

Iterative computed tomography reconstruction using deep learning

Alexandru Moraru

a.moraru@student.utwente.nl

Faculty of Electrical Engineering Mathematics and Computer Science

December 13, 2020

Abstract—In computed tomography it is important not only to obtain images of good quality but also to minimize the radiation dose given to the patient. Research efforts are dedicated to increase the quality of the reconstructed images and minimize the radiation exposure. This work addresses the problem of reducing the dose by using deep learning to correct the update term of the simultaneous iterative reconstruction technique (SIRT). The aim is to improve the output of the backprojection operator which does not rely on any prior knowledge about the object and distributes all the rays back into the volume under reconstruction uniformly. We propose a deep learning solution to correct the update term of the SIRT algorithm after the backprojection operator has been applied with the purpose to increase the image quality. We evaluate the quality of the images obtained with the proposed method using similarity measures between the low dose reconstructions obtained with the proposed method and the high dose reconstructions taken as ground truth. We also investigate whether the iterative scheme converges faster with the proposed modification. We obtained a structural similarity index (SSIM) of 0.725, a peak signal-to-noise ratio (PSNR) of 29.42 dB and a mean absolute error (MAE) of 92.69 HU which indicates that our method performs better than the classical SIRT algorithm. We also demonstrated that the proposed iterative scheme has the side benefit that it converges faster, achieving with three iterations the similarity that is obtained with the classical scheme with 115 iterations.

I. INTRODUCTION

Computed tomography (CT) is an imaging technique that aims at reconstructing patient's interior volume from the measurement of the X-ray beams that passed through the volume and were attenuated. This imaging modality was used clinically for the first time in the Atkinson Morley Hospital in London in 1972 [1], [2]. Since then, the number of CT scans performed every year has constantly grown and the CT has become an indispensable imaging tool for diagnosis. Nowadays, over 80 million CT scans are performed every year only in the United States [3]. Due to the increasing number of patients that undergo CT scans and the fact that exposure to radiation increases the risk of induced cancer [4], current research efforts are dedicated to improve the image quality while minimizing the exposure to radiation. It is always possible to improve the quality of the CT images by increasing the radiation dose, but this comes at the price of exposing the patients to a higher risk for their health. Thus, the research nowadays is focused on how to lower the radiation dose without affecting the image quality [5].

For many years the reconstruction method of choice was

filtered backprojection (FBP) [6]. This was preferred because it is computationally cheap and the quality of the images obtained with these algorithms was good enough to be used clinically [6]. However, in the case of some patients, especially the young ones which are more prone to be affected by ionizing radiation, the CT scans were being acquired with a lower X-ray intensity to avoid exposing them to a high radiation dose. This determined a reduction of the image quality. Once the computational power permitted, iterative reconstruction algorithms were adopted which made it possible to obtain images of the same quality as those obtained with FBP, but with up to 76% lower radiation dose [3]. In recent years, methods that combine iterative schemes with learned models started to be investigated. The reason why these methods are interesting is that the reconstruction task is partly performed by the classical steps of the scheme which greatly reduces the amount of work to be done by the neural networks and thus makes it easier to train them [5]. On the other hand, there is also the advantage that if the prediction of the neural networks at a certain iteration is wrong, the classical part of the algorithm will correct the wrong prediction during the next iterations.

In this work we address the problem of reducing the radiation dose and propose a modification of the simultaneous iterative reconstruction technique (SIRT) by using an independent deep neural network at each iteration to improve the quality of the low dose images. The neural network is designed to improve the output of the backprojection operator which has the drawback that does not rely on any prior knowledge about the volume and distributes all the rays back into it uniformly. We will use low dose projections as input in the reconstruction scheme and take images reconstructed from high dose projections as ground truth for training the neural networks. The quality of the reconstructed low dose images is improved by making them more similar to the reference high dose images. We perform optimization of the solution on a validation set investigating different loss functions and the use of transfer learning among iterations to possibly improve the model convergence and the performance. We also use the validation set to compare a number of network architectures in order to identify the one that gives the best performance indicators. Finally, we will evaluate the performance of the optimized solution on the test set and we will analyze whether the iterative algorithm converges faster

with the proposed correction.

II. RELATED WORK

Over the last few years there has been a growing interest in the application of deep learning to solve a large number of technical problems and in computed tomography there has been a number of proposed deep learning solutions which address the problem of dose reduction.

One possible way to employ deep learning in computed tomography is to use neural networks to postprocess the reconstructed image in order to improve the quality of the image. Kida et al. [7] proposed a deep convolutional neural network to improve the quality of the cone beam computed tomography (CBCT) images which are used to reassess the position of the tumor before the treatment is delivered. They used planning CT images (which are used for diagnostic and have a higher quality) as targets for the neural networks. Jiang et al. [8] proposed a symmetric residual convolution neural network (SRCNN) architecture to enhance the sharpness of edges and anatomical details for images reconstructed from under-sampled projections. The targets used for training the deep learning model were the fully sampled CBCT reconstructions. By using fewer projections, the X-ray radiation exposure of the patient is reduced. Maspero et al. [9] proposed a cycle-GAN architecture which generates synthetic-CT images from CBCT images reducing the CBCT artefacts and increasing similarity to CT. They proved the efficiency of the method by showing that it results in an accurate dose calculation for radiotherapy treatment.

Another possible way in which deep learning can be used in computed tomography is by implementing the convolution of a classical reconstruction algorithm as a layer in a neural network and learning the filter samples [10].

Last but not least, deep learning solutions were also proposed to modify iterative reconstruction schemes. Aggarwal et al. [11] proposed a convolutional neural network (CNN) to be used in the regularization term of the conjugate gradient optimization scheme. Although they applied the proposed method to MRI, the same idea is applicable to CT. Dilz et al. [12] focus on low dose CBCT application and propose an algorithm named learned SIRT. They rework the derivation of the SIRT formula so as to obtain a regularizer term which will be predicted by a neural network and demonstrate that their algorithm scales to clinically relevant problems.

III. COMPUTED TOMOGRAPHY BASICS

The problem addressed in X-ray Computed Tomography is to reconstruct a discrete volume (also called image) given the attenuation values of the X-ray radiation for a sufficient number of paths that pass through the volume. Each voxel (volume pixel) in the image is described by its X-ray attenuation coefficient. We use the terms *projection bin* to refer to the attenuation on a single path, *projection* to refer to all attenuation values measured for a single angular position and *set of projections* to refer to all attenuation values measured for all angular positions, respectively.

Figure 1 illustrated a simplified example of a CT scanner. Each projection bin represents the sum of the voxel's attenuation coefficients over the corresponding X-ray path through the volume (the total attenuation of the path is the product of the attenuation values of the path's voxels, but applying the logarithm transforms the product into a sum [13]).

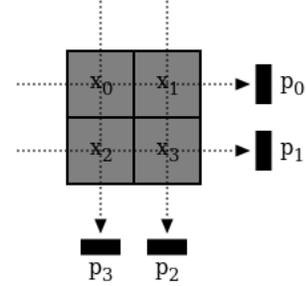


Fig. 1: The principle of the CT scanner: each projection bin represents the sum of the attenuation values of the voxels that radiation passes through.

From Figure 1 we obtain the following relations between the voxels and the projection bins:

$$\begin{aligned} x_0 + x_1 &= p_0 \\ x_2 + x_3 &= p_1 \\ x_1 + x_3 &= p_2 \\ x_0 + x_2 &= p_3 \end{aligned} \quad (1)$$

We can rewrite the system (1) in much more compact form using the system matrix A :

$$Ax = p \quad (2)$$

where

$$A = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix} \quad (3)$$

$$x = [x_0 \quad x_1 \quad x_2 \quad x_3]^T \quad (4)$$

and

$$p = [p_0 \quad p_1 \quad p_2 \quad p_3]^T \quad (5)$$

A can be thought as a *forward projection operator* which is applied to the image x to obtain the set of projections p and thus models how the radiation is attenuated in the volume x . In real CT scans, the number of voxels in an image and the number of projection bins differ and thus A is a non-square matrix of size (m, n) , where m is the number of projection bins and n is the number of voxels.

In order to be able to do reconstruction, we also need the *backprojection operator*. This operator takes the energy from each projection bin and distributes it equally to all the pixels encountered on the path (see Figure 2). From a mathematical point of view, this operator is the transpose of the forward projection operator A .

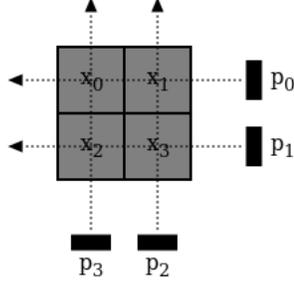


Fig. 2: The backprojection operator takes all the values from the projection space and redistributes them back on the path of the X-rays.

A. Classical SIRT scheme

The solution to the equation 2 can be solved exactly only under idealized conditions. However, in computed tomography we are dealing with real data which is affected by measurement noise. Also, in the real world CT scanners the number of voxels differs from the number of projection bins which makes it impossible to invert the matrix A . We therefore consider the least squares minimum norm solution to the equation 2 which can be found by minimizing the objective function

$$\chi^2 = \|Ax - p\|_2^2 \quad (6)$$

This objective function can be minimized by starting with an initial image $x^{(0)}$ having all the voxels set to 0 and applying the gradient update:

$$x^{(k+1)} = x^{(k)} - \alpha \nabla \|Ax - p\|_2^2 \quad (7)$$

where the gradient is:

$$\nabla \|Ax - p\|_2^2 = A^T (Ax^{(k)} - p) \quad (8)$$

Combining equations 7 and 8 we get the update formula of the SIRT algorithm:

$$x^{(k+1)} = x^{(k)} + \alpha A^T (p - Ax^{(k)}) \quad (9)$$

Although one can work with this SIRT variant, it is impractical due to the fact that the step size α has to be chosen manually. We will instead proceed as in [14] and use their variant of SIRT:

$$x^{(k+1)} = x^{(k)} + CA^T R (p - Ax^{(k)}) \quad (10)$$

$$C_{jj} = 1 / \sum_j a_{ij} \quad R_{ii} = 1 / \sum_i a_{ij}$$

where C and R are diagonal matrices and a_{ij} are the individual components of A . The matrices C and R are weighting matrices that contain the sum of the columns and

rows of the projection matrix, respectively and are necessary to guarantee the convergence of the scheme [14].

IV. MATERIALS AND METHODS

A. Proposed method

The drawback of the SIRT scheme is that the backprojection operator that is used in the update formula distributes the energy from each projection bin uniformly into the volume under reconstruction on the path of the corresponding ray without relying on any prior knowledge about the object. In Figure 3 we have shown an example in which the energy was misallocated during backprojection. Our goal is to train a neural network to provide this missing knowledge about where the energy should be allocated. This will be done by predicting for each voxel a value between 0 and 1 representing how much of the energy allocated by the backprojection operator should actually be allocated there and then using it to correct the allocation. In this way, we try to prevent the backprojection operator from allocating more energy that it should actually do.

In the following, we denote with \otimes the elementwise multiplication of two tensors. Also, we denote by $r^{(k)}$ the update term of the SIRT scheme:

$$r^{(k)} = CA^T R (p - Ax^{(k)}) \quad (11)$$

With this new notation, equation 10 becomes:

$$x^{(k+1)} = x^{(k)} + r^{(k)} \quad (12)$$

In order to prevent the SIRT scheme from allocating energy to the wrong voxels, we propose to introduce a multiplication tensor $\lambda^{(k)}$ in the SIRT formula so as to correct the update term $r^{(k)}$:

$$x^{(k+1)} = x^{(k)} + \lambda^{(k)} \otimes r^{(k)} \quad (13)$$

The tensor $\lambda^{(k)}$ is predicted by a neural network based on the image at the current iteration $x^{(k)}$ and the update term $r^{(k)}$:

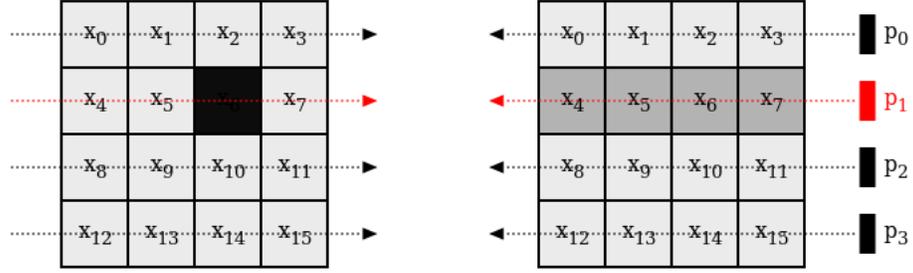
$$\lambda^{(k)} = \text{dnn}(x^{(k)}, r^{(k)}) \quad (14)$$

We make an assumption that in order not to allocate more energy than the classical SIRT scheme would allocate, we need that $0 \leq \lambda_{ij}^{(k)} \leq 1$. Note that if all the components of the tensor predicted by the neural network are 1, then the scheme is identical to the classical SIRT:

$$x^{(k+1)} = x^{(k)} + 1 \otimes r^{(k)} = x^{(k)} + r^{(k)} \quad (15)$$

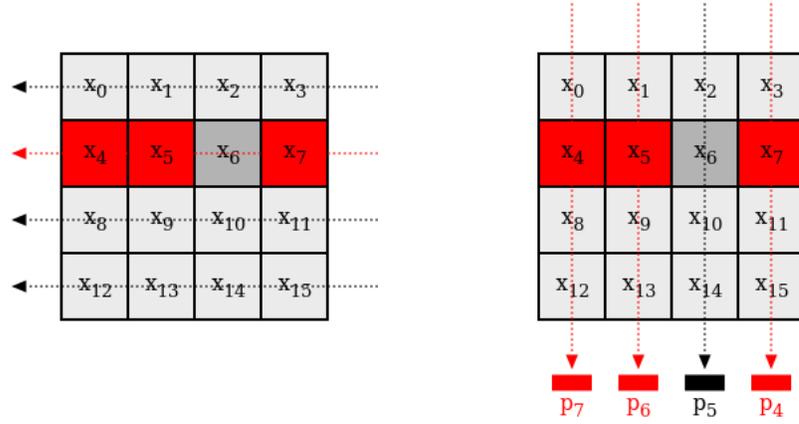
On the contrary, if the components of the predicted tensor are all 0, then the image is not varied. We hypothesise that this will prevent the image degradation due to misallocations.

In Figure 4 we illustrate how λ_{gt} is calculated. In Figure 4 (a) we have the image that is taken as ground truth. In Figure 4 (b) we can see the first update that the classical SIRT



(a) Image with a high attenuation voxel

(b) Redistributing back the energy



(c) Misallocated energy

(d) Other projection bins are affected

Fig. 3: The backprojection operator redistributes the energy back uniformly to all voxels. Suppose we have an image which has a single voxel (x_6) which has a high attenuation value and which results in a high value measured by the detector bin p_1 (see Figure 3 (a)). When backprojection is applied, the measured energy will be redistributed also to the voxels x_4 , x_5 and x_7 (see Figure 3 (b) and Figure 3 (c)). This will create, in turn, a further mismatch at the projection bins p_4 , p_6 and p_7 .

scheme would make to the initial image. Because the initial image is 0, the first update $r^{(0)}$ coincides with the image at iteration 1, $x^{(1)}$. For training, the target (ground truth) tensor $\lambda_{gt}^{(k)}$ is taken to be the one that brings the image $x^{(k)}$ as close as possible to the full dose image x_{fd} . To obtain it, we first calculate the update that would bring $x^{(k)}$ to x_{fd} :

$$r_{gt}^{(k)} = x_{fd} - x^{(k)} \quad (16)$$

We then calculate the $\lambda_{gt}^{(k)}$ tensor as:

$$\lambda_{gt}^{(k)} = \frac{r_{gt}^{(k)}}{r^{(k)}} \quad (17)$$

Finally, we clip $\lambda_{gt}^{(k)}$ values between 0 and 1 and train a neural network to predict the $\lambda^{(k)}$ tensor. In Figure 4 (c) we can see the resulted multiplication tensor $\lambda_{gt}^{(0)}$ and in Figure 4 (d) we can see the update term after it has been corrected. Comparing Figures 4 (b) and (d), we can see how the image $x^{(1)}$ would be improved if the prediction of $\lambda^{(0)}$ was perfect.

In equation 17 we performed a division operation. It might be that some of the components of the classical update term $r^{(k)}$ are close to zero which could cause numerical overflows. In order to avoid that, we check all the values of $r^{(k)}$ before performing the division and we force the pixels of $\lambda_{gt}^{(k)}$ to zero where the corresponding absolute values of the pixels

of $r^{(k)}$ are smaller than 10^{-15} .

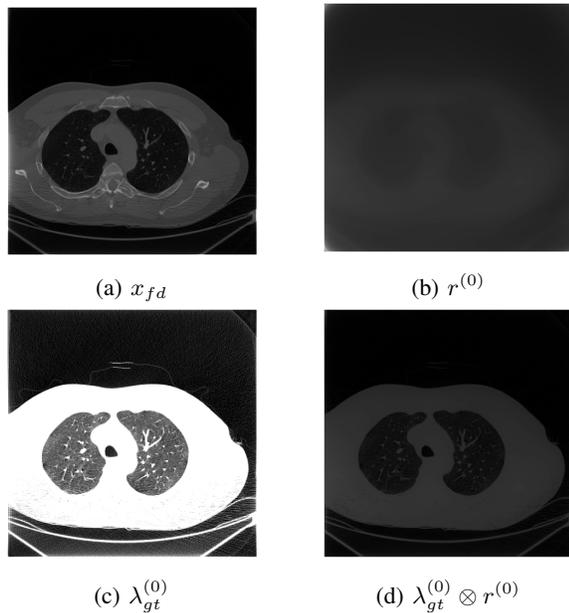


Fig. 4: How the first step of the SIRT scheme is modified: based on the ground truth (a) and the classical update term (b) we calculate the target tensor (c) that the neural network should predict. Applying the proposed multiplication, we obtain the corrected update in (d).

B. The dataset and preprocessing

In this work we used data from the Low Dose CT Grand Challenge which had the purpose to quantitatively assess the diagnostic performance of denoising and iterative reconstruction techniques on common low-dose patient CT datasets. The data contains full dose projections (real sensor data) and simulated low dose projections (obtained by adding noise to the full dose projections). From the available data we selected only the chest scans acquired with the GE LightSpeed VCT scanner model which corresponded to scans of 45 patients. Patients underwent a helical acquisition with a peak voltage=100 kV, tube current=300 mA, exposure time=500 ms and a collection field of view=500 mm. The patients were randomly divided in three sets: training (n=27), validation (n=9) and test (n=9).

Preprocessing of the Low Dose CT Grand Challenge data was performed to convert to circular trajectory, to be able to reconstruct the sinograms with available frameworks. In particular, the preprocessing consisted in dividing the volume to be reconstructed in slices perpendicular to the axial direction by:

- Creating a sinogram containing the projections from a complete rotation for each slice by interpolating at each angle step the two closest projections to the slice position; since the CT device uses a helical trajectory, the positions at which projections are made do not, in general, coincide with the slice position and thus interpolation is necessary to obtain the attenuation values at

that position. Interpolation was performed as explained in [6]. The slice thickness was set to 1.25 mm and the reconstruction space chosen covers an area of 377 mm by 377 mm divided in an array of 512 by 512 voxels.

- Rebinning the data from a curved detector array to a flat detector array; this step was necessary to make it possible to work with ASTRA Toolbox [15], [16].
- Reconstructing the full dose images from the corresponding sinograms using alternative direction of multipliers algorithm [17] implemented in the ODL framework [18] with ASTRA Toolbox backend; these images are necessary to provide the ground truth for the learning algorithm.

C. Architecture optimization and training

We implement the proposed method using an independent neural network at each SIRT iteration. We fixed the number of SIRT iterations to ten investigating whether ten iterations are sufficient to obtain a good performance. The input to the network has two channels: the image at the current iteration, $x^{(k)}$ and the classical update term $r^{(k)}$ calculated with equation 11 which were normalized so as to have values between 0 and 1 before applying the forward pass. The chosen architecture for the network is U-Net [19] because it proved to give very good results in situations when the output has the same dimensions as the input. The U-Net architecture was adopted as provided by the MONAI framework [20]. The number of up and downsampling blocks, number of channels and the use of residual units has been investigated as reported in the experiment section. The standard configuration for the experiments consists of kernel size=5 (both for up and downsampling), stride=2. The activation function used throughout the network was PReLU.

Training was performed with ADAM optimizer for 100 epochs with a batch size of 32. Early stopping was applied with a patience of 7 epochs, mean squared error and binary crossentropy loss functions over the whole image were considered during training, performing experiments to decide which one gives the best results.

D. Evaluation metrics

In order to assess the performance of the proposed method, we employ three performance indicators: structural similarity index measure (SSIM), mean absolute error (MAE) and peak signal-to-noise ratio (PSNR) with a data range of 4095 HU. Since the corners of the image are mostly dark areas, we calculate these metrics only for the voxels in region of interest illustrated in Figure 5. We will not explain these metrics here, but the interested reader can consult references [21] and [22] for a complete explanation on how they are computed.

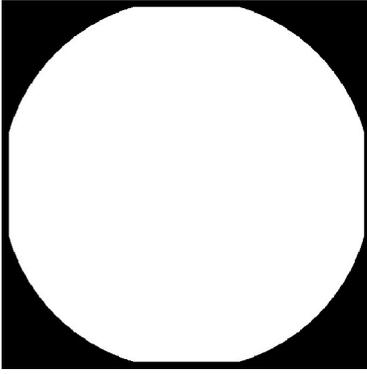


Fig. 5: The backprojection operator takes all the values from the projection space and redistributes them back on the path of the X-rays.

V. EXPERIMENTS AND RESULTS

In this section we present the results obtained with the proposed method. Since we are trying to improve an existing iterative algorithm, we will first obtain a baseline with the classical SIRT algorithm. We then perform three experiments to optimize the proposed solution and we use the validation set to take design solutions. Finally, we evaluate the optimized solution on the test set.

A. Baseline

In order to obtain a baseline, we ran the classical SIRT algorithm (initialized with filtered backprojection) for 400 iterations on the test set. In Figure 6 it is depicted the variation of the PSNR as the iteration number increases. We can observe that after a certain iteration (115th) the quality of the reconstructions starts to decrease. In TABLE I we compare the metric values after 115 iterations, when the maximum similarity is achieved, and after 400 iterations. These values show that after the 115th iteration the images better resemble the ground truth than after 400 iterations.

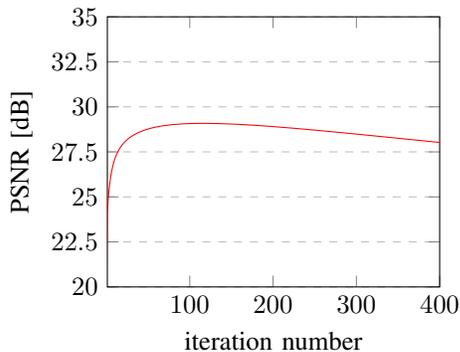


Fig. 6: Variation of PSNR over iteration number in the case of classical SIRT scheme with filtered backprojection initialization. It can be seen that after the 115th iteration the similarity to the ground truth starts to decrease.

	after 400 iterations	best value
SSIM [-]	0.634	0.669
PSNR [dB]	28.02	29.08
MAE [HU]	123.50	115.42

TABLE I: Performance indicators for the classical scheme. The highest similarity (the best metric values) is achieved with 115 iterations. After 115 iterations the similarity decreases.

B. Model optimization

In order to optimize the model, we propose three experiments. We decide which experiments are successful based on the performance indicators calculated for the validation set images after the tenth SIRT iteration has been performed. The experiments we propose are:

1) *Compare loss functions:* We compare the metrics when using the mean squared error (MSE) loss or when we use the binary crossentropy (BCE) loss. In Figure 7 we can see the evolution of the PSNR over the iteration number for the two loss functions used from which it can be seen that the mean squared error consistently performed better throughout all the iterations. In TABLE II we displayed the maximum values of the evaluation metrics on the validation set achieved during the iterative scheme. It can be observed that the mean squared error loss gave significantly better results than the binary crossentropy. In Figure 8 we can see how the correction looks for the case of the two losses investigated. We can observe that the proposed scheme is capable to identify the regions where the energy should be backprojected. It can also be seen that the correction in the case of the mean squared error loss is better at preserving the details. We therefore decide to use the mean squared error during training because it gives a better performance than the binary crossentropy loss.

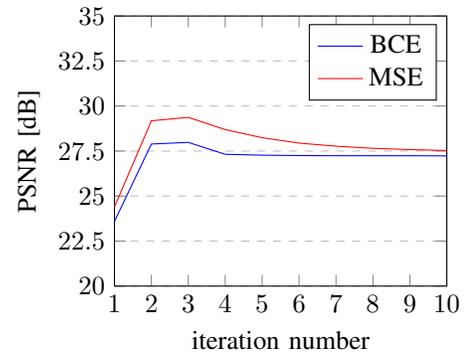


Fig. 7: PSNR for the BCE and MSE losses on the validation set.

Loss	SSIM [-]	PSNR [dB]	MAE [HU]
BCE	0.641	27.98	113.37
MSE	0.7187	29.37	87.79

TABLE II: The best values of the similarity metrics achieved during the iterative scheme on the validation set for the two losses used show that the MSE loss outperforms the BCE loss.

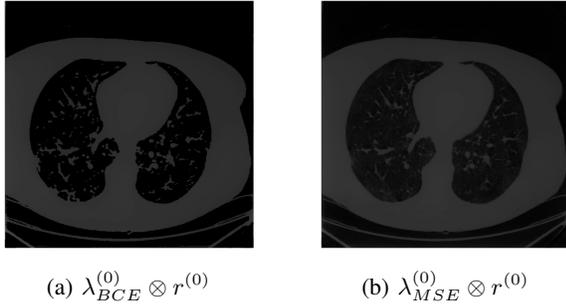


Fig. 8: Comparison between the update terms in the case of the two losses: (a) the classical update, (b) ideal update, (c) update in the case of BCE loss, (d) update in the case of MSE loss.

2) *Initialization with previous weights*: As we use a different network at each iteration, it is natural to ask whether initializing the weights of the network to be trained with the values of the weights of the network trained at the previous iteration improves the performance. In Figure 9 we can see the variation of the PSNR with the iteration number for the case when we loaded the weights from the previous network and for the case when we initialized the weights at random from which we can see that the difference in performance is not significant. In TABLE III we can see the values of the SSIM, PSNR and MAE evaluation metrics for the same experiment which also indicate that the performance does not differ significantly. As it turned out that there were minor improvements if we load the previous weights, we decide to continue to do so for the rest of the experiments.

Using previous weights	SSIM [-]	PSNR [dB]	MAE [HU]
no	0.7187	29.37	87.79
yes	0.723	29.52	87.24

TABLE III: Comparison of the evaluation metrics for case when we initialize with the weights previously learned and for the case when we initialize the weights at random.

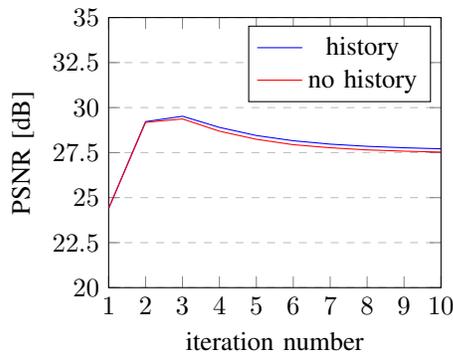


Fig. 9: PSNR for the case when we initialize with the weights previously learned and for the case when we initialize the weights at random.

3) *Tuning the hyperparameters of the U-Net*: We consider five of variations of the U-Net architecture. Since running

an experiment for ten SIRT iterations is computationally expensive, we will restrict ourselves to a suboptimal search and optimize only for the network used at the third SIRT iteration, which is the iteration for which the best similarity metrics were obtained during the previous experiments. We will then use the same network architecture at all iterations. In TABLE IV we enumerated the architectures of the five networks investigated and in TABLE V we have the evaluation of each of these networks from which we can see that there is no significant difference in performance between the five architectures investigated. As in the case of the previous experiment, we decide to adopt the architecture that gave the best performance even though the difference is not significant.

	# blocks	# channels	# residual units
Arch 1	5	8, 16, 32, 64, 128	0
Arch 2	5	4, 8, 16, 32, 64	0
Arch 3	6	2, 4, 8, 16, 32, 64	0
Arch 4	5	8, 16, 32, 64, 128	1
Arch 5	5	4, 8, 16, 32, 64	1

TABLE IV: Comparison of the evaluation metrics for five architectures considered.

	SSIM [-]	PSNR [dB]	MAE [HU]
Arch 1	0.7041	29.37	94.95
Arch 2	0.7007	29.33	95.54
Arch 3	0.7005	29.30	96.12
Arch 4	0.7045	29.39	94.75
Arch 5	0.7012	29.31	95.95

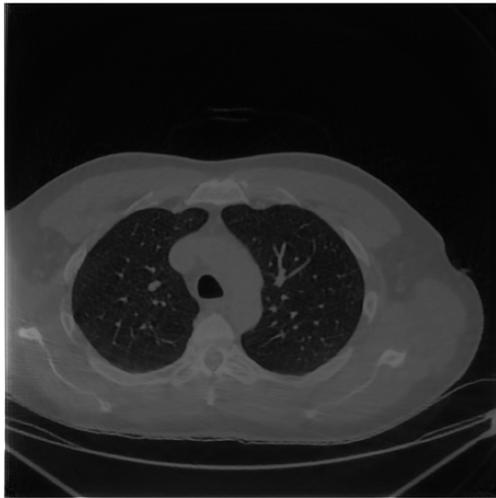
TABLE V: Result of the evaluation metrics for the five architectures considered.

C. Results on the test set

We use the test set to assess the performance of the solution found during the optimization stage. The resulted values of the evaluation metrics for the proposed scheme after ten iterations evaluated on the test set are:

- SSIM: 0.7254
- PSNR: 29.42
- MAE: 92.69

In Figures 11 - 13 we show the variation of these metrics over the iteration number for the proposed method and for the classical SIRT algorithm. We can observe that the proposed method outperforms the classical scheme throughout all the ten iterations. We can also observe that after the third iteration the similarity starts to decrease in the case of the proposed scheme. This was also observed in the case of the variation of the PSNR of the classical scheme in Figure 6, but there it was necessary to perform 115 iterations to achieve the maximum value whereas the proposed algorithm achieved the maximum similarity with only three iterations. In Figure 14 we depicted the variation of the objective function over the iteration number and we observe that it indeed decreases for all iterations.



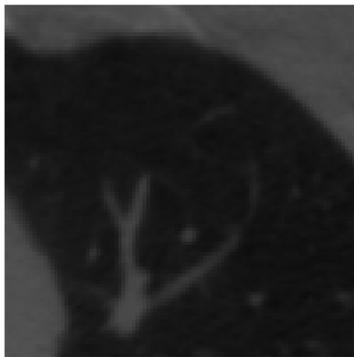
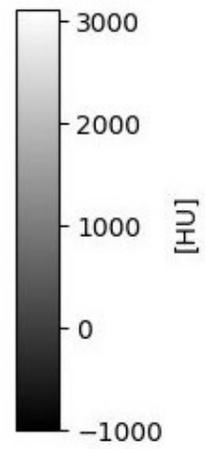
(a) proposed



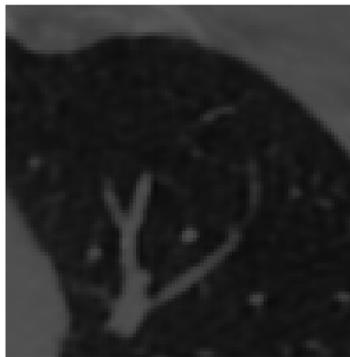
(b) classical



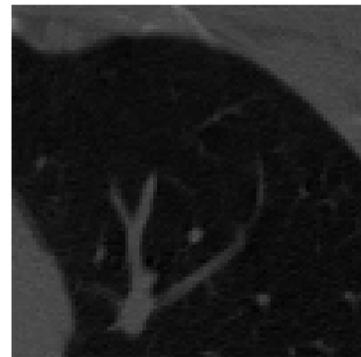
(c) ground truth



(d) classical



(e) proposed



(f) ground truth

Fig. 10: Comparison between the images obtained with the proposed scheme (a and e), with the classical SIRT scheme (b and d) and the ground truth (c and f).

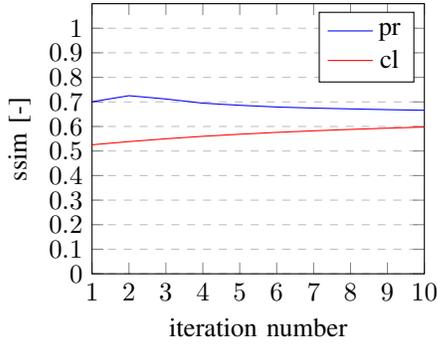


Fig. 11: SSIM for proposed solution (pr) and for the classical SIRT (cl) evaluated on the test set.

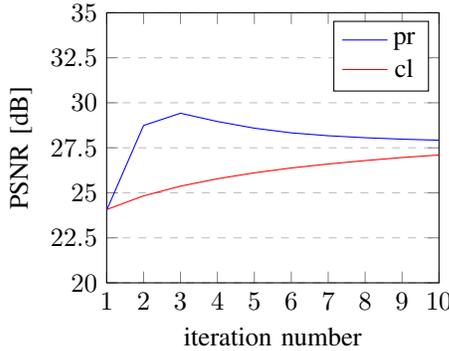


Fig. 12: PSNR for proposed solution (pr) and for the classical SIRT (cl) evaluated on the test set.

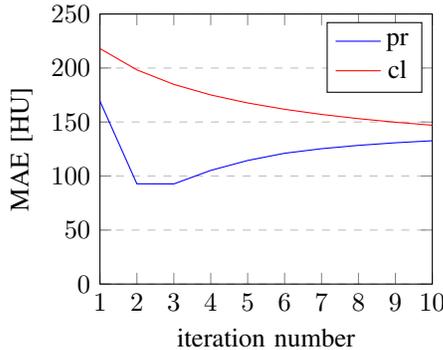


Fig. 13: MAE for proposed solution (pr) and for the classical SIRT (cl) evaluated on the test set.

In Figure 10 we can see a comparison between a reconstruction obtained with the proposed method (a and e), the same reconstruction obtained with the classical SIRT algorithm (b and d) and the ground truth used for evaluation (c and f). It can be observed that the proposed method succeeded to improve the quality of the image at the tenth iteration and that it reduced considerably the streaking artifacts. We can also observe that the details in the lung region were accurately reconstructed at the cost of having a poorer reconstruction in the bone region.

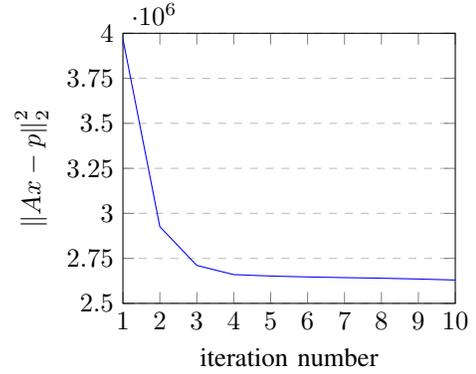


Fig. 14: Variation of the value of the objective function to be minimized by the SIRT algorithm over iteration number.

VI. DISCUSSION

The focus of this work is to explore whether neural networks can learn to correct the output of the backprojection operator which has drawback that it projects all the rays uniformly into the volume without using any prior knowledge and in this way to improve the SIRT scheme. Comparing the performance metric values obtained for the proposed method in section V-C with the ones obtained for the baseline in TABLE I we can conclude that the proposed method succeeded to improve the SIRT scheme. In addition, it turned out that the proposed iterative scheme has the side benefit that it converges very fast, achieving higher similarity to the ground truth than the classical SIRT with only three iterations. However, if we compare our method with other works found in literature, we will see that there is a considerable difference between results of the existing state of the art methods and the results we obtained. For example, Dilz et al. [12], who also proposed a modification to the SIRT algorithm, obtained a PSNR of 31.7 dB and an SSIM of 0.922 for a model trained on chest anatomy. However, they used a different dataset and possibly a different region of interest for metric evaluation.

We also analyzed how the values of the similarity measures vary during the iterative algorithm and we observed that the similarity decreases after the third iteration. We observed that this behaviour is present also in the classical SIRT. The possible reason for this is that the SIRT scheme is designed to minimize objective function 6 which is different from the metrics we use to evaluate the performance of the proposed method. In order to verify this, we analyzed the variation of the objective function over the ten iterations (see Figure 14) and observed that it decreases throughout the execution of the iterative algorithm. However, we can observe that it quickly reaches a plateau. This indicates that the objective function 6 can hardly be further minimized after the fourth iteration. Future work may investigate adding a regularization term to the objective function that contains the prior knowledge about the image features and constrains the iterative algorithm to converge to a solution that contains these features. The regularization can be provided by a deep

neural network which is very suitable for the task of learning features. This approach was proposed by Aggarwal et al. [11] who demonstrated the efficiency of the method obtaining a PSNR of 39.24 dB on a brain MRI dataset. Another limitation of the proposed method is that the neural network is used only after the backprojector has been applied to correct its output. Using an intelligent backprojector, one which is capable to learn how to allocate the energy, may be a better approach for a future work.

VII. CONCLUSION

In this work we addressed the problem of reducing the dose given to the patient during the CT scanning and proposed a modification to the simultaneous iterative reconstruction technique (SIRT) used in X-ray computed tomography. Our goal was to correct the output of the backprojector which redistributes back the rays from the detectors uniformly using no prior knowledge about the volume under reconstruction and in this way to improve the quality of the images reconstructed with the SIRT algorithm. We proposed experiments to identify which loss function is indicated to be used for training in order to obtain the best performance, to assess whether transferring the network weights from each iteration to the next one improves the performance. We tested a number of network architectures in order to identify the one that gives the best performance. We also compared the speed of convergence of the proposed scheme with that of the classical SIRT. We obtained an SSIM of 0.725, a PSNR of 29.42 dB and an MAE of 92.69 HU and we showed that this is an improvement over the classical SIRT algorithm. We also proved that the proposed method has the side benefit that it can greatly speed up the convergence of the SIRT algorithm achieving with three iterations the same quality that is obtained with 115 classical iterations.

REFERENCES

- [1] J. Ambrose and G. Hounsfield, "Computerized transverse axial tomography.," *The British journal of radiology*, vol. 46, no. 542, p. 148, 1973.
- [2] G. N. Hounsfield, "Computerized transverse axial scanning (tomography): Part 1. description of system," *The British journal of radiology*, vol. 46, no. 552, pp. 1016–1022, 1973.
- [3] M. J. Willemink and P. B. Noël, "The evolution of image reconstruction for ct—from filtered back projection to artificial intelligence," *European radiology*, vol. 29, no. 5, pp. 2185–2195, 2019.
- [4] D. Shah, R. Sachs, and D. Wilson, "Radiation-induced cancer: a modern view," *The British journal of radiology*, vol. 85, no. 1020, pp. e1166–e1173, 2012.
- [5] G. Wang, Y. Zhang, X. Ye, and X. Mou, *Machine learning for tomographic imaging*. IOP Publishing, 2019.
- [6] T. M. Buzug, "Computed tomography," in *Springer Handbook of Medical Technology*, pp. 311–342, Springer, 2011.
- [7] S. Kida, T. Nakamoto, M. Nakano, K. Nawa, A. Haga, J. Kotoku, H. Yamashita, and K. Nakagawa, "Cone beam computed tomography image quality improvement using a deep convolutional neural network," *Cureus*, vol. 10, no. 4, 2018.
- [8] Z. Jiang, Y. Chen, Y. Zhang, Y. Ge, F.-F. Yin, and L. Ren, "Augmentation of cbct reconstructed from under-sampled projections using deep learning," *IEEE transactions on medical imaging*, vol. 38, no. 11, pp. 2705–2715, 2019.

- [9] M. Maspero, A. C. Houweling, M. H. Savenije, T. C. van Heijst, J. J. Verhoeff, A. N. Kotte, and C. A. van den Berg, "A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer," *Physics and Imaging in Radiation Oncology*, vol. 14, pp. 24–31, 2020.
- [10] T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier, "Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1454–1463, 2018.
- [11] H. K. Aggarwal, M. P. Mani, and M. Jacob, "Modl: Model-based deep learning architecture for inverse problems," *IEEE transactions on medical imaging*, vol. 38, no. 2, pp. 394–405, 2018.
- [12] R. J. Dilz, L. Schröder, N. Moriakov, J.-J. Sonke, and J. Teuwen, "Learned sirt for cone beam computed tomography reconstruction," *arXiv preprint arXiv:1908.10715*, 2019.
- [13] W. Demtröder, *Experimentalphysik 3: Atome, Moleküle und Festkörper*. Springer-Verlag, 2010.
- [14] A. Van der Sluis and H. Van der Vorst, "Sirt-and cg-type methods for the iterative solution of sparse linear least-squares problems," *Linear Algebra and its Applications*, vol. 130, pp. 257–303, 1990.
- [15] W. Van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabrovolski, J. De Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the ASTRA toolbox," *Optics express*, vol. 24, no. 22, pp. 25129–25147, 2016.
- [16] W. Van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA toolbox: A platform for advanced algorithm development in electron tomography," *Ultramicroscopy*, vol. 157, pp. 35–47, 2015.
- [17] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 123–231, 2014.
- [18] H. Kohr and J. Adler, "ODL (operator discretization library)," 04 2017.
- [19] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [20] "Project MONAI, <https://docs.monai.io>."
- [21] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [22] A. Horé and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th International Conference on Pattern Recognition*, pp. 2366–2369, 2010.