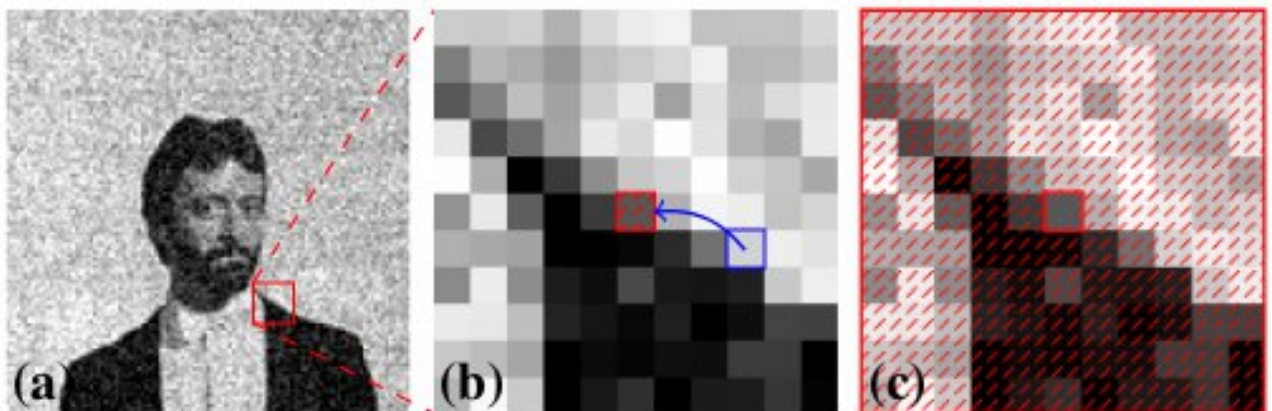
Fig. 9. Blind-spot creation scheme: a) Noisy image b) Input patch c) Target patch [21]

Based on figure 9, you can see that a blind-spot is created in the input patch by the replacing the centre pixel with a random one within the patch. The original unmodified patch acts a target, but the loss is calculated only for the blind-spot pixels. This teaches the model to estimate the centre pixel given the information in the surrounding pixels. N2V shows slightly lower performance in comparison to other approaches such as N2N. However, this is to be expected because N2V has access to less information during model training.

Another method that utilizes patches of the same noisy image for training is Noise2Self (N2S) [22]. It treats the patches within the same noisy image as independent instances of the same patch. The objective is to minimize the difference between the model's output and the original noisy patch. Just like N2N, N2S can be implemented with any type of underlying architecture. When applied to training neural networks, this technique has demonstrated superior performance compared to traditional unsupervised methods like NLM and BM3D, while achieving similar performance to N2N.

1.6. Overview

In summary, I covered a range of techniques for reducing metal artefacts, encompassing both supervised and unsupervised approaches. While supervised approaches are definitely powerful, their reliance on labeled data limits their applicability. Synthetic datasets are often used as substitutes, but they present their own challenges. Since it is difficult to create synthetic datasets that encompass the full spectrum of metal artefacts, the performance of supervised models on synthetic data may not translate well to real-world scenarios. This is where unsupervised methods offer advantages.

Despite substantial research into unsupervised techniques, there are still gaps to be addressed. One of which is exploiting dual domain information in a parallel fashion. While sequential approaches exist for exploiting sinogram information to address secondary artefacts, I could not find any that leverage this information in a parallel manner. Considering that supervised methods utilizing dual domain information have shown improved performance on average, studying unsupervised models that follow the same paradigm is a worthwhile pursuit.

2. Methods

This section introduces the dataset utilized for modelling purposes, along with an architecture specifically designed for metal artefact reduction. The implementation details of this architecture are provided together with details for ADN and RCNN, which were used as baseline models for comparison. The theory of Radon transforms and their role in medical imaging is briefly discussed as well.

2.1. Dataset

As mentioned before, there are two primary types of metal-induced artefacts [1]:
- Beam hardening – dark streaks with adjacent light streaks produced between two objects like bone or metal.
- Poisson noise – random, thin, bright and dark streaks strewn across an image.

Several examples of both are presented in figure 10.
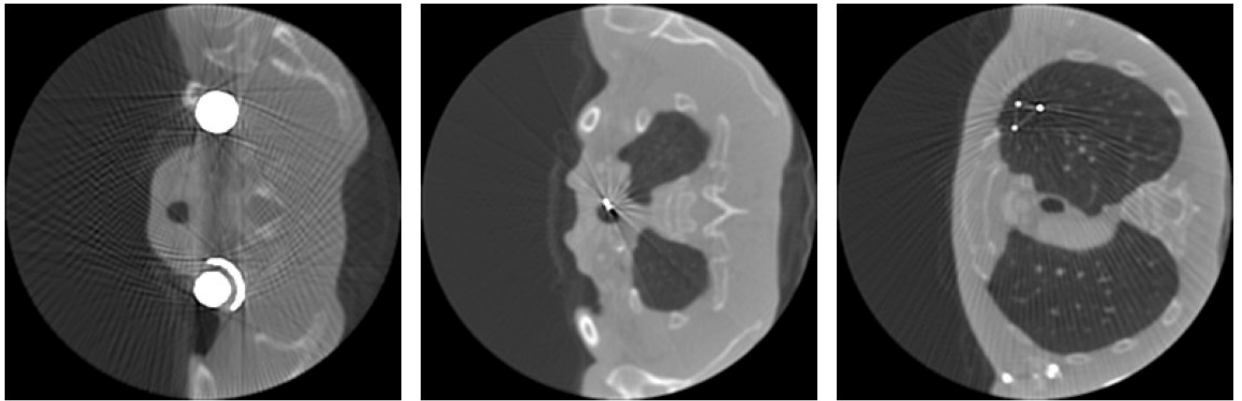
Fig. 10. Examples of metal induced artefacts

For experiments, a dataset of CT images was provided by the Kaunas University of Technology (Department of Applied Physics). The dataset consists of CT image data for three patients. The images are mostly unaffected by artefacts; however, patient three possesses a dental implant, which affects 77 slices. These were put aside for qualitative evaluation. A few examples are given in figure 11.
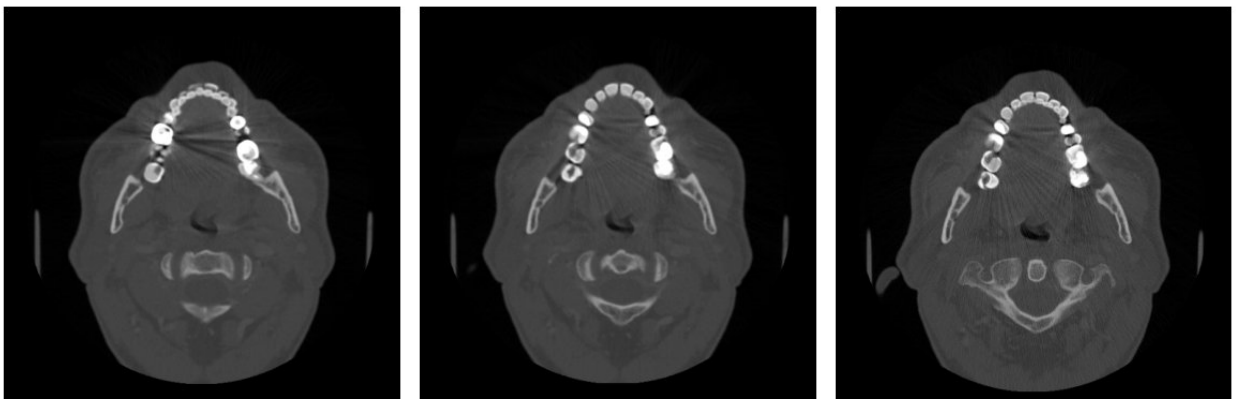
Fig. 11. Examples of naturally occurring artefacts

For the rest, I generate the metallic artefacts myself using the artefact synthesis code provided by the authors of [3]. The code uses metal masks to synthesize different types of metal artefacts. Thirty-six masks are used to produce the training set, while another six are reserved for the test set.

Since the generated dataset is too large for training, I randomly chose 10000 images from the first patient of the dataset for training, another 1000 from the second for validation, and 5000 from the third one for testing.

Model performance is evaluated with two metrics – peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). Distributions of these metrics are presented in figure 12.
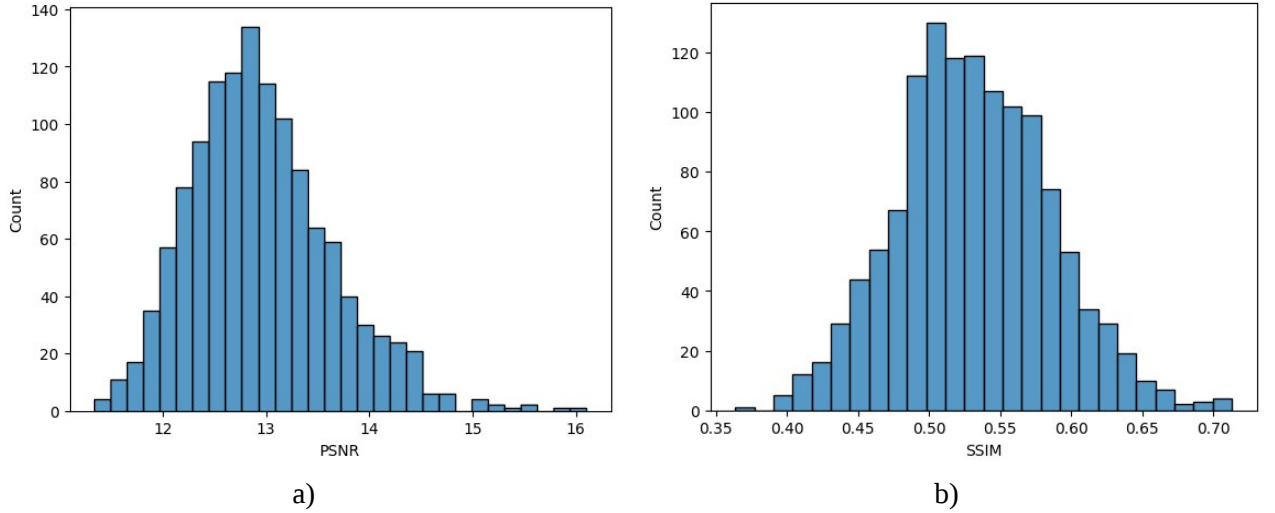

a)   b)

Fig. 12. Distribution of: a) PSNR b) SSIM

The base average values of PSNR and SSIM are 12.95 dB and 0.53. The higher these metrics, the better the model performance. So anything above the baselines is an improvement. The PSNR is calculated in the following way [23]:

$$PSNR\,(I,J) = 10\log_{10}\frac{MAX_I^2}{MSE(I,J)} \tag{1}$$

where I,J are two images and MAX_I is the maximum intensity of pixels in the image I, while MSE is the mean squared error calculated between pixels of the two images. The SSIM is computed by applying a sliding window, where patches of the images x and y are compared [23]:

$$SSIM\,(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{\left(\mu_x^2 + \mu_y^2 + c_1\right)\left(\sigma_x^2 + \sigma_y^2 + c_2\right)} \tag{2}$$

where μ is the pixel sample mean, σ is the variance, and σ_{xy} is the covariance of x and y. Additionally c_1 and c_2 are constants used to stabilize the denominator.

While both metrics capture image quality information, there are differences in how they relate to human perception. Since mean squared error (MSE) is part of the formula for PSNR, it exhibits a bias towards blurry images. On the other hand, SSIM takes into account structural information such as edges and textures, which provides a closer approximation to human perception, but it does a poorer job of estimating actual noise in the image. That is why both of these metrics are necessary to get a

complete picture of model performance. However, no metric is a perfect substitute for human perception, so qualitative evaluation is important as well.

Ideally, I would like to evaluate models strictly on natural images, but the lack of labels enables qualitative comparisons only. On the other hand, synthetic images allow the evaluation of a metric but may not accurately indicate performance in a real-life scenario. So I use both types of data to get a more accurate representation of model behaviour.

2.2. Artefact Disentanglement Network

As previously mentioned, the ADN architecture comprises multiple components, namely encoders, generators, and discriminators. Each of these components consists of blocks depicted in figure 13.
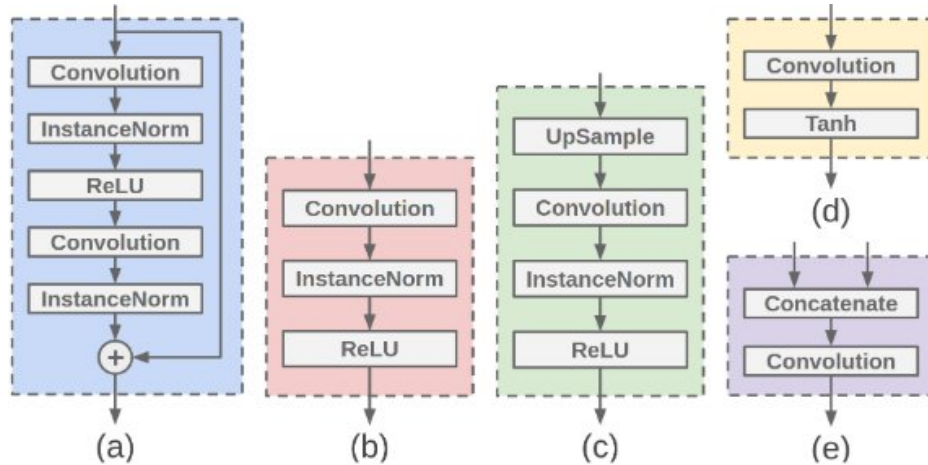
Fig. 13. Components of the ADN: a) Residual block b) Downsampling block c) Upsampling block d) Final block e) Merging block [9]

Furthermore, the artefact generator incorporates artefact code from the artefact encoder to produce artefact-affected images in a distinct manner. This process, referred to as artefact pyramid decoding (APD) in the paper, is illustrated in Figure 14.

Fig. 14. Artefact Pyramid Decoding [9]

As you can see in figure 14, the outputs of the artefact encoder layers are directly integrated into the upsampling process of the generator.

As mentioned before, the total loss of the model consists of five components – two adversarial losses as well as the reconstruction loss, self-reduction loss and artefact consistency loss. Formally, the adversarial losses can be written down as:

$$\mathscr{L} = E\left[\log D\left(y\right)\right] + E\left[1 - \log D\left(\hat{x}\right)\right]$$
$$\mathscr{L} = E\left[\log D\left(x^a\right)\right] + E\left[1 - \log D\left(\hat{y}^a\right)\right]$$
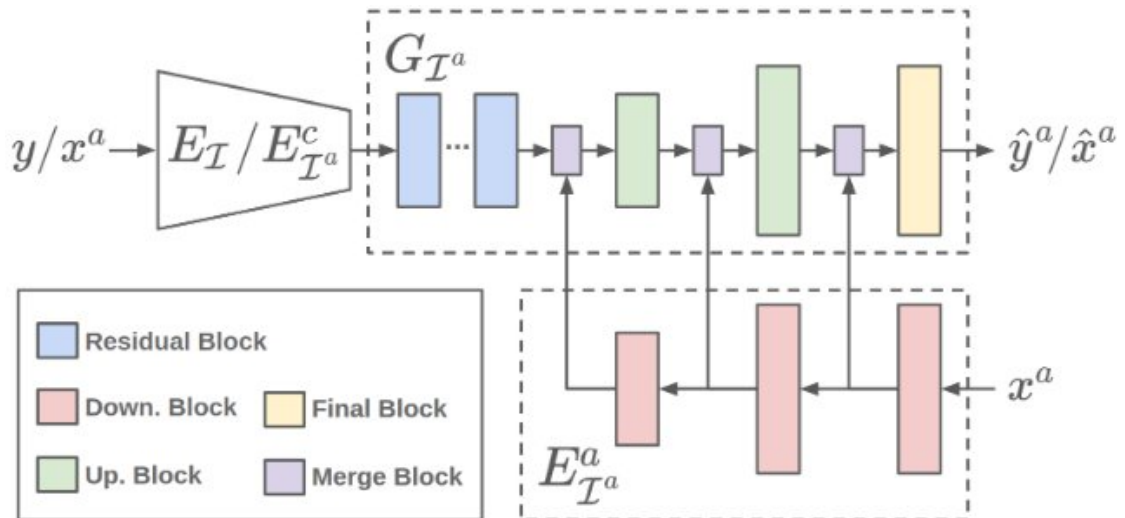
(3)

here y is the artefact-free input, x^a is the artefact-affected input, $\hat{x}$ is the artefact-free output and $\hat{y}^a$ is the artefact-affected output. The adversarial losses ensure that the generators produce realistic outputs.

The reconstruction loss is calculated in the following way:

$$\mathscr{L}_{rec} = E\left\|\hat{x}^a - x^a\right\|_1 + E\left\|\hat{y} - y\right\|_1$$

(4)

here $\hat{x}^a$ is the reconstructed artefact-affected input and $\hat{y}$ is the output of the artefact-free generator. The reconstruction loss guarantees that no information is lost during the process of encoding and decoding the image while also preventing the introduction of any additional artefacts.

The self-reduction loss is calculated as:

$$\mathscr{L}_{self} = E\left\|\tilde{y} - y\right\|_1$$

(5)

here $\tilde{y}$ is the output produced from y when artefacts are added and then removed from the image. The loss serves to improve artefact reduction.

The calculation of the artefact consistency loss is defined as follows:

$$\mathscr{L}_{art} = E\left\|(x^a - \hat{x}) - (\hat{y}^a - y)\right\|_1$$

(6)

It ensures that the model does not introduce additional artefacts by observing that the difference between x^a and $\hat{x}$ should be similar to the difference between $\hat{y}^a$ and y, as they are affected by the same artefact.

2.3. Reused Convolutional Neural Network

The precise implementation details of the RCNN are a lot sparser compared to ADN, so I chose the components as I saw fit. The encoders and decoders, as well as discriminators were reused from ADN. The generator is a standard U-net implemented based on the original paper [24].

The total loss consists of the same components as ADN, with adversarial loss calculated in the same exact way. However, other losses are calculated slightly differently. The reconstruction loss is defined as [10]:

$$\mathscr{L}_{rec} = E\left\|(\hat{x}_1 + a_1) - x^a\right\|_1 + E\left\|(\hat{x}_2 + a_2) - x^a\right\|_1 + E\left\|\hat{y}_1 - y\right\|_1 + E\left\|\hat{y}_2 - y\right\|_1$$

(7)

here x^a is the artefact affected input, while $\hat{x}_1$ is the corresponding cleaned output and $\hat{x}_2$ and a_2 are outputs of the second pass through the network, y is the artefact-free input and $\hat{y}_1$, $\hat{y}_2$ are the results of the first and second pass through the network respectively. As you can see, the main difference

here is that the artefact image in RCNN is constructed explicitly by combining clean outputs with the segmented artefact information.

The artefact consistency loss is evaluated as the L_1 difference between the two artefact segmentation passes:

$$\mathcal{L}_{art} = E \left\| a_1 - a_2 \right\|_1 \tag{8}$$

And the self-reduction loss is calculated as the L_1 difference between the cleaned output $\hat{x}_1$ and the output of the second pass through the network $\hat{x}_2$:

$$\mathcal{L}_{self} = E \left\| \hat{x}_1 - \hat{x}_2 \right\|_1 \tag{9}$$

2.4. Radon Transform

The core premise of this paper is that the performance of existing unsupervised models can be improved by incorporating sinogram information. A sinogram is a visualization of X-ray intensity losses captured by a CT scanner [25]. An example is presented in figure 15.
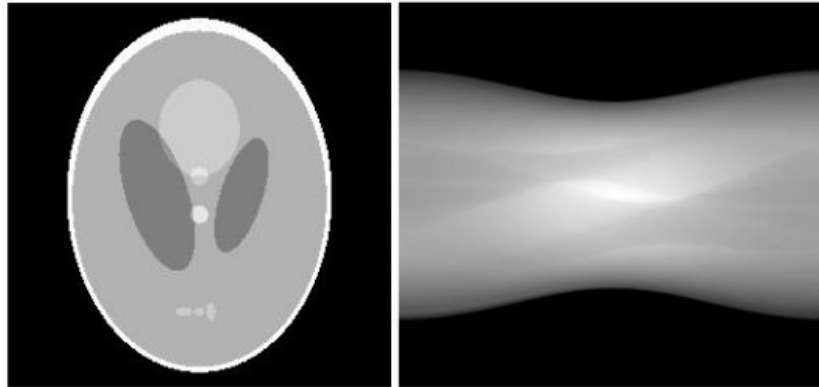
Fig. 15. Shepp-Logan phantom and the corresponding sinogram [25]

A CT scanner does not measure these intensities directly. Instead, it captures the initial intensity at the source and the final intensity once the beam has traversed the object. A pass from a single direction is not enough to reconstruct the image, so the orientation of the source and detector is changed to gain additional information about a slice. The concept is visualized in figure 16.

Fig. 16. The underlying principle of a CT scanner [25]

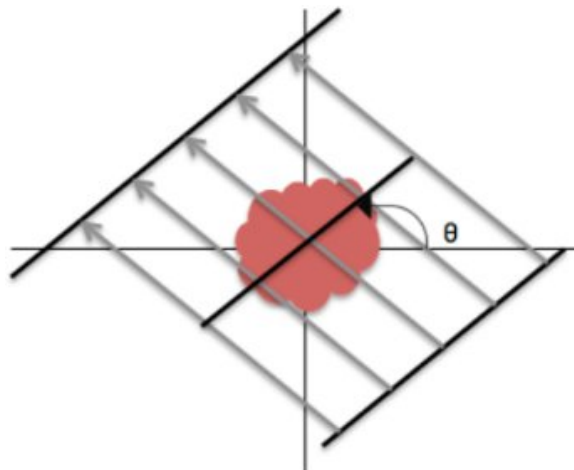Each point in the sinogram corresponds to an intensity difference at a particular distance and angle. The vertical axis represents the distance of the beams from the origin, and the horizontal axis is the angle at which the slice is measured. Moreover, a pixel value of zero corresponds to no change in intensity, while 255 means that the beam was completely absorbed. On its own, the sinogram lacks interpretability and must be first transformed into an image for interpretation.

As mentioned before, each beam is parametrized by two values – the distance from the origin t and the angle of the measurement θ. Given these parameters, a transformation from an image into a sinogram is defined as:

$$Rf(t,\theta) = \int_{-\infty}^{\infty} f(x(s),y(s))ds \tag{10}$$

where $t \in R^2$ and $\theta \in (0,2\pi]$. This transformation is known as the Radon transform. It provides the means to calculate the total density of a function f along a given line. This is illustrated in figure 17.
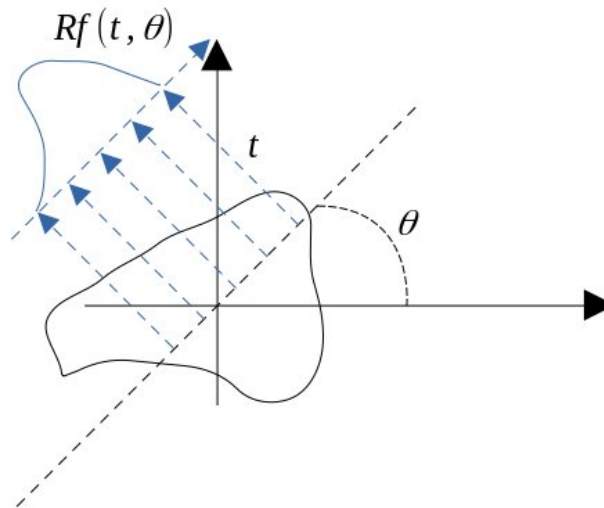

Fig. 17. The Radon transform

The function f in the context of medical imaging is the attenuation coefficient $A(x)$. It shows the size of photon absorption per millimetre of a medium at a distance from the origin x. To reconstruct an image from a sinogram, we need a way to calculate $A(x)$, which requires computing an inverse of the Radon transform. For this purpose, the central slice theorem plays a key role. It relates the two-dimensional Fourier transform to the one-dimensional Fourier transform of the Radon transform:

$$F_2 f(S\cos\theta, S\sin\theta) = F(Rf)(S,\theta) \tag{11}$$

where $S \in R$ and $\theta = [0,2\pi)$. Since we know the definition of the inverse of the two-dimensional Fourier transform, we can compute the original function as:

$$f(x,y) = F_2^{-1}F_2 f(x,y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F_2 f(X,Y) e^{i(xX+yY)}dxdy \tag{12}$$

where $X,Y \in R^2$. For the Radon transform, this formula can be rewritten as:

$$f(x,y) = \frac{1}{2}B(F^{-1}[\,|S|\,F(Rf)(S,\theta)])(x,y)$$

$$Bh(x,y) = \frac{1}{\pi}\int_0^\pi h(x\cos\theta + y\sin\theta,\theta)d\theta$$

$$|S| = det\begin{bmatrix} \dfrac{\partial X}{\partial S} & \dfrac{\partial X}{\partial \theta} \\ \dfrac{\partial Y}{\partial S} & \dfrac{\partial Y}{\partial \theta} \end{bmatrix}$$

(13)

where the operator B is called a back-projection and $|S|$ is called a filter. That is why the method is referred to as filtered back projection (FBP).

The aforementioned formulas are defined for continuous sets; however, in practice, the amount of available data is finite. Therefore, discrete versions of the functions are needed to compute the attenuation coefficients.

Suppose measurements are taken at P different angles θ and that for each angle there are $2 \cdot M + 1$ beams spaced a distance d apart. Then the specific values of θ and t can be defined as:

$$\theta_k = \left\{\frac{k\pi}{P} : 0 \leq k \leq P - 1\right\}$$

$$t_j = \{j \cdot d : -M \leq j \leq M\}$$

(14)

This is enough to define a particular line l_{t_j,θ_k}. The discrete back-projection formula can then be defined as:

$$B_D h(x,y) = \frac{1}{N}\sum_{k=0}^{N-1} h\left(x\cos\frac{k\pi}{N} + y\sin\frac{k\pi}{N}, \frac{k\pi}{N}\right)$$

(15)

where N is the periodicity of the function. As you can see, the formula requires summing over various points in the Cartesian plane, but the locations of the beams were defined in a space parametrized by θ and t. This creates a discrepancy, so the missing data points have to be interpolated. The function to be interpolated is the argument of the back-projection, which in discrete space is defined as:

$$f(x,y) = (F_D^{-1}S) * (R_D f)(jd,k\pi/N)$$

(15)

where F_D and R_D are the discrete versions of Fourier and Radon transforms and $*$ denotes the convolution operation. Since the particulars of the interpolating function are not relevant, we can simply denote it as I. The discrete filtered back-projection can then be defined as:

$$f(x_m,y_n) \approx \frac{1}{2}B_D\left(\left(F_D^{-1}S\right) * (R_D f)\right)\left(jd, \frac{k\pi}{N}\right) \approx \frac{1}{2}B_D I(x_m,y_n)$$

(16)

The final equation addresses the discrete nature of data and fixes issues that arise due to that with interpolation.

2.5. Hounsfield Scale

As mentioned in the previous section, each material has a different attenuation coefficient associated with it. Attenuation coefficients associated with metals are very high compared to typical human tissue or bone, which makes metallic objects easy to distinguish. Typically tissue densities in images are expressed through the Hounsfield unit scale (HU) [26]. The relationships between Hounsfield units and the attenuation coefficients measured by a CT scanner can be expressed as:

$$HU = p \cdot m + b \tag{17}$$

where p is the pixel value, m is the intercept, and b is the offset. The intercept and the offset are parameters of the CT scanner and their values are usually equal to 1 and -1024 respectively.

The Hounsfield scale is a standard CT measurement scale that varies for different tissues. The densities of water and air correspond to 0 HU -1000 HU, respectively. Less dense structures like soft tissue will have values closer to zero, e.g., 0 to +100. Denser structures like bone will have values of 1000 to 2000 HU. Metals are even denser, with values greater than 2000 HU. Different metals have different densities and so different HU values. These values for some metals are presented in table 1 [27].

Table 1: HU values for various metals

Type of Metal	HU
Aluminum	2100
Copper	14000
Steel	20000
Gold	30000

Titanium and stainless steel implants are the most prevalent, but other metals can appear too especially in forensic applications. However, it is crucial to note that these values can vary based on the particular CT scanner and imaging parameters used.

To underline different structures in the image, a technique called windowing is used. First, the values of the image are clamped to:

$$\max = c + \frac{w}{2}$$
$$\min = c - \frac{w}{2} \tag{18}$$

where w is the width of the window and c is the centre of the window, often also referred to as a level. This means that any values below the minimum are set to the minimum, and any values above the maximum are set to that maximum. How different windows highlight different tissues within an image is illustrated in figure 18.
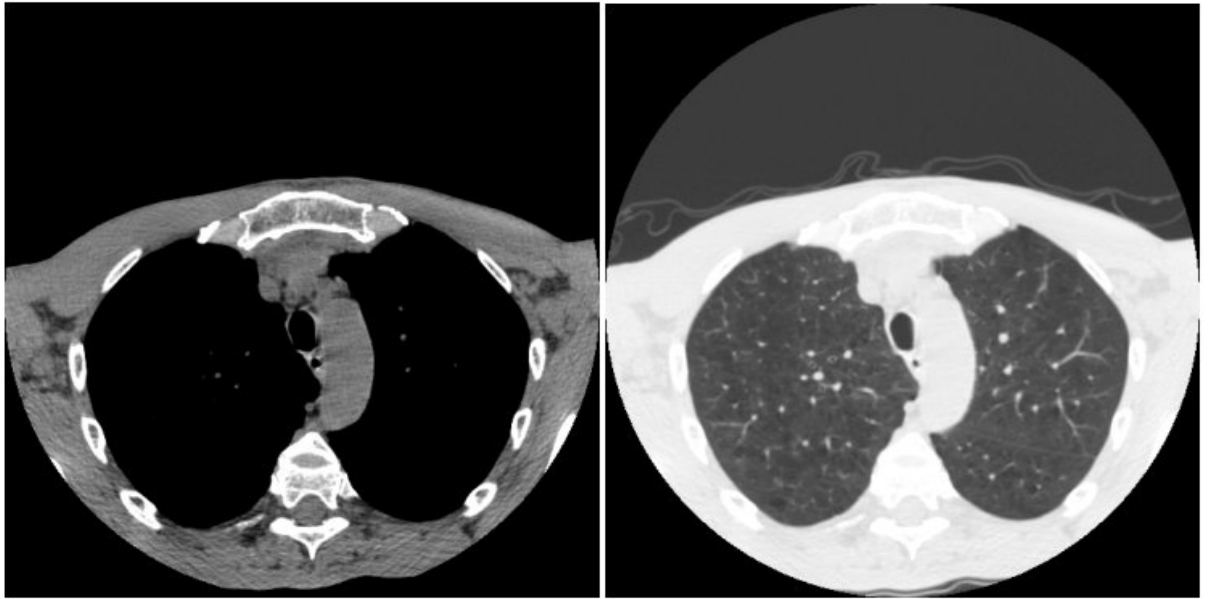
Fig. 18. Influence of windowing on perception

For the image on the left, a window centred around 50 with window width of 350 is used, while for the image on the right, a centre of -600 and a width of 1400 is used. As you can see, lung tissue is much clearer in the second image.

2.6. Proposed Modification

The architecture proposed in this paper is similar to ADN and RCNN; however, unlike ADN and RCNN it integrates sinogram domain information with image domain information to boost performance. This is achieved by executing two versions of RCNN simultaneously in parallel – one for sinograms and the other for corresponding images. The concept for a generic module is illustrated in Figure 19.
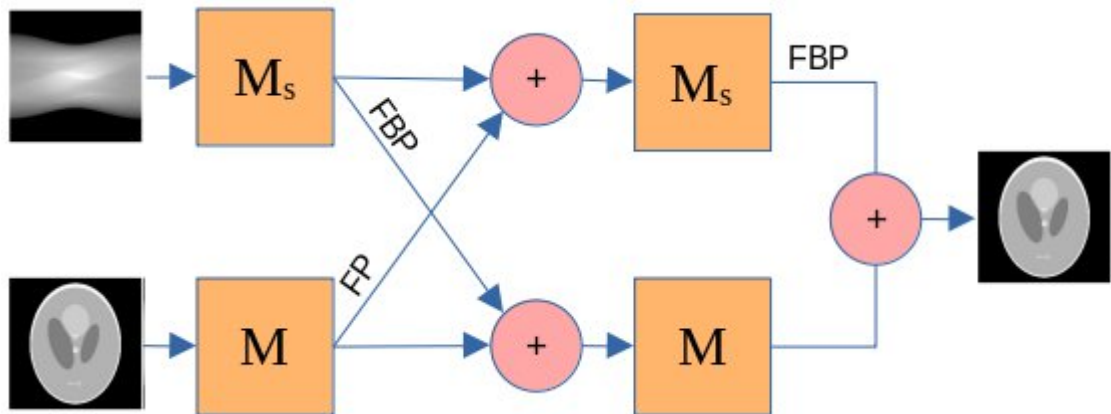

Fig. 19. An illustration of processing image and sinogram information in parallel

Figure 19 depicts the process where the module outputs are projected into the opposite domain. Images are projected with forward projection (FP), and sinograms are projected using FBP. The results are added up with the outputs of the corresponding image or sinogram model. At the very end, the outputs are combined to produce the final image.

3. Results

In this section, the results for each model are presented. Since the dataset and training split have already been described in section 2.1, this information is not repeated here.

3.1. Artefact Disentanglement Network

ADN [2.2] was trained with the standard hyperparameters described in [9] except for the loss weights. As per the paper, the weight for the adversarial losses was kept constant at 1, while the other losses shared a weight parameter. Three values for the weight parameter were tested by training the model for five epochs. The results are presented in figure 20.
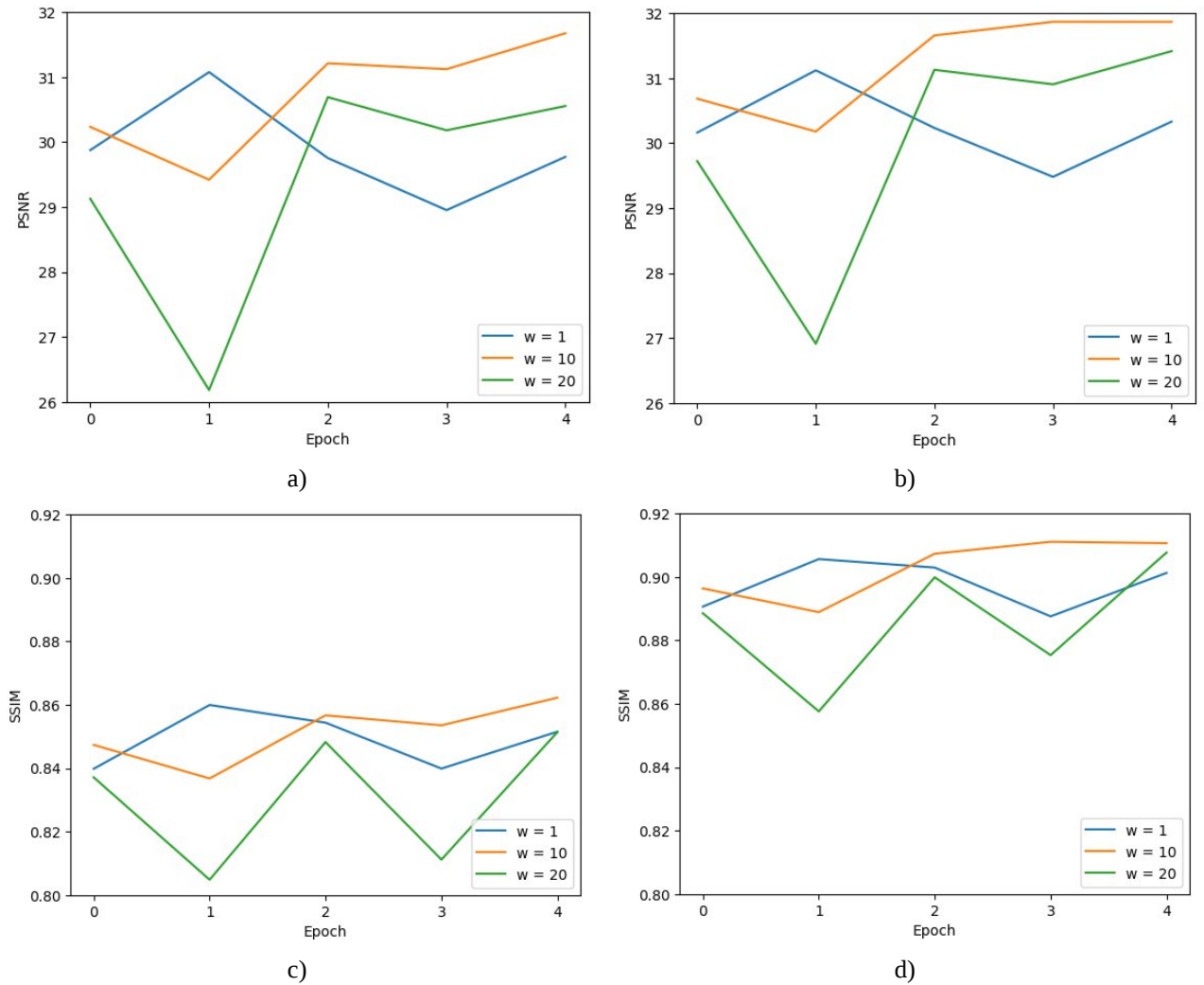

Fig. 20. Results of parameter tuning for ADN: a) Training PSRN b) Validation PSNR c) Training SSIM d) Validation SSIM

As you can see in figure 20, the model with a weight set to 10 stands out as an outlier in terms of performance. On the other hand, the PSNR values for the other two weights appear to be similar, although the SSIM for the model with weight value of 1 is lower. This suggests that it might introduce a greater degree of image blur and potentially struggle to preserve edge information. In order to investigate this further, a comparison of model outputs is provided in Figure 21.
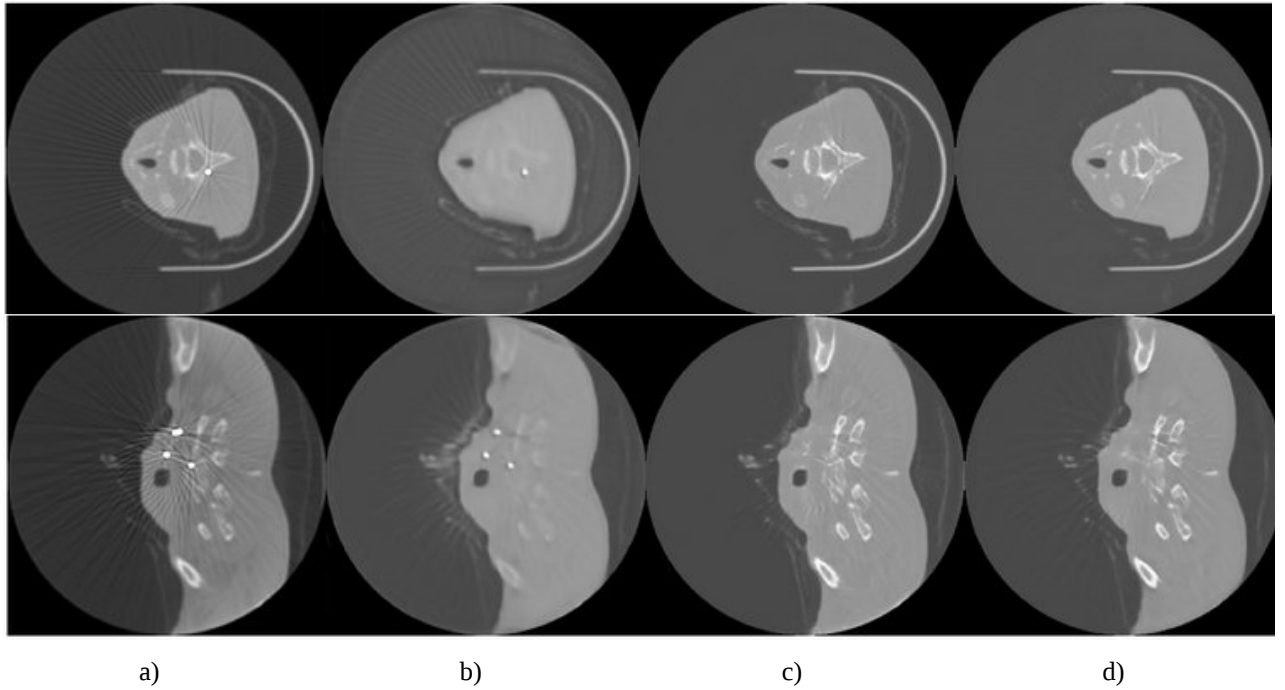
a)   b)   c)   d)

Fig. 21. ADN outputs for different weight parameters: a) Original image b) 1 c) 10 d) 20

Based on figure 21, you can tell that the initial assessment was correct, and setting the weight to 1 does not preserve enough structural information in the image. On the other hand, there does not seem to be a huge qualitative difference between weights 10 and 20, so to decide between them, we can look at their loss progressions. They are presented in figure 22.
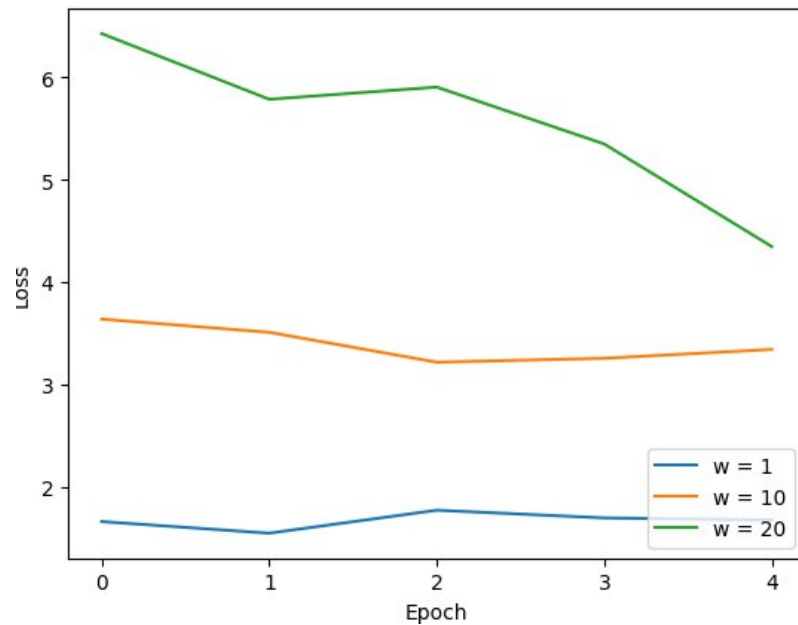

Fig. 22: ADN losses for different weights

As you can see in figure 22, the loss for the model with weight 20 behaves the most desirably because it decreases consistently. The actual values of the losses do not matter in this plot since the losses themselves are not normalized, so losses for models with larger weight parameters will be naturally higher. Based on what was discussed, the weight value of 20 seems to be the correct choice.

Next, to better understand overall performance of the model, the histograms for PSNR and SSIM are looked at. These are presented in figure 23.
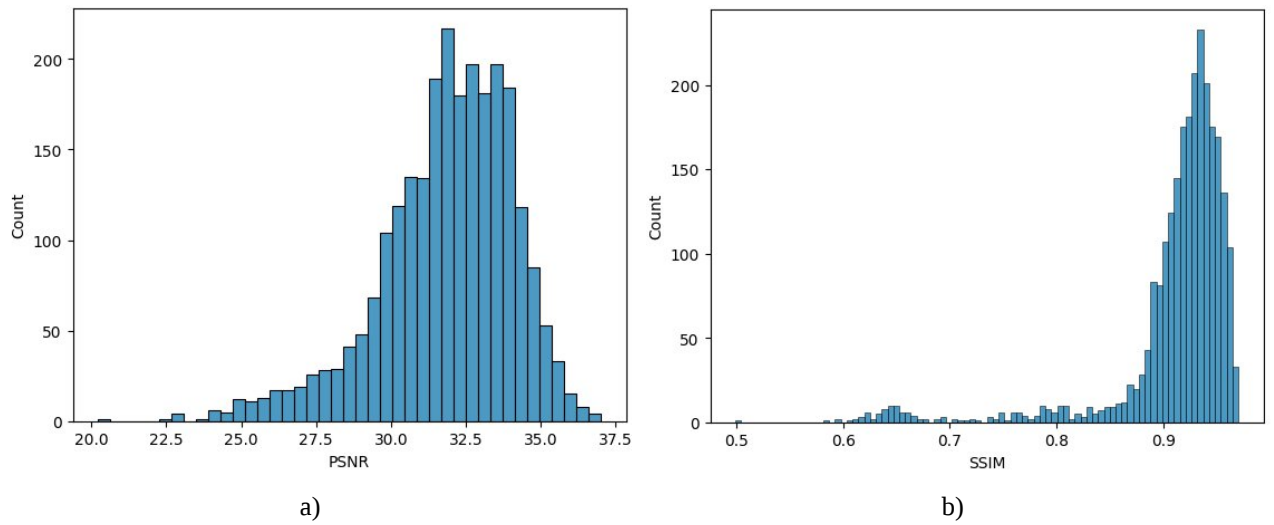

a)   b)

Fig. 23. Distribution of: a) PSNR b) SSIM for ADN

Based on figure 23, you can see that the model performance is not entirely consistent. There seems to be a at least a small subset of images that the model performs relatively poorly on. To understand why that is, some examples of ADN outputs with PSNR lower than 25 and SSIM lower than 0.5 are presented in figure 24.
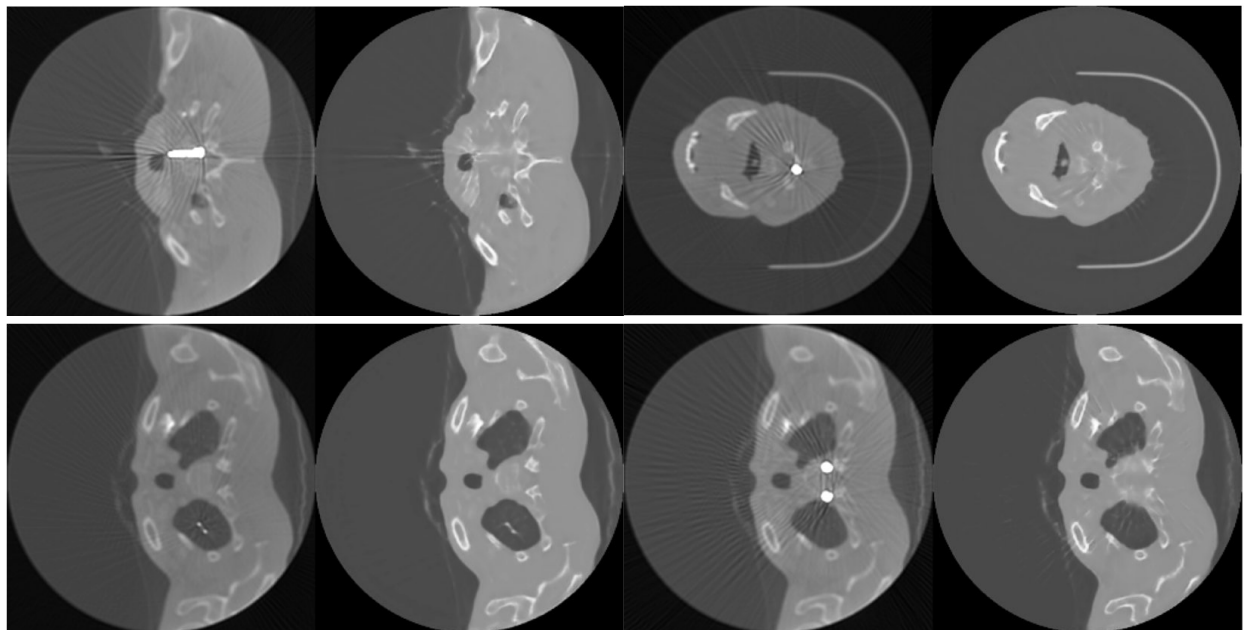

Fig. 24. Examples of poor ADN outputs

As you can see in figure 24, even in the worst case scenario the model has no issues dealing with beam hardening artefacts. However, in some cases, the model struggles to completely mitigate Poisson noise, resulting in residual smearing near the metallic implant area. Nonetheless, the model accurately reconstructs the structural details of the images without excessive loss of edge information.

3.2.   Reused Convolutional Neural Network

For RCNN [2.3], same as ADN, three values for the weight parameter were tested by training the model for five epochs. The results are visualized in figure 25.
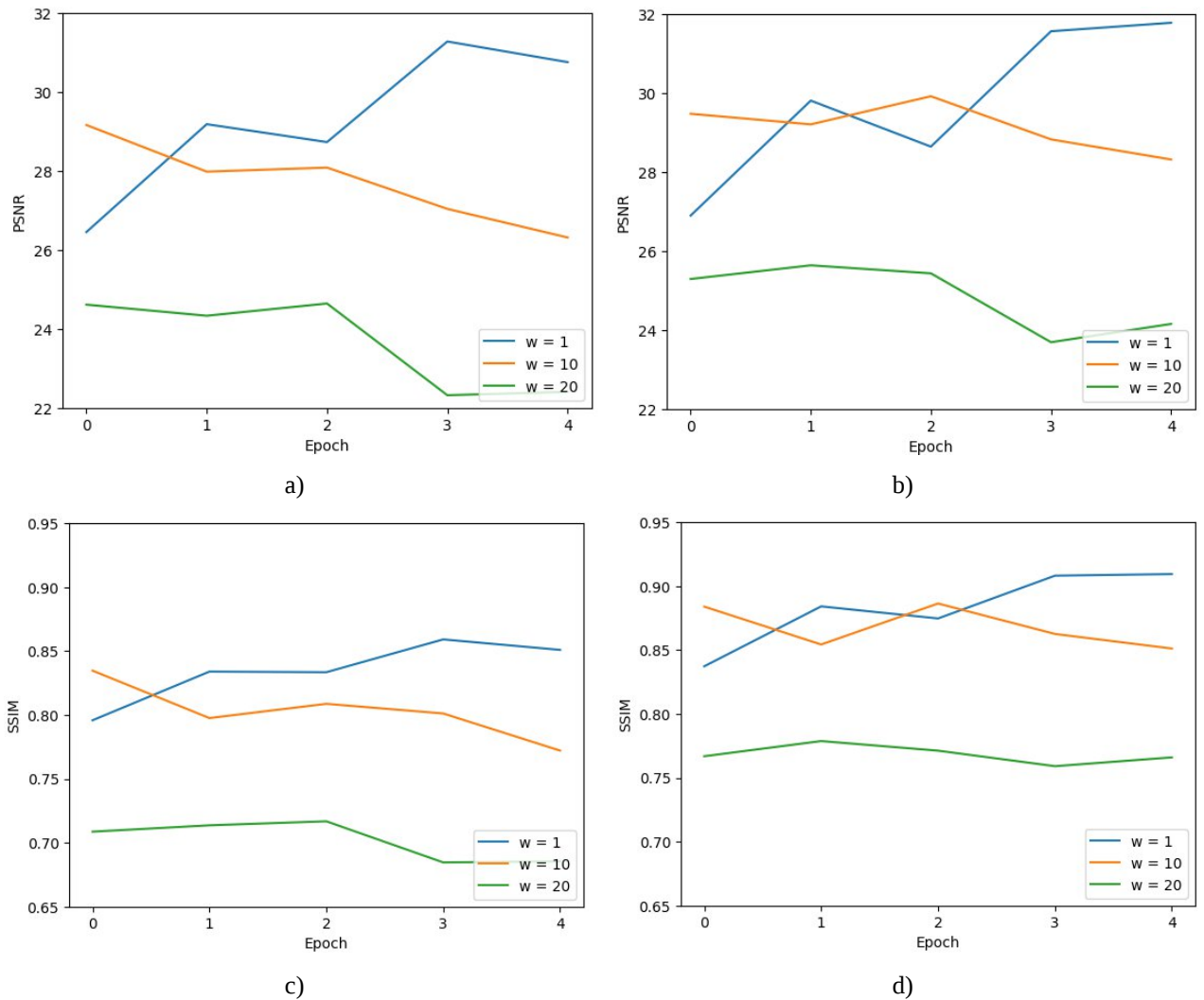

Fig. 25. Results of parameter tuning for RCNN: a) Training PSRN b) Validation PSNR c) Training SSIM d) Validation SSIM

As you can see in figure 25, the model performs best, in contrast to ADN, for smaller weight values, with optimal performance achieved for the weight value of 1. The performance for weight 1, however is quite close to performance for weight 10, with weight 20 being the obvious underperformer. To further examine the differences between these models, we can compare their results qualitatively. The outputs of these models are presented in figure 26.
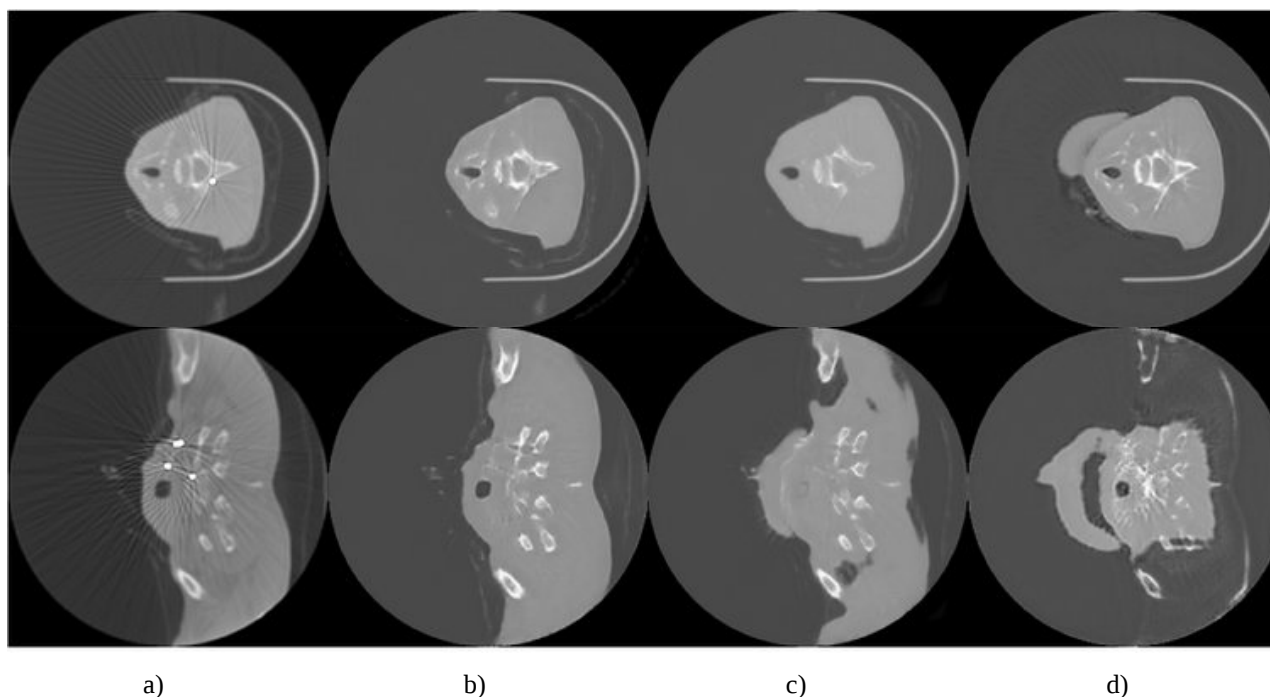
Fig. 26. RCNN outputs for different weight parameters: a) Original image b) 1 c) 10 d) 20

Figure 26 support previous findings with images for the weight value of 1 being the most clear. On the other hand, weights 10 and 20 fail to even reconstruct the images properly. The losses for different weights are presented in figure 28.

Fig. 27. RCNN losses for different weights
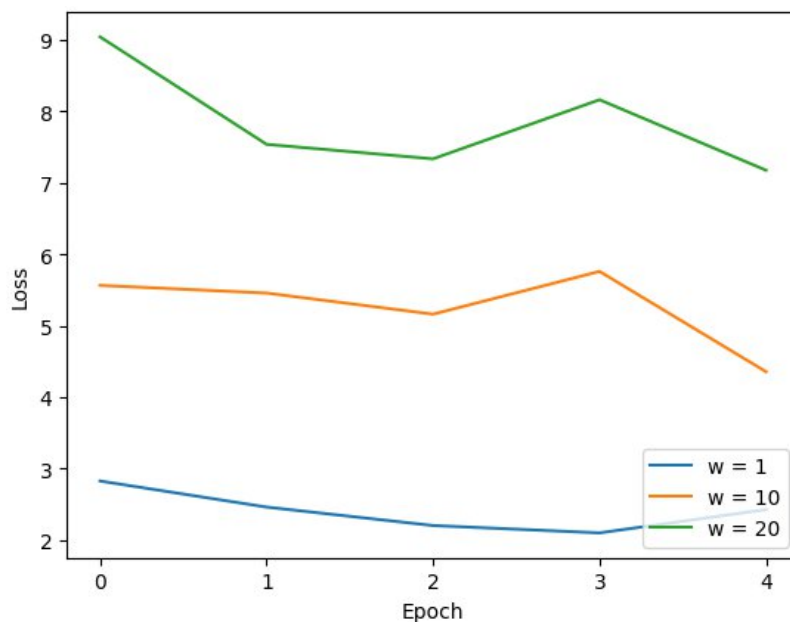
As you can see in figure 27, the difference between loss progressions for weights 10 and 20 is small. The decrease for weight 1 is the most consistent increasing only in the final epoch, so it is unsurprising that it has the best performance.

As before, to better understand performance of the model, the histograms for PSNR and SSIM are examined. These are presented in figure 28.
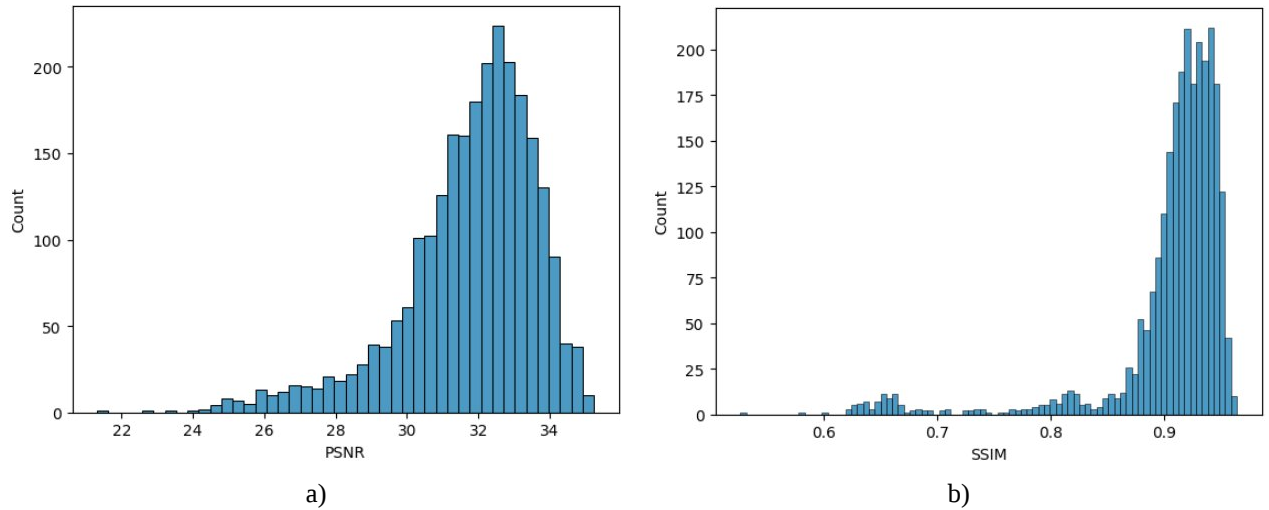
Fig. 28. Distribution of: a) PSNR b) SSIM for RCNN

Based on figure 28, you can see that the metric distributions for RCNN are very similar to those of ADN, e.g., there are several images for which the model exhibits poor performance. These are examined in figure 29.

Fig. 29. Examples of poor RCNN outputs

As you can see in figure 29, the deficiencies in RCNN images are much more subtle compared to ADN. While the RCNN model successfully eliminates beam hardening artefacts, some remnants of Poisson noise remain. Furthermore, certain reconstructed features within the images exhibit blurred edges, lacking the desired level of sharpness.

3.3. Radon Enhanced Reused Convolutional Neural Net

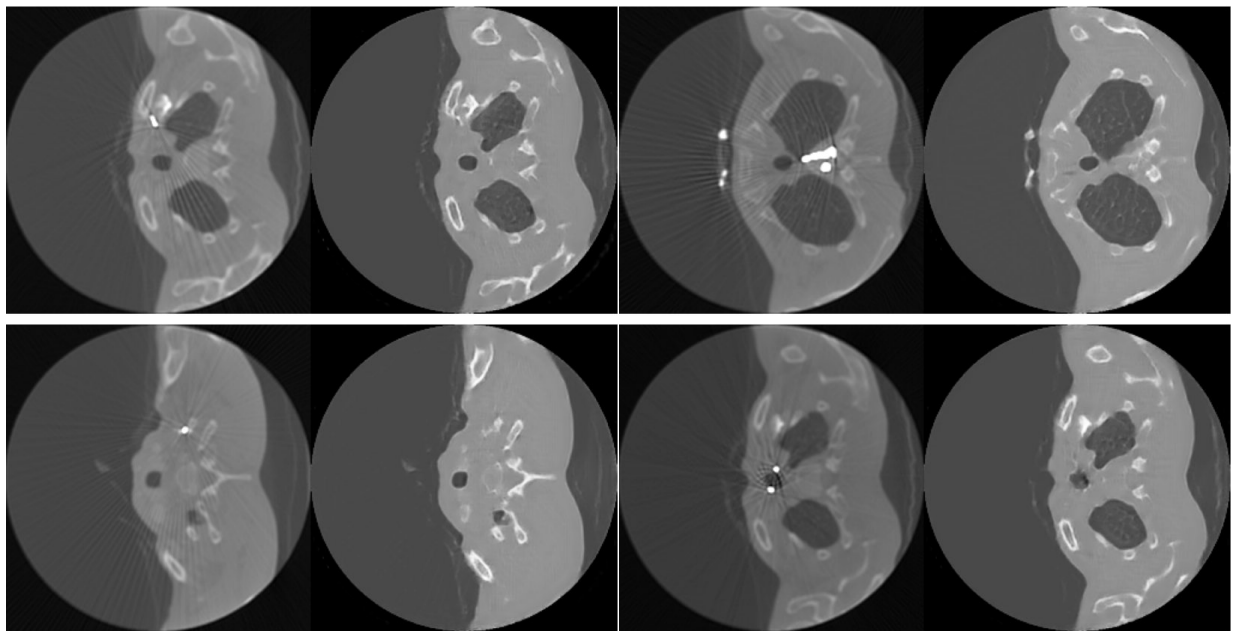Since this model [2.6] is a variation of RCNN, and RCNN performed best for the value of 1, it was tested only on that value for direct comparison. The plots for training and validation PSNR and SSIM are presented in figure 30.

a)   b)
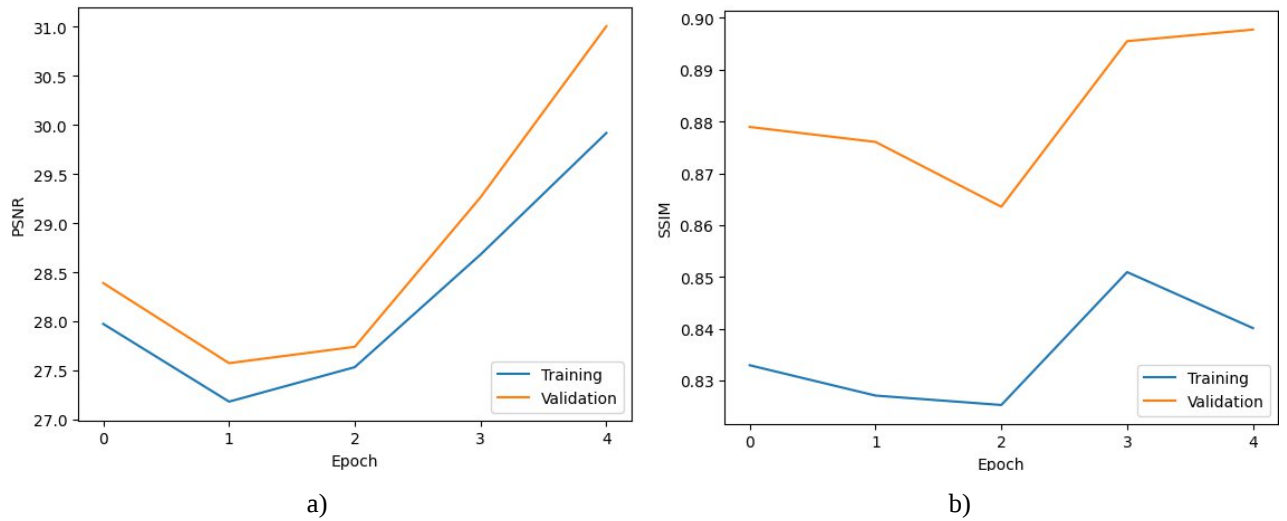
Fig. 30. Training results for Radon enhanced RCNN: a) PSNR b) SSIM

As you can see in figure 30, both PSNR and SSIM show fairly consistent improvements with peak values of 31 and 30 PSNR of training and validation sets, while SSIM for the validation set reaches 0.9 and SSIM for the training set ends up being 0.84. The model is examined from a qualitative perspective in figure 31.
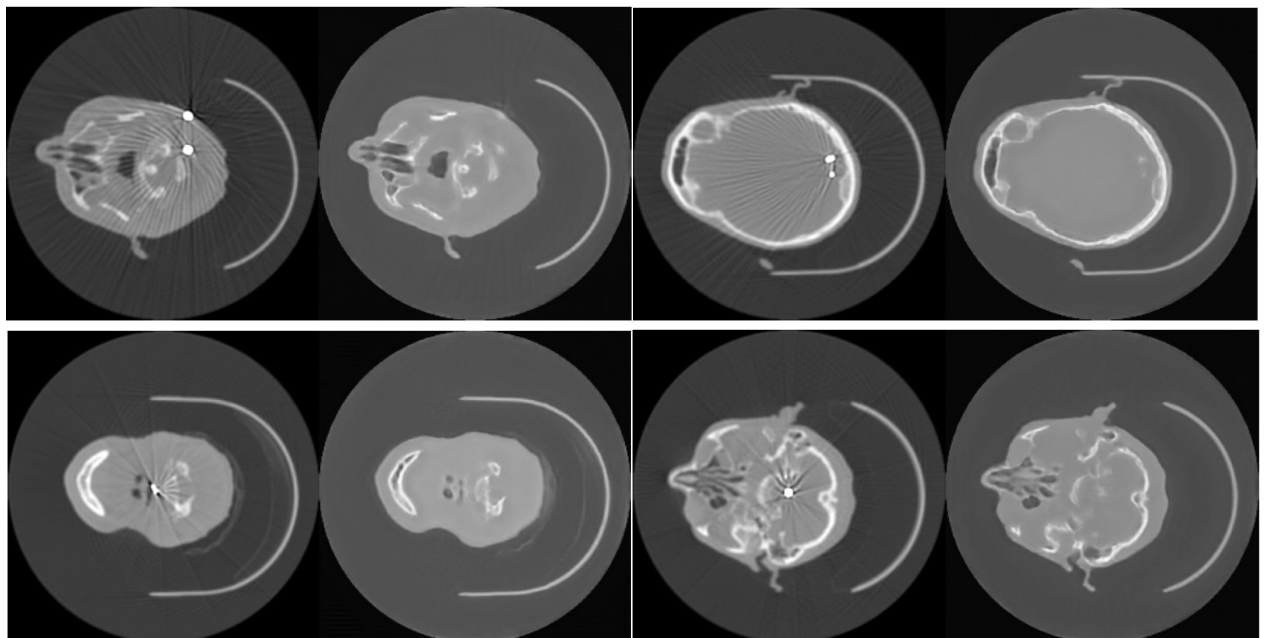
Fig. 31. Radon RCNN outputs

Based on figure 31, you can see that this model has no problems dealing with Poisson noise and beam hardening artefacts, however, in some images e.g. top right, vestiges of the metallic implants remain.

a)   b)
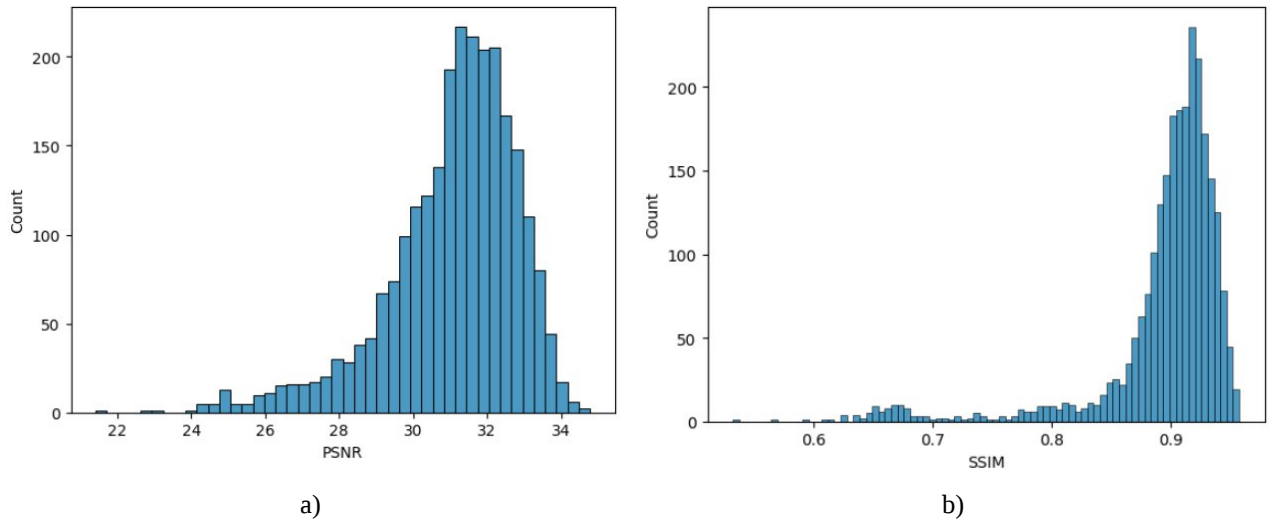
Fig. 32. Distribution of: a) PSNR b) SSIM for Radon RCNN

As you can see in figure 32, the plots are nearly identical to those of RCNN. Some examples of low quality outputs are presented in figure 33.
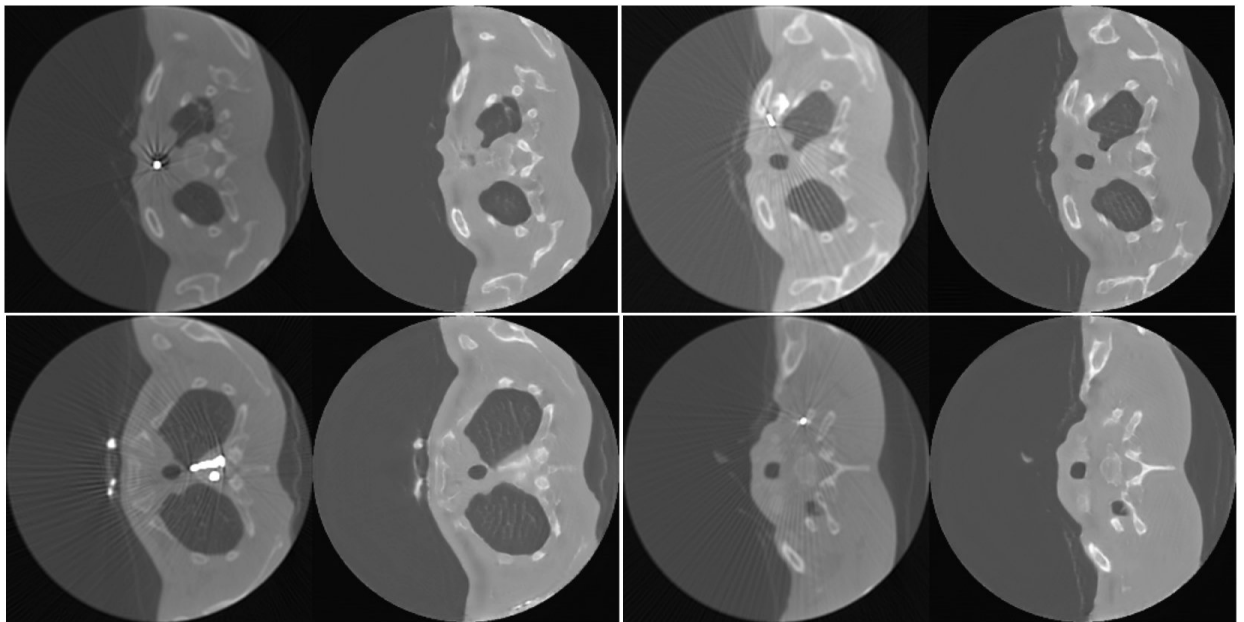
Fig. 33. Examples of poor Radon RCNN outputs

Based on figure 33, you can see that Radon RCNN faces difficulties in completely eliminating Poisson noise, and although it successfully removes beam hardening artefacts, there are instances where residual traces of metallic implants remain visible.

3.4. Overview

Finally, all three methods were compared on the test set. The results are presented in table 2.

Table 2: Test results

Model	PSNR	SSIM
ADN	31.9455	0.9152
RCNN	31.8814	0.9105
Radon RCNN	31.0629	0.8952

As you can see in table 2, all three models have comparable performance, with ADN slightly outperforming the rest. However, when quantitative measurements are this close, qualitative factors are a better way to compare model performance. Figure 34 displays the synthetic data based performance comparison.

a)	b)	c)	d)

Fig. 34. Model performance on synthetic image data: a) Original image b) ADN c) RCNN d) Radon RCNN

Based on figure 34, you can see that the outputs of RCNN contain less noise and have slightly sharper features on average.

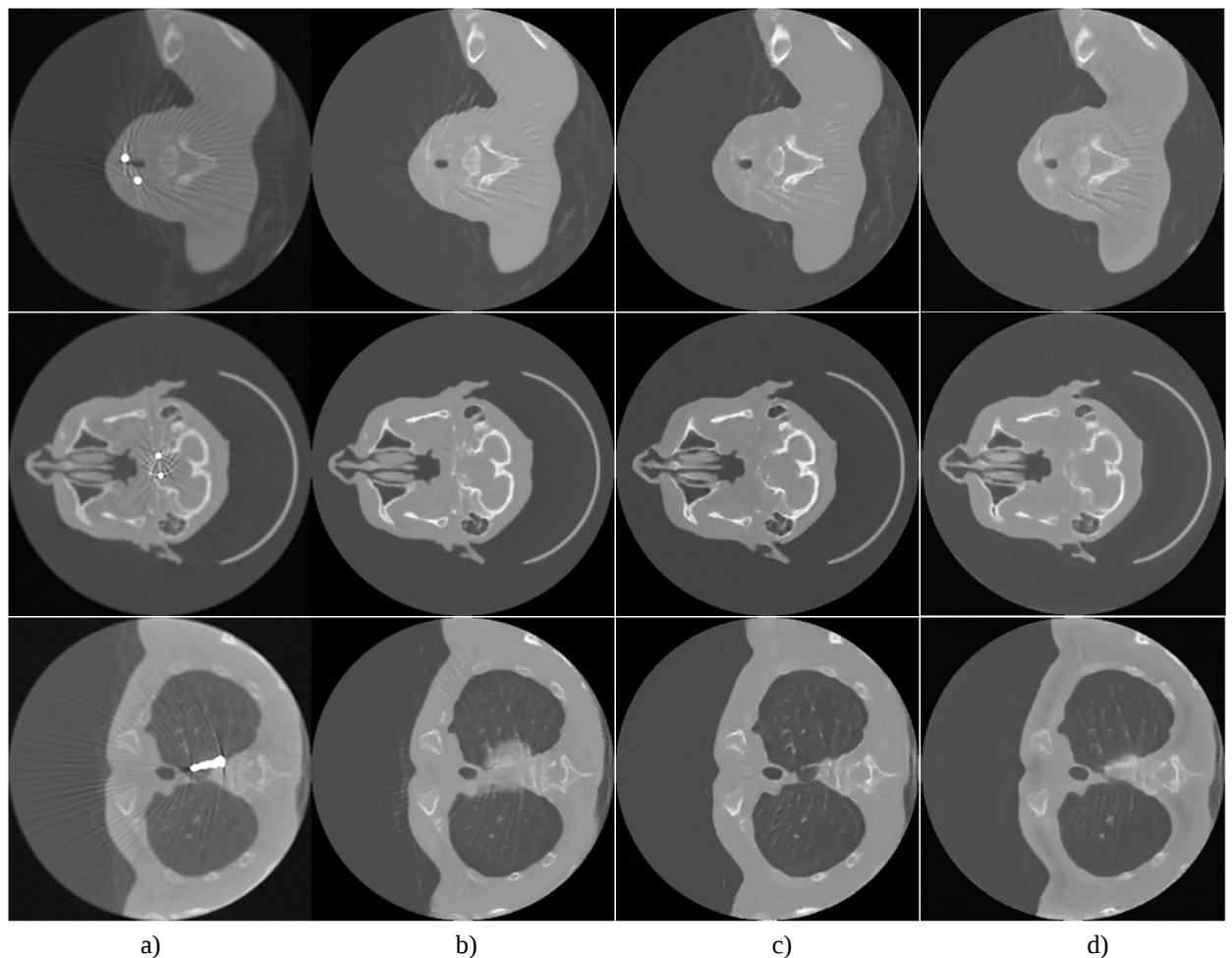To assess model performance on real world data, their outputs for naturally noisy images of the dataset are compared in figure 35.

a)   b)   c)   d)
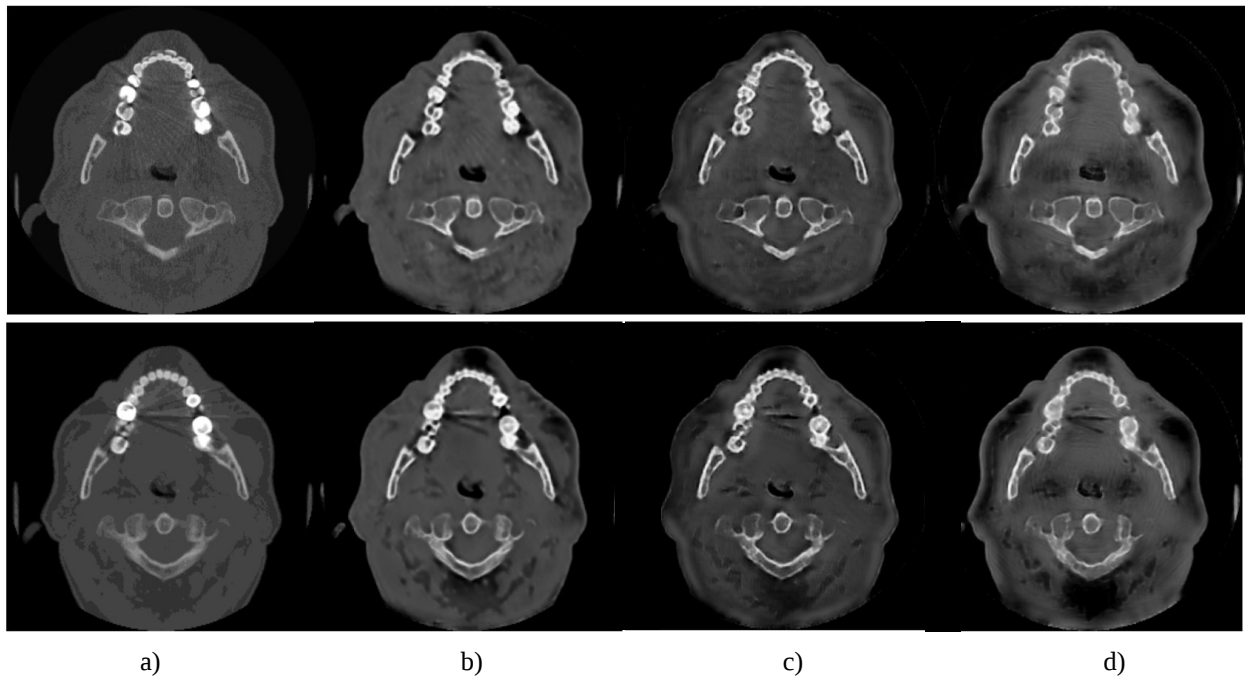
Fig. 35. Model performance on natural image data: a) Original image b) ADN c) RCNN d) Radon RCNN

As you can see in figure 35, the performance of the models on synthetic images does not extrapolate well to natural images. Among the models, RCNN consistently demonstrates the best performance. In the case of the first image, ADN struggles to completely eliminate Poisson noise and beam hardening artefacts, resulting in excessive blurring of the image. Radon RCNN, on the other hand, successfully removes all artefacts but introduces an additional ring-like artefact in the centre while also causing some tissues to smear together. RCNN stands out by not only effectively removing all noise but also preserving the details of the image. However, there is still some additional smearing of the surrounding tissues.

For the second image, none of the three models fully remove the beam hardening artefacts. Nevertheless, ADN produces the highest quality reconstruction for the image, while Radon RCNN once again introduces an additional ring-like artefact in the centre.

3.5. Limitations and Recommendations

To effectively denoise natural images using models trained on synthetic data, it is necessary to employ large-scale models trained over numerous iterations. For instance, in the ADN paper, the model was trained on a dataset of 350,000 images for 50 epochs. Unfortunately, replicating such extensive training is not feasible in this context, which may account for the relatively lower quality results observed with natural images. To conduct a comprehensive comparison and study of real-world applications, training larger models would be essential.

Moreover, the full potential of unsupervised models remains unexplored in this study due to the limited availability of data. One notable advantage of unsupervised models is their ability to be trained on natural images. Intuitively, the quality of a model relies on the quality of the dataset it is trained on. Therefore, it would be worthwhile to explore the possibility of training purely on natural image data, thereby leveraging the potential of unsupervised models to their fullest extent.

Conclusions

1. A comprehensive review of the literature on artefact reduction was conducted, which allowed me to understand the current trends and shape the focus of research in this paper.
2. A synthetic CT dataset was successfully generated, which allowed quantitative model evaluation.
3. The incorporation of sinogram information into the RCNN model did not yield a performance improvement, with the constructed model having the lowest PSNR and SSIM out of the three tested models. Furthermore, it was found that the modified model introduced additional secondary artefacts when applied to natural images. I can conclude that, at least in this case, combining sinogram and image domain information in the proposed fashion is ineffective.
4. ADN achieved the highest quantitative metrics, with a PSNR of 31.9455 dB and an SSIM of 0.9152. RCNN followed closely behind with a PSNR of 31.8814 dB and an SSIM of 0.8995. On the other hand, Radon RCNN performed the poorest among the models, lagging behind RCNN by 0.8185 in terms of PSNR and 0.0153 in terms of SSIM. However, qualitatively, RCNN demonstrated superior performance compared to the other models. It generated outputs with reduced noise while preserving a larger amount of the original feature information.

List of References

1. BOAS, F. Edward, and Dominik FLEISCHMANN. CT artifacts: causes and reduction techniques. *Imaging in Medicine* [online]. 2012, 4(2), 229–240 [viewed 23 May 2023]. ISSN 1755-5205. Available from: doi:10.2217/iim.12.13

2. MEYER, Esther, et al. Normalized metal artifact reduction (NMAR) in computed tomography. *Medical Physics* [online]. 2010, 37(10), 5482–5493 [viewed 23 May 2023]. ISSN 0094-2405. Available from: doi:10.1118/1.3484090

3. ZHANG, Yanbo, and Hengyong YU. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE Transactions on Medical Imaging* [online]. 2018, 37(6), 1370–1381 [viewed 23 May 2023]. ISSN 1558-254X. Available from: doi:10.1109/tmi.2018.2823083

4. YU, Lequan, et al. Deep sinogram completion with image prior for metal artifact reduction in CT images. *IEEE Transactions on Medical Imaging* [online]. 2021, 40(1), 228–238 [viewed 23 May 2023]. ISSN 1558-254X. Available from: doi:10.1109/tmi.2020.3025064

5. WANG, Tao, et al. DAN-Net: Dual-domain adaptive-scaling non-local network for CT metal artifact reduction. *Physics in Medicine & Biology* [online]. 2021, 66(15), 155009 [viewed 23 May 2023]. ISSN 1361-6560. Available from: doi:10.1088/1361-6560/ac1156

6. HAI, Chao, et al. Dual-domain metal trace inpainting network for metal artifact reduction in baggage CT images. *Measurement* [online]. 2023, 207, 112420 [viewed 23 May 2023]. ISSN 0263-2241. Available from: doi:10.1016/j.measurement.2022.112420

7. WANG, Tao, et al. IDOL-Net: an interactive dual-domain parallel network for CT metal artifact reduction. *IEEE Transactions on Radiation and Plasma Medical Sciences* [online]. 2022, 1 [viewed 23 May 2023]. ISSN 2469-7303. Available from: doi:10.1109/trpms.2022.3171440

8. ZHU, Yulin, et al. Sinogram domain metal artifact correction of CT via deep learning. *Computers in Biology and Medicine* [online]. 2023, 155, 106710 [viewed 23 May 2023]. ISSN 0010-4825. Available from: doi:10.1016/j.compbiomed.2023.106710

9. LIAO, Haofu, et al. ADN: artifact disentanglement network for unsupervised metal artifact reduction. *IEEE Transactions on Medical Imaging* [online]. 2020, 39(3), 634–643 [viewed 23 May 2023]. ISSN 1558-254X. Available from: doi:10.1109/tmi.2019.2933425

10. ZHAO, Binyu, et al. Unsupervised reused convolutional network for metal artifact reduction. In: Haiqin YANG, et al., eds. *Neural information processing* [online]. Cham: Springer International Publishing, 2020, pp. 589——596 [viewed 23 May 2023]. ISBN 978-3-030-63820-7. Available from: doi:10.1007/978-3-030-63820-7_67

11. LEE, Junghyun, Jawook GU, and Jong Chul YE. Unsupervised CT Metal Artifact Learning using Attention-guided β-CycleGAN. *IEEE Transactions on Medical Imaging* [online]. 2021, 1 [viewed 23 May 2023]. ISSN 1558-254X. Available from: doi:10.1109/tmi.2021.3101363

12. ZHAO, Binyu, et al. Attention-Embedded decomposed network with unpaired CT images prior for metal artifact reduction. In: *ICASSP 2021 - 2021 IEEE international conference on acoustics, speech and signal processing (ICASSP)* [online]. IEEE, 2021 [viewed 23 May 2023]. Available from: doi:10.1109/icassp39728.2021.9413578

13. LYU, Yuanyuan, et al. U-DuDoNet: unpaired dual-domain network for CT metal artifact reduction. In: Marleen DE BRUIJNE, et al., eds. *Medical image computing and computer assisted intervention -- MICCAI 2021* [online]. Cham: Springer International Publishing, 2021, pp. 296—

—306 [viewed 23 May 2023]. ISBN 978-3-030-87231-1. Available from: doi:10.1007/978-3-030-87231-1_29

14. ZHU, Manman, et al. Physics-informed sinorgram completion for metal artifact reduction in CT imaging. *Physics in Medicine & Biology* [online]. 2023 [viewed 23 May 2023]. ISSN 1361-6560. Available from: doi:10.1088/1361-6560/acbddf

15. ULYANOV, Dmitry, Andrea VEDALDI, and Victor LEMPITSKY. Deep image prior. *International Journal of Computer Vision* [online]. 2020, 128(7), 1867–1888 [viewed 23 May 2023]. ISSN 1573-1405. Available from: doi:10.1007/s11263-020-01303-4

16. MATAEV, Gary, Michael ELAD, and Peyman MILANFAR. DeepRED: deep image prior powered by {RED}. *CoRR* [online]. 2019, abs/1903.10176 [viewed 23 May 2023]. Available from: http://arxiv.org/abs/1903.10176

17. DABOV, Kostadin, et al. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing* [online]. 2007, 16(8), 2080–2095 [viewed 23 May 2023]. ISSN 1941-0042. Available from: doi:10.1109/tip.2007.901238

18. BUADES, A., B. COLL, and J.-M. MOREL. A non-local algorithm for image denoising. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)* [online]. 2005, pp. 60–65 [viewed 23 May 2023]. Available from: doi:10.1109/CVPR.2005.38

19. HECKEL, Reinhard, and Paul HAND. Deep decoder: concise image representations from untrained non-convolutionalnetworks. *CoRR* [online]. 2018, abs/1810.03982 [viewed 23 May 2023]. Available from: http://arxiv.org/abs/1810.03982

20. LEHTINEN, Jaakko, et al. Noise2Noise: learning image restoration without clean data. *CoRR* [online]. 2018, abs/1803.04189 [viewed 23 May 2023]. Available from: http://arxiv.org/abs/1803.04189

21. KRULL, Alexander, Tim{-}Oliver BUCHHOLZ, and Florian JUG. Noise2Void - learning denoising from single noisy images. *CoRR* [online]. 2018, abs/1811.10980 [viewed 23 May 2023]. Available from: http://arxiv.org/abs/1811.10980

22. BATSON, Joshua, and Loic ROYER. Noise2Self: blind denoising by self-supervision. *CoRR* [online]. 2019, abs/1901.11365 [viewed 23 May 2023]. Available from: http://arxiv.org/abs/1901.11365

23. HORE, Alain, and Djemel ZIOU. Image quality metrics: PSNR vs. SSIM. In: *2010 20th international conference on pattern recognition (ICPR)* [online]. IEEE, 2010 [viewed 23 May 2023]. ISBN 9781424475421. Available from: doi:10.1109/icpr.2010.579

24. RONNEBERGER, Olaf, Philipp FISCHER, and Thomas BROX. U-Net: convolutional networks for biomedical image segmentation. *CoRR* [online]. 2015, abs/1505.04597 [viewed 23 May 2023]. Available from: http://arxiv.org/abs/1505.04597

25. BEATTY, J. A. The radon transform and the mathematics of medical imaging. 2012.

26. IAEA. *Diagnostic radiology physics: a handbook for teachers and students*. International Atomic Energy Agency, 2014. ISBN 9789201310101.

27. BOLLIGER, Stephan A., et al. Is differentiation of frequently encountered foreign bodies in corpses possible by Hounsfield density measurement? *Journal of Forensic Sciences* [online]. 2009, 54(5), 1119–1122 [viewed 24 May 2023]. ISSN 1556-4029. Available from: doi:10.1111/j.1556-4029.2009.01100.x