



# Sparse-View CT Reconstruction Using Wasserstein GANs

Franz Thaler<sup>1</sup>(✉), Kerstin Hammernik<sup>1</sup>, Christian Payer<sup>1</sup>, Martin Urschler<sup>2</sup>,  
and Darko Štern<sup>1,2</sup>

<sup>1</sup> Institute of Computer Graphics and Vision, Graz University of Technology,  
Graz, Austria

`f.thaler@student.tugraz.at`, `stern@icg.tugraz.at`

<sup>2</sup> Ludwig Boltzmann Institute for Clinical Forensic Imaging, Graz, Austria

**Abstract.** We propose a 2D computed tomography (CT) slice image reconstruction method from a limited number of projection images using Wasserstein generative adversarial networks (wGAN). Our wGAN optimizes the 2D CT image reconstruction by utilizing an adversarial loss to improve the perceived image quality as well as an  $L_1$  content loss to enforce structural similarity to the target image. We evaluate our wGANs using different weight factors between the two loss functions and compare to a convolutional neural network (CNN) optimized on  $L_1$  and the Filtered Backprojection (FBP) method. The evaluation shows that the results generated by the machine learning based approaches are substantially better than those from the FBP method. In contrast to the blurrier looking images generated by the CNNs trained on  $L_1$ , the wGANs results appear sharper and seem to contain more structural information. We show that a certain amount of projection data is needed to get a correct representation of the anatomical correspondences.

**Keywords:** Computed tomography · Sparse-view reconstruction  
Convolutional neural networks · Generative adversarial networks  
L1 loss

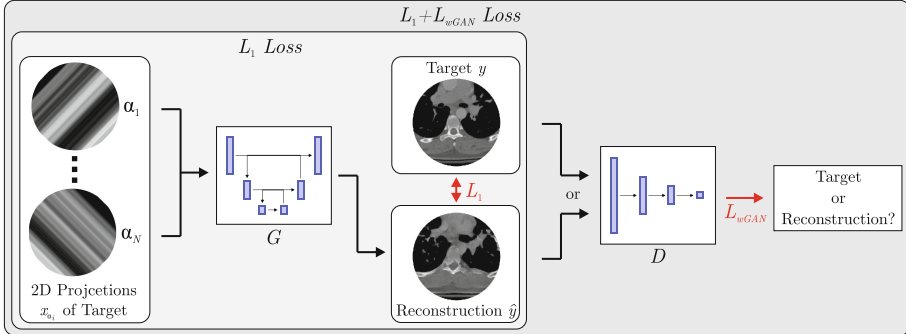
## 1 Introduction

Computed tomography (CT) is a non-invasive image modality to visualize the interior body structure, enabling fast acquisition and high image quality. To generate a three dimensional (3D) CT image, multiple 2D X-ray projection images of the subject are acquired from different angles on the axial plane and used for reconstruction. The Filtered Backprojection (FBP) is a well established method for 3D CT reconstruction. However, the quality of the reconstructed image using FBP heavily depends on the number of projection images, which correlates to the amount of ionizing radiation exposed.

---

This work was supported by the Austrian Science Fund (FWF): P28078-N33 and by the Austrian Science Fund (FWF) under the START project BIVISION, No. Y729.

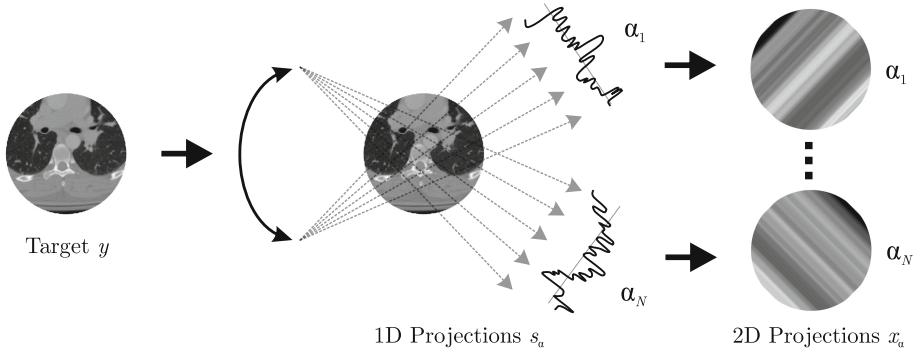
As the risk of cancer is increased by radiation exposure, different approaches exist to decrease the radiation dose. Two popular approaches to decrease radiation dose are tube current reduction, resulting in degraded image quality, and beam blocking, which restricts the amount of X-rays reaching the subject in a physical way, resulting in streaking artifacts. Recent promising results for ionizing dose reduction were achieved by utilizing convolutional neural networks (CNN) [3, 4, 13] and made deep learning also attractive for image reconstruction.



**Fig. 1.** Reconstruction  $\hat{y}$  of a target image  $y$  from a limited number of 2D projection images  $x_{\alpha_i}$  generated from  $y$  with different angles  $\alpha_i$  using a combination of a wGAN loss  $L_{wGAN}$  and an additional content loss  $L_1$ . The generator  $G$  is based on the U-Net [7], the discriminator  $D$  results in a single scalar value.

Reducing the number of X-ray image views acquired and used for CT reconstruction is another approach to decrease the amount of radiation exposed. Sparse-view CT reconstruction becomes important during minimally invasive and image guided surgeries, where multiple X-ray images are acquired repeatedly during intervention to precisely locate the instruments, leading to an exposure to ionizing radiation for both the patient and medical staff. In a recent CNN based approach [10], residual learning is used to extract the artifacts from the FBP image which are then subtracted from the FBP image to obtain the clean reconstruction. In contrast to other CNN based approaches that learn the transformation from a low quality, FBP based reconstructed CT image to a high quality CT image, in our previous work [9], we learned a direct mapping from 3D digitally reconstructed radiographs (DRR) to the full 3D CT reconstruction using a U-Net architecture. However, the downside of this approach is that the reconstructed images look blurry due to the used  $L_1$  loss. This observation suggests to improve on the loss function used for training.

Generative adversarial networks (GANs), which can generate realistically looking images, have a great potential to improve also the reconstruction quality of medical images. A GAN requires two networks to be trained: a generator, which has the goal to create images coming from a target distribution, and a discriminator, which has to distinguish between the generated and the real target



**Fig. 2.** Generation of 1D projections  $s_\alpha$  from a target 2D CT slice image  $y$  for a number of  $N$  fixed angles  $\alpha_i$ . All  $s_\alpha$  are further processed by repeating in the direction of the respective  $\alpha$  yielding the 2D projection images  $x_\alpha$  used as network input.

distribution. However, GANs are inherently hard to train and often suffer from stability issues. Wasserstein GANs (wGANs) [1], which were further improved by utilizing a gradient penalty [2], provide a way to stabilize the training. Combined with a content loss such as  $L_1$ , state-of-the-art results were achieved for super resolution [5] and in medical imaging [6, 8, 11, 12].

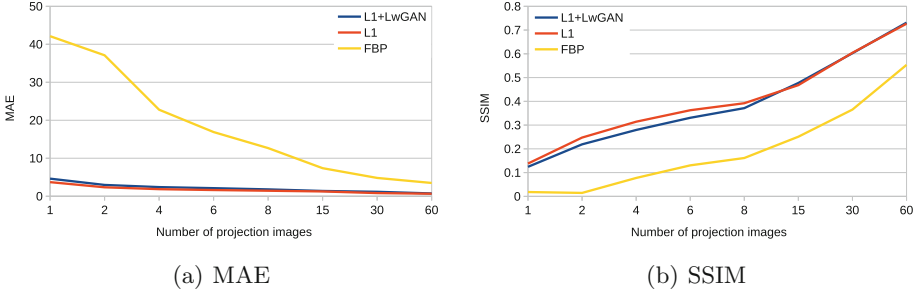
However, as GANs were initially proposed to generate new images from noise, its applicability to medical applications is an open question. In this work, we want to gain insights in the applicability of wGANs for improving the image quality for 2D CT image slice reconstruction from a limited number of projection images. We investigate the role of an additional content loss for improved reconstruction quality and provide insights in the amount of projection images that are necessary for anatomically correct reconstructions.

## 2 Method

In our deep learning based method we utilize wGANs with gradient penalty in combination with a content loss  $L_1$  to improve the reconstruction of 2D axial CT slices, see Fig. 1. Our method is trained to reconstruct the target 2D CT axial slice directly from a small number of 2D projection images generated by extending 1D projections of the target image, see Fig. 2.

**Projection Image Generation:** We generated a 1D sum projection  $s_{\alpha_i}$  from a target 2D axial CT slice  $y \in Y$  for different angles  $\alpha_i, i \in \{1, \dots, N\}$ , see Fig. 2. The angles  $\alpha$  are uniformly distributed in the range of  $0^\circ$  to  $180^\circ$  with a fixed angle between them. With the same size as  $y$ , the 2D projection image  $x_{\alpha_i}$  is generated by repeating  $s_{\alpha_i}$  in the direction of  $\alpha_i$ .

**wGAN Architecture:** Based on the U-Net [7], the generator  $G$  of wGAN uses a set of 2D projection images  $x_\alpha$  to generate a 2D image  $\hat{y} \in \hat{Y}$ , which is as similar as possible to  $y \in Y$ . Alternately receiving an image from  $Y$  and  $\hat{Y}$ ,



**Fig. 3.** Mean absolute error (MAE) and structural similarity index metric (SSIM) of our wGAN trained using  $L_1 + L_{wGAN}$  with  $\lambda = 10^{-3}$ , only  $L_1$  and the FBP method compared to the ground truth for a different number of projection images

the task of the discriminator  $D$  of wGAN is to recognize from which of these two distributions the currently observed image is coming. The architecture of  $D$  consists of consecutive 2D convolution layers and 2D max pooling layers, which are followed by a fully connected layer resulting in a single scalar value.

**Loss Functions:** The discriminator’s loss is defined as

$$L_D = -D(y) + D(\hat{y}) + \rho, \quad (1)$$

where  $D(y)$  is the discriminator’s predicted probability for  $y$  coming from  $Y$ ,  $D(\hat{y})$  is the predicted probability for  $\hat{y}$  coming also from  $Y$  and  $\rho$  is the gradient penalty, which is used to stabilize the training of the wGAN [2].

The generator’s loss is defined as

$$L_G = L_1 - \lambda \cdot D(\hat{y}) = L_1 + \lambda \cdot L_{wGAN}, \quad (2)$$

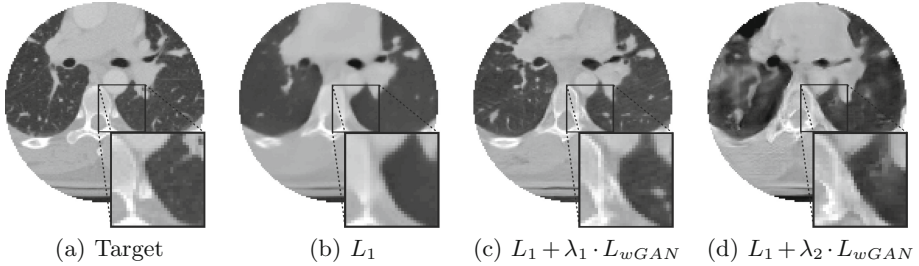
where  $\lambda$  is used as a weight between the adversarial loss  $L_{wGAN} = -D(\hat{y})$  and  $L_1$  loss, which is defined as

$$L_1 = \frac{1}{|M|} \sum_{m \in M} |\hat{y}_m - y_m|, \quad (3)$$

where  $m \in M$  are corresponding pixels in  $\hat{y}$  and  $y$ , and  $M$  is the set of all pixels.

## 2.1 Experimental Setup

Our data set consists of 10 3D CT images containing information from neck to pelvis. To decrease the training time, we downsampled the axial slices for all images to a size of  $128 \times 128$ . We separated the 3D CT images into eight training and two testing images. During training, the 2D target image is selected as a random axial slice from a training 3D CT image that is augmented on the fly by random translation, rotation and scaling coming from a uniform distribution. To prevent the problem of different amounts of image data present in projection



**Fig. 4.** The target compared to reconstruction results for eight projections generated by  $L_1$  and  $L_1 + L_{wGAN}$  with two different values for  $\lambda$ .  $\lambda_1 = 10^{-3}$  (default),  $\lambda_2 = 10^{-1}$ .

images from different angles when generated from a square shaped target image, all targets are masked by a circle. We used the same mask when the loss is calculated. We experiment with a different number  $N = \{1, 2, 4, 6, 8, 15, 30, 60\}$  of projection images used for reconstruction of 2D CT axial slice images. The results are compared quantitatively to the FBP method by calculating the mean absolute error (MAE) and the structural similarity index metric (SSIM). When results are compared qualitatively, all images share the same brightness setting, but some values are truncated to give a better contrast.

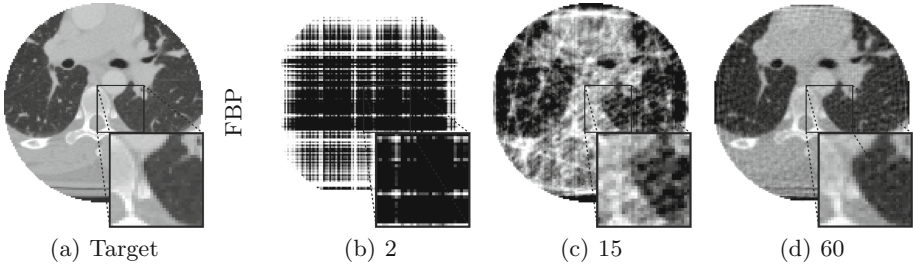
All networks were trained using a mini-batch size of 16 and 80.000 iterations, while the discriminator was trained five times for each iteration. We used Adam as an optimizer for all networks with a learning rate of 0.0001,  $\beta_1 = 0.5$  and  $\beta_2 = 0.9$ . We used a four level deep U-Net [7] as our generator. For both the generator and the discriminator we used a kernel size of  $3 \times 3$  and 64 intermediate convolutional filters. As activation function, we used ReLU for the generator and Leaky ReLU for the discriminator.

### 3 Results

Our results for a different number of projection images used for reconstruction of 2D CT axial slice images are presented quantitatively as MAE in Fig. 3(a) and as SSIM in Fig. 3(b). Qualitative results using eight projection images and a different weight factor  $\lambda$  are shown in Fig. 4. For a different number  $N \in \{2, 15, 60\}$  of projection images, Fig. 5 shows the qualitative results for the FBP method and Fig. 6 for using only  $L_1$  loss ( $\lambda = 0$ ) and  $L_1 + L_{wGAN}$  loss ( $\lambda = 10^{-3}$ ).

### 4 Discussion and Conclusion

In this work we investigated the potential use of wGANs for sparse-view CT slice reconstruction, which is motivated by a reduction of ionizing radiation exposure to the patient. While a content loss  $L_1$  enforces similarity to the target image, our U-net based CNN is optimized using a combination of the  $L_1$  and an adversarial



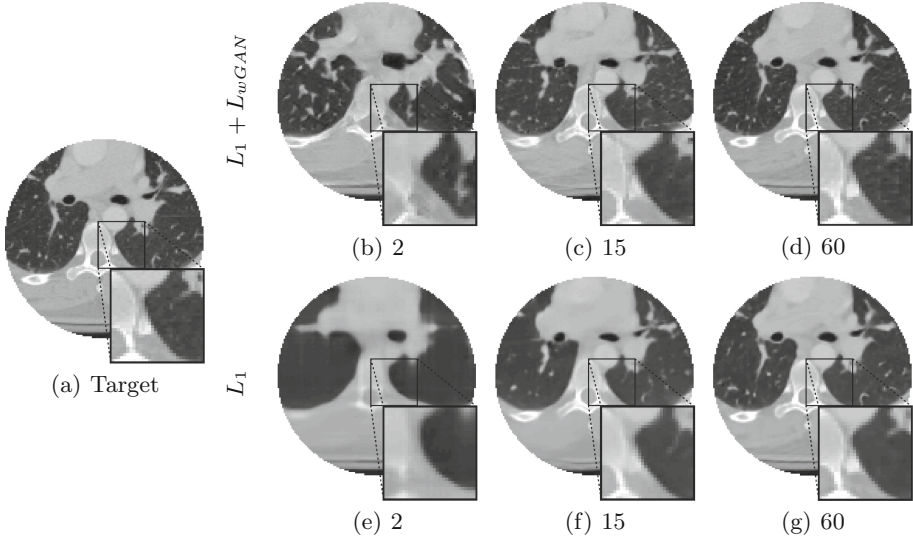
**Fig. 5.** The target image (a) compared to reconstruction results generated by the FBP method for two (b), 15 (c) and 60 (d) projection images.

loss  $L_{wGAN}$  (Eq. (2)) to reconstruct more realistically looking images. In contrast to other machine learning based approaches in which the reconstruction of a high quality CT image is learned from the previously reconstructed low quality CT image [3, 4, 13], in our approach the CNN learns the reconstruction directly from a limited number of projection images, see Fig. 1.

When a different number of projection images is used to train our CNNs, our quantitative results show that the learning based methods perform substantially better than the FBP, see Fig. 3, which is to be expected, since the FBP does not utilize any prior knowledge in contrast to the CNN based approaches. In terms of the MAE, the CNN trained on  $L_1$ -only performs slightly better than the wGAN trained on the combination of  $L_1$  and adversarial loss ( $L_1 + L_{wGAN}$ ). This was expected, since  $L_1$  loss is optimized to minimize MAE. By comparing the SSIM results, we can see that the CNN trained on  $L_1$ -only gives better results up to eight projection images, but from that point on the results from  $L_1$ -only and  $L_1 + L_{wGAN}$  can be considered equal. Although the quantitative results indicate that the CNNs trained on  $L_1$ -only provide a better reconstruction than on  $L_1 + L_{wGAN}$ , they have to be considered with caution, since MAE and SSIM do not represent the human perception of image quality well.

When training CNNs on  $L_1$ -only loss using a sparse number of projection images, the qualitative results show that the reconstructed image is blurry without fine structures and clear edges, see Fig. 4(b). Using an additional adversarial loss, the images contain fine structures and clear edges, see Fig. 4(c). However, when the adversarial loss dominates in the loss function, anatomical structures without correspondence to the target image can be introduced, see Fig. 4(d). We investigated the effect of  $\lambda$  by utilizing different orders of magnitude  $\lambda = 10^{\{-4, -3, -2, -1, 0\}}$  and found  $\lambda = 10^{-3}$  to be the optimum. While  $10^{-4}$  leads to results very similar to  $L_1$ -only and seemingly without an influence of  $L_{wGAN}$ , the results using  $10^{\{-2, -1, 0\}}$  lead to a clear reduction of structural similarity and thus a loss of anatomical correspondence to the target.

Our results using a different number of projection images in Fig. 5 confirm that the FBP method is not able to produce clinically meaningful images without a proper number of projections. On the other side, our machine learning based approach is able to reconstruct the main anatomical structures of the target



**Fig. 6.** The target image (a) compared to reconstruction results generated by  $L_1 + L_{wGAN}$  with  $\lambda = 10^{-3}$  (b, c and d) as well as by  $L_1$  (e, f and g) for two (b, e), 15 (c, f) and 60 (d, g) projection images.

image already from two projection images, see Fig. 6. While using  $L_1$ -only loss generates images that give the impression of a heavily blurred target image, the reconstructed image by  $L_1 + L_{wGAN}$  loss looks optically more realistic. However, for both reconstructions, the anatomical structures do not always correspond to the target due to a huge amount of missing information making them unsuitable for use in clinical practice. In our experiments we found that 15 projection images are sufficient for our CNN based approaches to achieve a qualitatively good reconstruction. However, the results generated by  $L_1 + L_{wGAN}$  are sharper and give more textural information compared to  $L_1$ -only loss. The results generated from 60 projection images provide a similar amount of fine details as the target image. Nevertheless, the  $L_1 + L_{wGAN}$  result is still slightly sharper than  $L_1$ -only loss, especially the fine details in the lung region are visible.

We showed that the combination of an adversarial loss  $L_{wGAN}$  and a content loss  $L_1$  improves the visual reconstruction quality. The reconstructions using  $L_1 + L_{wGAN}$  appear sharper and more structured compared to the CNN results trained on  $L_1$ -only. However, the tradeoff  $\lambda$  is crucial to reduce the amount of newly introduced information by the wGAN and guide the reconstruction in a direction close to the target image. While images generated by the CNNs trained on  $L_1$ -only appear blurry, the additional information present in the wGAN results trained on  $L_1 + L_{wGAN}$  can potentially lead to misinterpretation in a clinical relevant context if not enough data is available for reconstruction.

In conclusion, the wGANs have a potential to improve the perceived image quality even from a huge amount of missing information, however, it is dependent

on the application and domain, whether the kind of artifacts introduced are tolerable, which is an open question in medical imaging. To further evaluate anatomical correspondence, in our future work we will validate the perceived image quality of our approach by expert radiologists and also compare to other state-of-the-art methods based on compressed sensing.

## References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International Conference on Machine Learning, pp. 214–223 (2017)
2. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of Wasserstein GANs. In: Advances in Neural Information Processing Systems, pp. 5769–5779 (2017)
3. Jin, K.H., McCann, M.T., Froustey, E., Unser, M.: Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Process.* **26**(9), 4509–4522 (2017)
4. Kang, E., Min, J., Ye, J.C.: A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Med. Phys.* **44**(10), e360–e375 (2017)
5. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4681–4690 (2017)
6. Mardani, M., et al.: Deep generative adversarial networks for compressed sensing automates MRI. Preprint [arXiv:1706.00051](https://arxiv.org/abs/1706.00051) (2017)
7. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015 Part III. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
8. Seitzer, M., et al.: Adversarial and perceptual refinement for compressed sensing MRI reconstruction. Accepted at International Conference on Medical Image Computing and Computer-Assisted Intervention (2018)
9. Thaler, F., Payer, C., Štern, D.: Volumetric reconstruction from a limited number of digitally reconstructed radiographs using CNNs. In: Proceedings of the OAGM Workshop 2018, pp. 13–19. Verlag der TU Graz (2018)
10. Xie, S., et al.: Artifact removal using improved GoogLeNet for sparse-view CT reconstruction. *Sci. Rep.* **8**(1), 6700 (2018)
11. Yang, G., et al.: Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Trans. Med. Imaging* **37**(6), 1310–1321 (2018)
12. Yang, Q., et al.: Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE Trans. Med. Imaging* **37**(6), 1348–1357 (2018)
13. Yang, X., et al.: Low-dose X-ray tomography through a deep convolutional neural network. *Sci. Rep.* **8**(1), 2575 (2018)