Solving Low-Dose CT Reconstruction via GAN with Local Coherence

Wenjie Liu and Hu Ding[(✉)]

School of Computer Science and Technology, University of Science and Technology of China, Hefei, Anhui, China
lwj1217@mail.ustc.edu.cn, huding@ustc.edu.cn

Abstract. The Computed Tomography (CT) for diagnosis of lesions in human internal organs is one of the most fundamental topics in medical imaging. Low-dose CT, which offers reduced radiation exposure, is preferred over standard-dose CT, and therefore its reconstruction approaches have been extensively studied. However, current low-dose CT reconstruction techniques mainly rely on model-based methods or deep-learning-based techniques, which often ignore the coherence and smoothness for sequential CT slices. To address this issue, we propose a novel approach using generative adversarial networks (GANs) with enhanced local coherence. The proposed method can capture the local coherence of adjacent images by optical flow, which yields significant improvements in the precision and stability of the constructed images. We evaluate our proposed method on real datasets and the experimental results suggest that it can outperform existing state-of-the-art reconstruction approaches significantly.

Keywords: CT reconstruction · Low-dose · Generative adversarial networks · Local coherence · Optical flow

1 Introduction

Computed Tomography (CT) is one of the most widely used technologies in medical imaging, which can assist doctors for diagnosing the lesions in human internal organs. Due to harmful radiation exposure of standard-dose CT, the low dose CT is more preferable in clinical application [4,6,34]. However, when the dose is low together with the issues like sparse-view or limited angles, it becomes quite challenging to reconstruct high-quality CT images. The high-quality CT images are important to improve the performance of diagnosis in clinic [27]. In mathematics, we model the CT imaging as the following procedure

$$\mathbf{y} = \mathcal{T}(\mathbf{x^r}) + \delta, \tag{1}$$

where $\mathbf{x^r} \in \mathbb{R}^d$ denotes the **unknown** ground-truth picture, $\mathbf{y} \in \mathbb{R}^m$ denotes the received measurement, and δ is the noise. The function $\mathcal{T}$ represents the forward operator that is analogous to the Radon transform, which is widely used in medical imaging [23,28]. The problem of CT reconstruction is to recover $\mathbf{x^r}$ from the received $\mathbf{y}$.

Solving the inverse problem of (1) is often very challenging if there is no any additional information. If the forward operator $\mathcal{T}$ is well-posed and δ is neglectable, we know that an approximate $\mathbf{x^r}$ can be easily obtained by directly computing $\mathcal{T}^{-1}(\mathbf{y})$. However, $\mathcal{T}$ is often ill-posed, which means the inverse function $\mathcal{T}^{-1}$ does not exist and the inverse problem of (1) may have multiple solutions. Moreover, when the CT imaging is low-dose, the filter backward projection (FBP) [11] can produce serious detrimental artifact. Therefore, most of existing approaches usually incorporate some prior knowledge during the reconstruction [14,17,26]. For example, a commonly used method is based on regularization:

$$\mathbf{x} = argmin_{\mathbf{x}} \|\mathcal{T}(\mathbf{x}) - \mathbf{y}\|_p + \lambda\mathcal{R}(\mathbf{x}), \tag{2}$$

where $\|\cdot\|_p$ denotes the p-norm and $\mathcal{R}(\mathbf{x})$ denotes the penalty item from some prior knowledge.

In the past years, a number of methods have been proposed for designing the regularization $\mathcal{R}$. The traditional model-based algorithms, e.g., the ones using total variation [3,26], usually apply the sparse gradient assumptions and run an iterative algorithm to learn the regularizers [12,18,24,29]. Another popular line for learning the regularizers comes from deep learning [13,17]; the advantage of the deep learning methods is that they can achieve an end-to-end recovery of the true image $\mathbf{x^r}$ from the measurement $\mathbf{y}$ [1,21]. Recent researches reveal that convolutional neural networks (CNNs) are quite effective for image denoising, e.g., the CNN based algorithms [10,34] can directly learn the reconstructed mapping from initial measurement reconstructions (e.g., FBP) to the ground-truth images. The dual-domain network that combines the sinograms with reconstructed low-dose CT images were also proposed to enhance the generalizability [15,30].

A major drawback of the aforementioned reconstruction methods is that they deal with the input CT 2D slices independently (note that the goal of CT reconstruction is to build the 3D model of the organ). Namely, the neighborhood correlations among the 2D slices are often ignored, which may affect the reconstruction performance in practice. In the field of computer vision, "optical flow" is a common technique for tracking the motion of object between consecutive frames, which has been applied to many different tasks like video generation [35], prediction of next frames [22] and super resolution synthesis [5,31]. To estimate the optical flow field, existing approaches include the traditional brightness gradient methods [2] and the deep networks [7]. The idea of optical flow has also been used for tracking the organs movement in medical imaging [16,20,33]. However, to the best of our knowledge, there is no work considering GANs with using optical flow to capture neighbor slices coherence for low dose 3D CT reconstruction.

In this paper, we propose a novel optical flow based generative adversarial network for 3D CT reconstruction. Our intuition is as follows. When a patient is located in a CT equipment, a set of consecutive cross-sectional images are generated. If the vertical axial sampling space of transverse planes is small, the corresponding CT slices should be highly similar. So we apply optical flow, though there exist several technical issues waiting to solve for the design and implementation, to capture the local coherence of adjacent CT images for reducing the artifacts in low-dose CT reconstruction. Our contributions are summarized below:

1. We introduce the "local coherence" by characterizing the correlation of consecutive CT images, which plays a key role for suppressing the artifacts.
2. Together with the local coherence, our proposed generative adversarial networks (GANs) can yield significant improvement for texture quality and stability of the reconstructed images.
3. To illustrate the efficiency of our proposed approach, we conduct rigorous experiments on several real clinical datasets; the experimental results reveal the advantages of our approach over several state-of-the-art CT reconstruction methods.

2   Preliminaries

In this section, we briefly review the framework of the ordinary generative adversarial network, and also introduce the local coherence of CT slices.

Generative Adversarial Network. Traditional generative adversarial network [8] consists of two main modules, a generator and a discriminator. The generator $\mathcal{G}$ is a mapping from a latent-space Gaussian distribution $\mathbb{P}_Z$ to the synthetic sample distribution $\mathbb{P}_{X_G}$, which is expected to be close to the real sample distribution $\mathbb{P}_X$. On the other hand, the discriminator $\mathcal{D}$ aims to maximize the distance between the distributions $\mathbb{P}_{X_G}$ and $\mathbb{P}_X$. The game between the generator and discriminator actually is an adversarial process, where the overall optimization objective follows a min-max principle:

$$\min_{\mathcal{G}} \max_{\mathcal{D}} \mathbb{E}_{\mathbf{x^r} \sim \mathbb{P}_X, \mathbf{z} \sim \mathbb{P}_Z} (\log(\mathcal{D}(\mathbf{x^r})) + \log(1 - \mathcal{D}(\mathcal{G}(\mathbf{z})))). \tag{3}$$

Local Coherence. As mentioned in Sect. 1, optical flow can capture the temporal coherence of object movements, which plays a crucial role in many video-related tasks. More specifically, the optical flow refers to the instantaneous velocity of pixels of moving objects on consecutive frames over a short period of time [2]. The main idea relies on the practical assumptions that the brightness of the object more likely remains stable across consecutive frames, and the brightness of the pixels in a local region are usually changed consistently [9].

Based on these assumptions, the brightness of optical flow can be described by the following equation:

$$\nabla I_w \cdot v_w + \nabla I_h \cdot v_h + \nabla I_t = 0, \tag{4}$$

where $v = (v_w, v_h)$ represents the optical flow of the position (w, h) in the image. $\nabla I = (\nabla I_w, \nabla I_h)$ denotes spatial gradients of image brightness, and ∇I_t denotes the temporal partial derivative of the corresponding region.

Following the Eq. (4), we consider the question that whether the optical flow idea can be applied to 3D CT reconstruction. In practice, the brightness of adjacent CT images often has very tiny difference, due to the inherent continuity and structural integrity of human body. Therefore, we introduce the "local coherence" that indicates the correlation between adjacent images of a tissue. Namely, adjacent CT images often exhibit significant similarities within a certain local range along the vertical axis of the human body. Due to the local coherence, the noticeable variations observed in CT slices within the local range often occur at the edges of organs. We can substitute the temporal partial derivative ∇I_t by the vertical axial partial derivative ∇I_z in the Eq. (4), where "z" indicates the index of the vertical axis. As illustrated in Fig. 1, the local coherence can be captured by the optical flow between adjacent CT slices.
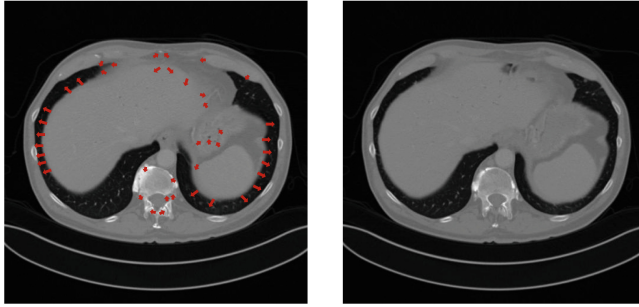

Fig. 1. The optical flow between two adjacent CT slices. The scanning window of X-ray slides from the position of the left image to the position of the right image. The directions and lengths of the red arrows represent the optical flow field. The left and right images share the local coherence and thus the optical flows are small. (Color figure online)

3   GANs with Local Coherence

In this section, we introduce our low-dose CT image generation framework with local coherence in detail.

The Framework of Our Network. The proposed framework comprises three components, including a generator $\mathcal{G}$, a discriminator $\mathcal{D}$ and an optical flow estimator $\mathcal{F}$. The generator is the core component, and the flow estimator provides auxiliary warping images for the generation process.

Suppose we have a sequence of measurements $\mathbf{y_1}, \mathbf{y_2}, \cdots, \mathbf{y_n}$; for each $\mathbf{y_i}$, $1 \leq i \leq n$, we want to reconstruct its ground truth image $\mathbf{x_i^r}$ as the Eq. (1). Before performing the reconstruction in the generator $\mathcal{G}$, we apply some prior knowledge in physics and run filter backward projection on the measurement $\mathbf{y_i}$ in Eq. (1) to obtain an initial recovery solution $\mathbf{s_i}$. Usually $\mathbf{s_i}$ contains significant noise comparing with the ground truth $\mathbf{x_i^r}$. Then the network has two input components, i.e., the initial backward projected image $\mathbf{s_i}$ that serves as an approximation of the ground truth $\mathbf{x_i^r}$, and a set of neighbor CT slices $\mathcal{N}(\mathbf{s_i}) = \{\mathbf{s_{i-1}}, \mathbf{s_{i+1}}\}^1$ for preserving the local coherence. The overall structure of our framework is shown in Fig. 2. Below, we introduce the three key parts of our framework separately.
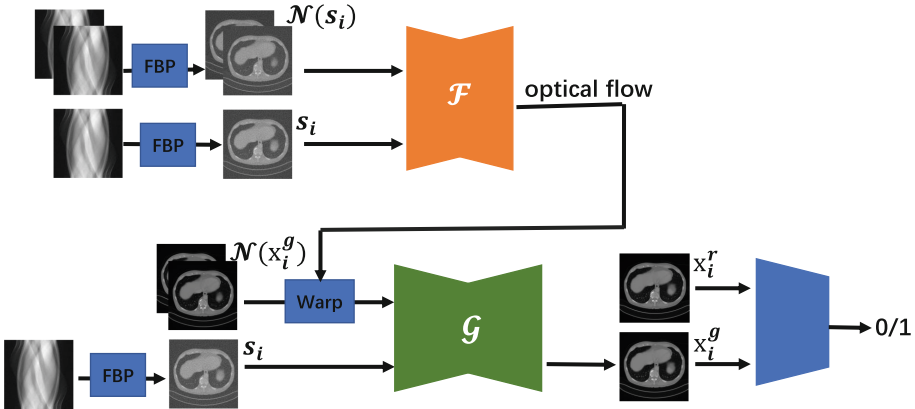

Fig. 2. The framework of our generate adversarial network with local coherence for CT reconstruction.

Optical Flow Estimator. The optical flow $\mathcal{F}(\mathcal{N}(\mathbf{s_i}), \mathbf{s_i})$ denotes the brightness changes of pixels from $\mathcal{N}(\mathbf{s_i})$ to $\mathbf{s_i}$, where it captures their local coherence. The estimator is derived by the network architecture of FlowNet [7]. The FlowNet is an autoencoder architecture with extraction of features of two input frames to learn the corresponding flow, which is consist of 6 (de)convolutional layers for both encoder and decoder.

Discriminator. The discriminator $\mathcal{D}$ assigns the label "1" to real standard-dose CT images and "0" to generated images. The goal of $\mathcal{D}$ is to maximize the separation between the distributions of real images and generated images:

---

^{1} If $i = 1$, $\mathcal{N}(\mathbf{s_i}) = \{\mathbf{s_2}\}$; if $i = n$, $\mathcal{N}(\mathbf{s_i}) = \{\mathbf{s_{n-1}}\}$.

$$\mathcal{L}_{\mathcal{D}} = \sum_{i=1}^{n} -(\log(\mathcal{D}(\mathbf{x_i^r})) + \log(1 - \mathcal{D}(\mathbf{x_i^g}))), \qquad (5)$$

where $\mathbf{x_i^g}$ is the image generated by $\mathcal{G}$ (the formal definition for $\mathbf{x_i^g}$ will be introduced below). The discriminator includes 3 residual blocks, with 4 convolutional layers in each residual block.

Generator. We use the generator $\mathcal{G}$ to reconstruct the high-quality CT image for the ground truth $\mathbf{x_i^r}$ from the low-dose image $\mathbf{s_i}$. The generated image is obtained by

$$\begin{aligned} \mathbf{x_i^g} &= \mathcal{G}(\mathbf{s_i}, \mathcal{W}(\mathcal{N}(\mathbf{x_i^g}))); \\ \mathcal{N}(\mathbf{x_i^g}) &= \mathcal{G}(\mathcal{N}(\mathbf{s_i})), \end{aligned} \qquad (6)$$

where $\mathcal{W}(\cdot)$ is the warping operator. Before generating $\mathbf{x_i^g}$, $\mathcal{N}(\mathbf{x_i^g})$ is reconstructed from $\mathcal{N}(\mathbf{s_i})$ by the generator without considering local coherence. Subsequently, according to the optical flow $\mathcal{F}(\mathcal{N}(\mathbf{s_i}), \mathbf{s_i})$, we warp the reconstructed images $\mathcal{N}(\mathbf{x_i^g})$ to align with the current slice by adjusting the brightness values. The warping operator $\mathcal{W}$ utilizes bi-linear interpolation to obtain $\mathcal{W}(\mathcal{N}(\mathbf{x_i^g}))$, which enables the model to capture subtle variations in the tissue from the generated $\mathcal{N}(\mathbf{x_i^g})$; also, the warping operator can reduce the influence of artifacts for the reconstruction. Finally, $\mathbf{x_i^g}$ is generated by combining $\mathbf{s_i}$ and $\mathcal{W}(\mathcal{N}(\mathbf{x_i^g}))$. Since $\mathbf{x_i^r}$ is our target for reconstruction in the i-th batch, we consider the difference between $\mathbf{x_i^g}$ and $\mathbf{x_i^r}$ in the loss. Our generator is mainly based on the network architecture of Unet [25]. Partly inspired by the loss in [5], the optimization objective of the generator G comprises three items with the coefficients $\lambda_{\mathtt{pix}}, \lambda_{\mathtt{adv}}, \lambda_{\mathtt{per}} \in (0, 1)$:

$$\mathcal{L}_{\mathcal{G}} = \lambda_{\mathtt{pix}}\mathcal{L}_{\mathtt{pixel}} + \lambda_{\mathtt{adv}}\mathcal{L}_{\mathtt{adv}} + \lambda_{\mathtt{per}}\mathcal{L}_{\mathtt{percept}}. \qquad (7)$$

In (7), "$\mathcal{L}_{\mathtt{pixel}}$" is the loss measuring the pixel-wise mean square error of the generated image $\mathbf{x_i^g}$ with respect to the ground-truth $\mathbf{x_i^r}$. "$\mathcal{L}_{\mathtt{adv}}$" represents the adversarial loss of the discriminator $\mathcal{D}$, which is designed to minimize the distance between the generated standard-dose CT image distribution $\mathbb{P}_{X_G}$ and the real standard-dose CT image distribution $\mathbb{P}_X$. "$\mathcal{L}_{\mathtt{percept}}$" denotes the perceptual loss, which quantifies the dissimilarity between the feature maps of $\mathbf{x_i^r}$ and $\mathbf{x_i^g}$; the feature maps denote the feature representation extracted from the hidden layers in the discriminator $\mathcal{D}$ (suppose there are t hidden layers):

$$\mathcal{L}_{percept} = \sum_{i=1}^{n} \sum_{j=1}^{t} \|\mathcal{D}_j(\mathbf{x_i^r}) - \mathcal{D}_j(\mathbf{x_i^g})\|_1 \qquad (8)$$

where $\mathcal{D}_j(\cdot)$ refers to the feature extraction performed on the j-th hidden layer. Through capturing the high frequency differences in CT images, $\mathcal{L}_{\mathtt{percept}}$ can enhance the sharpness for edges and increase the contrast for the reconstructed images. $\mathcal{L}_{\mathtt{pixel}}$ and $\mathcal{L}_{\mathtt{adv}}$ are designed to recover global structure, and $\mathcal{L}_{\mathtt{percept}}$ is utilized to incorporate additional texture details into the reconstruction process.

4 Experiment

Datasets. First, our proposed approaches are evaluated on the "Mayo-Clinic low-dose CT Grand Challenge" (Mayo-Clinic) dataset of lung CT images [19]. The dataset contains 2250 two dimensional slices from 9 patients for training, and the remaining 128 slices from 1 patient are reserved for testing. The low-dose measurements are simulated by parallel-beam X-ray with 200 (or 150) uniform views, i.e., $N_v = 200$ (or $N_v = 150$), and 400 (or 300) detectors, i.e., $N_d = 400$ (or $N_d = 300$). In order to further verify the denoising ability of our approaches, we add the Gaussian noise with standard deviation $\sigma = 2.0$ to the sinograms after X-ray projection in 50% of the experiments. To evaluate the generalization of our model, we also consider another dataset RIDER with non-small cell lung cancer under two CT scans [36] for testing. We randomly select 4 patients with 1827 slices from the dataset. The simulation process is identical to that of Mayo-Clinic. The proposed networks were implemented in the PyTorch framework and trained on Nvidia 3090 GPU with 100 epochs.

Baselines and Evaluation Metrics. We consider several existing popular algorithms for comparison. (1) **FBP** [11]: the classical filter backward projection on low-dose sinograms. (2) **FBPConvNet** [10]: a direct inversion network followed by the CNN after initial **FBP** reconstruction. (3) **LPD** [1]: a deep learning method based on proximal primal-dual optimization. (4) **UAR** [21]: an end-to-end reconstruction method based on learning unrolled reconstruction operators and adversarial regularizers. Our proposed method is denoted by **GAN-LC**. We set $\lambda_{\texttt{pix}} = 1.0, \lambda_{\texttt{adv}} = 0.01$ and $\lambda_{\texttt{per}} = 1.0$ for the optimization objective in Eq. (7) during our training process. Following most of the previous articles on 3D CT reconstruction, we evaluate the experimental performance by two metrics: the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM) [32]. PSNR measures the pixel differences of two images, which is negatively correlated with mean square error. SSIM measures the structure similarity between two images, which is related to the variances of the input images. For both two measures, the higher the better.

Results. Table 1 presents the results on the Mayo-Clinic dataset, where the first row represents different parameter settings (i.e., the number of uniform views N_v, the number of detectors N_d and the standard deviation of Gaussian noise σ) for simulating low-dose sinograms. Our proposed approach **GAN-LC** consistently outperforms the baselines under almost all the low-dose parameter settings. The methods **FBP** and **UAR** are very sensitive to noise; the performance of **LPD** is relatively stable but with low reconstruction accuracy. **FBPConvNet** has

Table 1. Experimental results for Mayo-Clinic dataset. The value in first row of the table represents N_v, N_d and σ for simulating low-dose sinograms, respectively.

Sinograms	200, 400, 0.0		200, 400, 2.0		150, 300, 0.0		150, 300, 2.0	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FBP	26.449	0.721	13.517	0.191	21.460	0.616	12.593	0.168
FBPConvNet	38.213	0.918	30.148	0.743	35.263	0.869	29.095	0.723
LPD	28.050	0.844	28.357	0.794	28.376	0.826	27.409	**0.801**
UAR	33.248	0.902	22.048	0.272	29.829	0.848	21.227	0.238
GAN-LC	**39.548**	**0.950**	**32.437**	**0.819**	**36.542**	**0.899**	**31.586**	0.725

Table 2. Experimental results for RIDER dataset. The value in first row of the table represents N_v, N_d and σ for simulating low-dose sinograms, respectively.

Sinograms	200, 400, 0.0		200, 400, 2.0		150, 300, 0.0		150, 300, 2.0	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
FBP	21.398	0.647	15.609	0.233	19.49	0.597	14.845	0.203
FBPConvNet	27.256	0.671	19.520	0.444	27.504	0.650	18.517	0.431
LPD	22.341	0.615	12.196	0.466	22.172	0.556	12.215	0.455
UAR	24.915	0.667	20.943	0.207	21.136	0.557	**19.873**	0.176
GAN-LC	**28.861**	**0.721**	**22.624**	**0.517**	**29.171**	**0.705**	19.607	**0.470**

a similar increasing trend with our approach across different settings but has worse reconstruction quality. To evaluate the stability and generalization of our model and the baselines trained on Mayo-Clinic dataset, we also test them on the RIDER dataset. The results are shown in Table 2. Due to the bias in the datasets collected from different facilities, the performances of all the models are declined to some extents. But our proposed approach still outperforms the other models for most testing cases.

To illustrate the reconstruction performances more clearly, we also show the reconstruction results for testing images in Fig. 3. We can see that our network can reconstruct the CT image with higher quality. Due to the space limit, the experimental results of different views N_v and more visualized results are placed in our supplementary material.
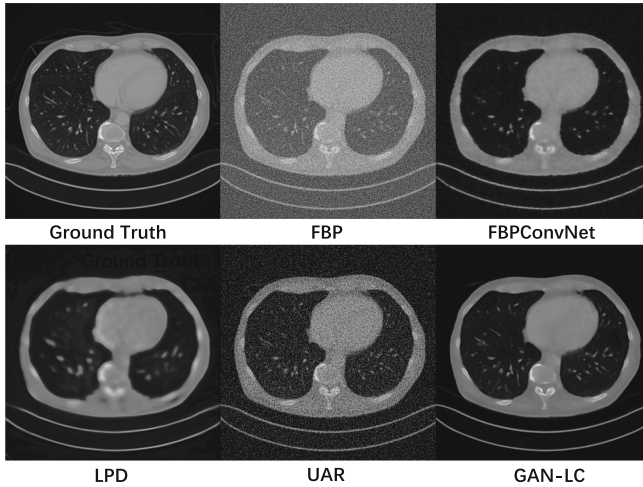
Fig. 3. Reconstruction results on Mayo-Clinic dataset. The sparse view setting of sinograms is $N_v = 200$, $N_d = 400$ and $\sigma = 2.0$. "Ground Truth" is the standard-dose CT image.

5   Conclusion

In this paper, we propose a novel approach for low-dose CT reconstruction using generative adversarial networks with local coherence. By considering the inherent continuity of human body, local coherence can be captured through optical flow, which is small deformations and structural differences between consecutive CT slices. The experimental results on real datasets demonstrate the advantages of our proposed network over several popular approaches. In future, we will evaluate our network on real-world CT images from local hospital and use the reconstructed images to support doctors for the diagnosis and recognition of lung nodules. Our code is publicly available at https://github.com/lwjie595/GANLC.

References

1. Adler, J., Öktem, O.: Learned primal-dual reconstruction. IEEE Trans. Med. Imaging **37**(6), 1322–1332 (2018)
2. Beauchemin, S.S., Barron, J.L.: The computation of optical flow. ACM Comput. Surv. (CSUR) **27**(3), 433–466 (1995)
3. Chambolle, A.: An algorithm for total variation minimization and applications. J. Math. Imaging Vis. **20**(1), 89–97 (2004)
4. Chen, H., et al.: Low-dose CT via convolutional neural network. Biomed. Opt. Express **8**(2), 679–694 (2017)

5. Chu, M., Xie, Y., Mayer, J., Leal-Taixé, L., Thuerey, N.: Learning temporal coherence via self-supervision for GAN-based video generation. ACM Trans. Graph. (TOG) **39**(4), 75-1 (2020)

6. Ding, Q., Nan, Y., Gao, H., Ji, H.: Deep learning with adaptive hyper-parameters for low-dose CT image reconstruction. IEEE Trans. Comput. Imaging **7**, 648–660 (2021)

7. Dosovitskiy, A., et al.: Flownet: learning optical flow with convolutional networks. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 2758–2766 (2015)

8. Goodfellow, I.J., et al.: Generative adversarial networks. CoRR abs/1406.2661 (2014). https://arxiv.org/abs/1406.2661

9. Horn, B.K., Schunck, B.G.: Determining optical flow. Artif. Intell. **17**(1–3), 185–203 (1981)

10. 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)

11. Kak, A.C., Slaney, M.: Principles of computerized tomographic imaging. SIAM (2001)

12. Knoll, F., Bredies, K., Pock, T., Stollberger, R.: Second order total generalized variation (TGV) for MRI. Magn. Reson. Med. **65**(2), 480–491 (2011)

13. Kobler, E., Effland, A., Kunisch, K., Pock, T.: Total deep variation for linear inverse problems. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7549–7558 (2020)

14. Li, H., Schwab, J., Antholzer, S., Haltmeier, M.: NETT: solving inverse problems with deep neural networks. Inverse Prob. **36**(6), 065005 (2020)

15. Lin, W.A., et al.: Dudonet: dual domain network for CT metal artifact reduction. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10512–10521 (2019)

16. Liu, H., Lin, Y., Ibragimov, B., Zhang, C.: Low dose 4D-CT super-resolution reconstruction via inter-plane motion estimation based on optical flow. Biomed. Signal Process. Control **62**, 102085 (2020)

17. Lunz, S., Öktem, O., Schönlieb, C.B.: Adversarial regularizers in inverse problems. In: Advances in Neural Information Processing Systems, vol. 31 (2018)

18. McCann, M.T., Nilchian, M., Stampanoni, M., Unser, M.: Fast 3D reconstruction method for differential phase contrast X-ray CT. Opt. Express **24**(13), 14564–14581 (2016)

19. McCollough, C.: TU-FG-207A-04: overview of the low dose CT grand challenge. Med. Phys. **43**(6Part35), 3759–3760 (2016)

20. Mira, C., Moya-Albor, E., Escalante-Ramírez, B., Olveres, J., Brieva, J., Vallejo, E.: 3D hermite transform optical flow estimation in left ventricle CT sequences. Sensors **20**(3), 595 (2020)

21. Mukherjee, S., Carioni, M., Öktem, O., Schönlieb, C.B.: End-to-end reconstruction meets data-driven regularization for inverse problems. Adv. Neural. Inf. Process. Syst. **34**, 21413–21425 (2021)

22. Patraucean, V., Handa, A., Cipolla, R.: Spatio-temporal video autoencoder with differentiable memory. arXiv preprint arXiv:1511.06309 (2015)

23. Ramm, A.G., Katsevich, A.I.: The Radon Transform and Local Tomography. CRC Press, Boca Raton (2020)

24. Romano, Y., Elad, M., Milanfar, P.: The little engine that could: regularization by denoising (RED). SIAM J. Imag. Sci. **10**(4), 1804–1844 (2017)

25. 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. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

26. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)

27. Sori, W.J., Feng, J., Godana, A.W., Liu, S., Gelmecha, D.J.: DFD-Net: lung cancer detection from denoised CT scan image using deep learning. Front. Comp. Sci. **15**, 1–13 (2021)

28. Toft, P.: The radon transform. Theory and Implementation (Ph.D. dissertation), Technical University of Denmark, Copenhagen (1996)

29. Venkatakrishnan, S.V., Bouman, C.A., Wohlberg, B.: Plug-and-play priors for model based reconstruction. In: 2013 IEEE Global Conference on Signal and Information Processing, pp. 945–948. IEEE (2013)

30. Wang, C., Shang, K., Zhang, H., Li, Q., Zhou, S.K.: DuDoTrans: dual-domain transformer for sparse-view CT reconstruction. In: Haq, N., Johnson, P., Maier, A., Qin, C., Würfl, T., Yoo, J. (eds.) MLMIR 2022. LNCS, vol. 13587, pp. 84–94. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-17247-2_9

31. Wang, T.C., et al.: Video-to-video synthesis. arXiv preprint arXiv:1808.06601 (2018)

32. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)

33. Weng, N., Yang, Y.H., Pierson, R.: Three-dimensional surface reconstruction using optical flow for medical imaging. IEEE Trans. Med. Imaging **16**(5), 630–641 (1997)

34. Wolterink, J.M., Leiner, T., Viergever, M.A., Išgum, I.: Generative adversarial networks for noise reduction in low-dose CT. IEEE Trans. Med. Imaging **36**(12), 2536–2545 (2017)

35. Xue, T., Wu, J., Bouman, K., Freeman, B.: Visual dynamics: probabilistic future frame synthesis via cross convolutional networks. In: Advances in Neural Information Processing Systems, vol. 29 (2016)

36. Zhao, B., et al.: Evaluating variability in tumor measurements from same-day repeat CT scans of patients with non-small cell lung cancer. Radiology **252**(1), 263–272 (2009)

Image Registration