X2Vision: 3D CT Reconstruction from Biplanar X-Rays with Deep Structure Prior

Alexandre Cafaro[1,2,3(✉)], Quentin Spinat[1], Amaury Leroy[1,2,3], Pauline Maury[2], Alexandre Munoz[3], Guillaume Beldjoudi[3], Charlotte Robert[2], Eric Deutsch[2], Vincent Grégoire[3], Vincent Lepetit[4], and Nikos Paragios[1]

[1] TheraPanacea, Paris, France
a.cafaro@therapanacea.eu
[2] Gustave Roussy, Inserm 1030, Paris-Saclay University, Villejuif, France
[3] Department of Radiation Oncology, Centre Léon Bérard, Lyon, France
[4] LIGM, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Paris, France

Abstract. We propose an unsupervised deep learning method to reconstruct a 3D tomographic image from biplanar X-rays, to reduce the number of required projections, the patient dose, and the acquisition time. To address this ill-posed problem, we introduce prior knowledge of anatomic structures by training a generative model on 3D CTs of head and neck. We optimize the latent vectors of the generative model to recover a volume that both integrates this prior knowledge and ensures consistency between the reconstructed image and input projections. Our method outperforms recent methods in terms of reconstruction error while being faster and less radiating than current clinical workflow. We evaluate our method in a clinical configuration for radiotherapy.

Keywords: Image reconstruction · Inverse problem · Sparse sampling · Deep generative model · CT

1 Introduction

Tomographic imaging estimates body density using hundreds of X-ray projections, but it's slow and harmful to patients. Acquisition time may be too high for certain applications, and each projection adds dose to the patient. A quick, low-cost 3D estimation of internal structures using only bi-planar X-rays can revolutionize radiology, benefiting dental imaging, orthopedics, neurology, and more. This can improve image-guided therapies and preoperative planning, especially for radiotherapy, which requires precise patient positioning with minimal radiation exposure.

However, this task is an ill-posed inverse problem: X-ray measurements are the result of attenuation integration across the body, which makes them very

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43999-5_66.

H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14229, pp. 699–709, 2023.
https://doi.org/10.1007/978-3-031-43999-5_66

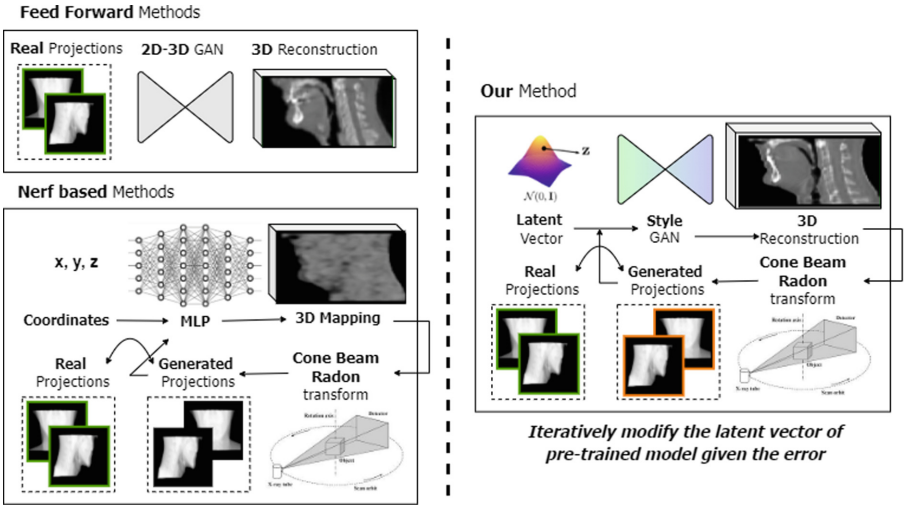

Fig. 1. Current methods vs our method. Feed-forward methods do not manage to predict a detailed and matching tomographic volume from a few projections. Iterative methods based on neural radiance fields lack prior for good reconstruction. By learning an embedding for the possible volumes, we can recover an accurate volume from very few projections with an optimization based on a Bayesian formulation.

ambiguous. Traditional reconstruction methods require hundreds of projections to get sufficient constraints on the internal structures. With very few projections, it is very difficult to disentangle the structures for even coarse 3D estimation. In other words, many 3D volumes may have generated such projections *a priori*.

Classical analytical and iterative methods [8] fail when very few projections are available. Several works have attempted to largely decrease the number of projections needed for an accurate volumetric reconstruction. Some deep learning methods [7,12,24,25,30] predict directly a 3D volume in a forward way from very few projections. The volume is however not guaranteed to be consistent with the projections and it is not clear which solution is retrieved. Other recent methods have adapted NeRFs [20] to tomographic reconstruction [23,31]. These non-learning methods show good results when the number of input projections remains higher than a dozen but fail when very few projections are provided, as our experiments in Sect. 3.3 show.

As illustrated in Fig. 1, to be able to reconstruct a volume accurately given as low as two projections only, we first learn a prior on the volume. To do this, we leverage the potential of generative models to learn a low-dimensional manifold of the target body part. Given projections, we find by a Bayesian formulation the intermediate latent vectors conditioning the generative model that minimize the error between synthesized projections of our reconstruction and these input projections. Our work builds on Hong et al. [10]'s 3D style-based generative model, which we extend via a more complex network and training framework.

Compared to other 3D GANs, it is proven to provide the best disentanglement of the feature space related to semantic features [2].

By contrast with feed-forward methods, our approach does not require paired projections-reconstructions, which are very tedious to acquire, and it can be used with different numbers of projections and different projection geometries without retraining. Compared to NeRF-based methods, our method exploits prior knowledge from many patients to require only two projections. We evaluate our method on reconstructing cancer patients' head-and-neck CTs, which involves intricate and complicated structures. We perform several experiments to compare our method with a feed-forward-based method [30] and a recent NeRF-based method [23], which are the previous state-of-the-art methods for the very few or few projections cases, respectively.

We show that our method allows to retrieve results with the finest reconstructions and better matching structures, for a variety of number of projections. To summarize, our contributions are two-fold: (i) A new paradigm for 3D reconstruction with biplanar X-rays: instead of learning to invert the measurements, we leverage a 3D style-based generative model to learn deep image priors of anatomic structures and optimize over the latent space to match the input projections; (ii) A novel unsupervised method, fast and robust to sampling ratio, source energy, angles and geometry of projections, all of which making it general for downstream applications and imaging systems.

2 Method

Figure 2 gives an overview of the pipeline we propose. We first learn the low-dimensional manifold of CT volumes of a target body region. At inference, we estimate the Maximum A Posteriori (MAP) volume on this manifold given very few projections: we find the latent vectors that minimize the error between the synthetic projections from the corresponding volume on the manifold and the real ones. In this section, we formalize the problem, describe how we learn the manifold, and detail how we optimize the latent vectors.

2.1 Problem Formulation

Given a small set of projections $\{I_i\}_i$, possibly as few as two, we would like to reconstruct the 3D tomographic volume v that generates these projections. This is a hard ill-posed problem, and to solve it, we need prior knowledge about the possible volumes. To do this, we look for the maximum *a posteriori* (MAP) estimate given the projections $\{I_i\}_i$:

$$v^* = \underset{v}{\operatorname{argmax}}\, p(v|\{I_i\}_i) = \underset{v}{\operatorname{argmax}}\, p(v)p(\{I_i\}_i|v) = \underset{v}{\operatorname{argmin}} \sum_i \mathcal{L}(v, I_i) + R(v)\,. \tag{1}$$

Term $\mathcal{L}(v, I_i)$ is a log-likelihood. We take it as:

$$\mathcal{L}(v, I_i) = \lambda_2 \left\| A_i \circ v - I_i \right\|_2 + \lambda_p \mathcal{L}_p(A_i \circ v, I_i)\,, \tag{2}$$

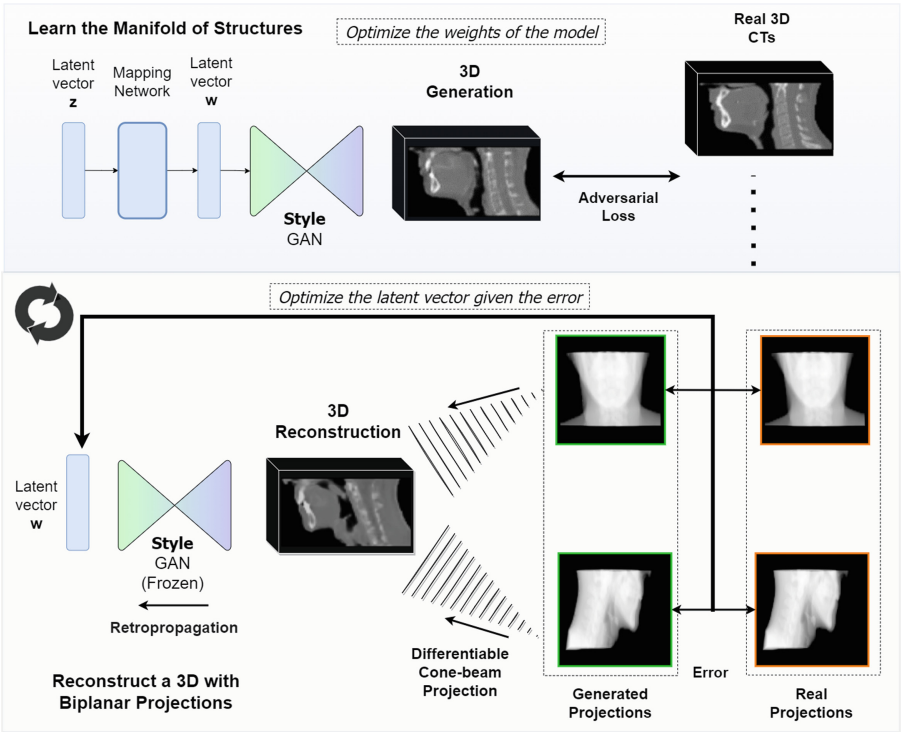

Fig. 2. Our pipeline. We first learn the low-dimensional manifold of 3D structures using a generative model. Then, given projections, we find the latent vectors that minimize the error between the projections of our generation and the input projections.

where A_i is an operator that projects volume v under view i. We provide more details about operator A in Sect. 2.3. $\mathcal{L}_p$ is the perceptual loss [13] between projection of v and the observed projection I_i. Term $R(v)$ is a regularization term. It is crucial as it is the term that embodies prior knowledge about the volume to reconstruct. As discussed in the introduction, we rely on a generative model, which we describe in the next section. Then, we describe how exactly we use this generative model for regularization term $R(v)$ and how this changes our optimization problem.

2.2 Manifold Learning

To regularize the domain space of solutions, we leverage a style-based generative model to learn deep priors of anatomic structures. Our model relies on StyleGAN2 [15] that we extend in 3D by changing the 2D convolutions into 3D ones as done in 3DStyleGAN [10] except that we start from the StyleGAN2 architecture.

Our generator G generates a volume v given a latent vector $\mathbf{w}$ and Gaussian noise vectors $\mathbf{n} = \{\mathbf{n}_j\}_j$: $v = G(\mathbf{w}, \mathbf{n})$. Latent vector $\mathbf{w} \in \mathcal{N}(\mathbf{w}|\mu, \sigma)$ is computed

from an initial latent vector $\mathbf{z} \in \mathcal{N}(0, I)$ mapped using a learned network m: $\mathbf{w} = m(\mathbf{z})$. $\mathbf{w}$ controls the global structure of the predicted volumes at different scales by its components $\mathbf{w}_i$, while the noise vectors $\mathbf{n}$ allow more fine-grained details. The mean μ and standard deviation σ of the mapped latent space can be computed by mapping over initial latent space $\mathcal{N}(0, I)$ after training. The mapping network learns to disentangle the initial latent space relatively to semantic features which is crucial for the inverse problem. We train this model using the non-saturating logistic loss [5] and path length regularization [15]. For the discriminator, we use the non-saturating logistic loss with R1 regularization [19]. We implement adaptive discriminator augmentation from StyleGAN-ADA [14] to improve learning of the model's manifold with limited medical imaging data.

2.3 Reconstruction from Biplanar Projections

Since our generative model provides a volume v as a function of vectors $\mathbf{w}$ and $\mathbf{n}$, we can reparameterize our optimization from Eq. (1) into:

$$\mathbf{w}^*, \mathbf{n}^* = \underset{\mathbf{w},\mathbf{n}}{\operatorname{argmin}} \sum_i \mathcal{L}(G(\mathbf{w}, \mathbf{n}), I_i) + R(\mathbf{w}, \mathbf{n}) \,. \tag{3}$$

Note that by contrast with [18] for example, we optimize on the noise vectors $\mathbf{n}$ as well: as we discovered in our early experiments, the $\mathbf{n}$ are also useful to embed high-resolution details. We take our regularization term $R(\mathbf{w}, \mathbf{n})$ as:

$$R(\mathbf{w}, \mathbf{n}) = \lambda_w \mathcal{L}_w(\mathbf{w}) + \lambda_c \mathcal{L}_c(\mathbf{w}) + \lambda_n \mathcal{L}_n(\mathbf{n}) \,. \tag{4}$$

Term $\mathcal{L}_w(\mathbf{w}) = -\sum_k \log \mathcal{N}(\mathbf{w}_k | \mu, \sigma)$ ensures that $\mathbf{w}$ lies on the same distribution as during training. $\mathcal{N}(\cdot | \mu, \sigma)$ represents the density of the standard normal distribution of mean μ and standard deviation σ.

Term $\mathcal{L}_c(\mathbf{w}) = -\sum_{i,j} \log \mathcal{M}(\theta_{i,j} | 0, \kappa)$ encourages the $\mathbf{w}_i$ vectors to be collinear so to keep the generation of coarse-to-fine structures coherent. $\mathcal{M}(\cdot; \mu, \kappa)$ is the density of the Von Mises distribution of mean μ and scale κ, which we take fixed, and $\theta_{i,j} = \arccos(\frac{\mathbf{w}_i \cdot \mathbf{w}_j}{\|\mathbf{w}_i\| \|\mathbf{w}_j\|})$ is the angle between vectors $\mathbf{w}_i$ and $\mathbf{w}_j$.

Term $\mathcal{L}_n(\mathbf{n}) = -\sum_j \log \mathcal{N}(\mathbf{n}_j | \mathbf{0}, I)$ ensures that the $\mathbf{n}_j$ lie on the same distribution as during training, *i.e.*, a multivariate standard normal distribution. The λ_* are fixed weights.

Projection Operator. In practice, we take operator A as a 3D cone beam projection that simulates X-ray attenuation across the patient, adapted from [21,27]. We model a realistic X-ray attenuation as a ray tracing projection using material and spectrum awareness:

$$\mathcal{I}_{\text{atten}} = \sum_E \mathcal{I}_0 e^{-\sum_m \mu(m,E) t_m} \,, \tag{5}$$

with $\mu(m, E)$ the linear attenuation coefficient of material m at energy state E that is known [11], t_m the material thickness, $\mathcal{I}_0$ the intensity of the source X-ray.

For materials, we consider the bones and tissues that we separate by threshold on electron density. A inverts the attenuation intensities $\mathcal{I}_{\text{atten}}$ to generate an X-ray along few directions successively. We make A differentiable using [21] to allow end-to-end optimization for reconstruction.

3 Experiments and Results

3.1 Dataset and Preprocessing

Manifold Learning. We trained our model with a large dataset of 3500 CTs of patients with head-and-neck cancer, more exactly 2297 patients from the publicly available The Cancer Imaging Archive (TCIA) [1,6,16,17,28,32] and 1203 from private internal data, after obtention of ethical approbations. We split this data into 3000 cases for training, 250 for validation, and 250 for testing. We focused CT scans on the head and neck region above shoulders, with a resolution of $80 \times 96 \times 112$, and centered on the mouth after automatic segmentation using a pre-trained U-Net [22]. The CTs were preprocessed by min-max normalization after clipping between -1024 and 2000 Hounsfield Units (HU).

3D Reconstruction. To evaluate our approach, we used an external private cohort of 80 patients who had undergone radiotherapy for head-and-neck cancer, with their consent. Planning CT scans were obtained for dose preparation, and CBCT scans were obtained at each treatment fraction for positioning with full gantry acquisition. As can be seen in Fig. 3 and the supplementary material, all these cases are challenging as there are large changes between the original CT scan and the CBCT scans. We identified these cases automatically by comparing the CBCTs with the planning CTs. To compare our reconstruction in the calibrated HU space, we registered the planning CTs on the CBCTs by deformable registration with MRF minimization [4]. We hence obtained 3D volumes as virtual CTs we considered as ground truths for our reconstructions after normalization. From these volumes, we generated projections using the projection module described in Sect. 2.3.

3.2 Implementation Details

Manifold Learning. We used Pytorch to implement our model, based on StyleGAN2 [15]. It has a starting base layer of $256 \times 5 \times 6 \times 7$ and includes four upsamplings with 3D convolutions and filter maps of $256, 128, 64, 32$. We also used 8 fully-convolutional layers with dimension 512 and an input latent vector of dimension 512, with tanh function as output activation. To optimize our model, we used lazy regularization [15] and style mixing [15], and added a 0.2 probability for generating images without Gaussian noise to focus on embedding the most information. We augmented the discriminator with vertical and depth-oriented flips, rotation, scaling, motion blur and Gaussian noise at a probability of 0.2. Our training used mixed precision on a single GPU Nvidia Geforce GTX

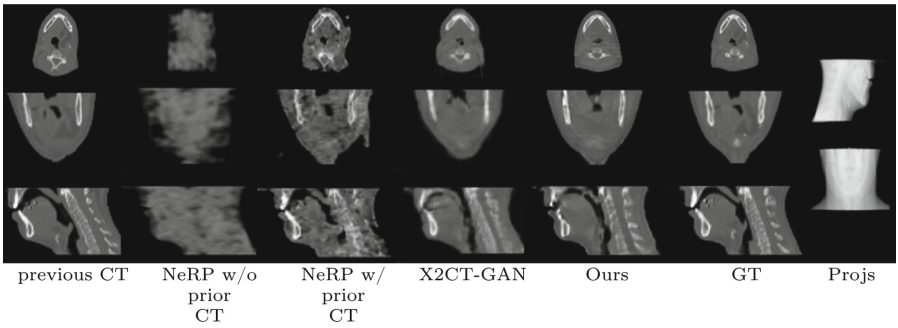

Fig. 3. Visual comparison of 3D reconstruction from biplanar projections by our model and baselines. Without a previous CT volume, NeRP fails by lack of constraints. When initialized with an earlier CT (left), NeRP tends to create artefacts to match the projections rather than really change the anatomy. Our method produces better matching structures than X2CT-GAN, almost matching the CT volume deformed on the CBCT volume (GT, right).

3090 with a batch size of 6, and we optimized the generator, discriminator, and mapping networks using Adam at learning rates 6e−5 and 1e−5 to avoid mode collapse and unstable training. After training for 4 weeks, we achieved stabilization of the Fréchet Inception Distance (FID) [9] and Multi-scale Structural Similarity (MS-SSIM) [29] on the validation set.

3D Reconstruction. For the reconstruction, we performed the optimization on GPU V100 PCI-E using Adam, with learning rate of 1e−3. By grid search on the validation set, we selected the best weights that well balance between structure and fine-grained details, $\lambda_2 = 10$, $\lambda_p = 0.1$, $\lambda_w = 0.1$, $\lambda_c = 0.05$, $\lambda_n = 10$. We perform 100 optimization steps starting from the mean of the mapped latent space, which takes 25 s, enabling clinical use.

3.3 Results and Discussion

Manifold Learning. We tested our model's ability to learn the low-dimensional manifold. We used FID [9] to measure the distance between the distribution of generated volumes and real volumes, and MS-SSIM [29] to evaluate volumes' diversity and quality. We obtained a 3D FID of 46 and a MS-SSIM of 0.92. For reference, compared to 3DStyleGAN [10], our model achieved half their FID score on another brain MRI dataset, with comparable MS-SSIM. This may be due to a more complex architecture, discriminator augmentation, or simpler anatomy.

Baselines. We compared our method against the main feed-forward method X2CT-GAN [30] and the neural radiance fields with prior image embedding method NeRP [23] meant for modest sparsely-sampled reconstruction. Recent methods like [24] and [12] were excluded because they provide only minor

Table 1. Metrics for our method and baselines, for reconstruction from 1 to 8 cone beam projections. Standard deviations are provided in parentheses.

Method	1 Projection		2 Projections	
	PSNR (dB)↑	SSIM↑	PSNR (dB)↑	SSIM↑
NeRP (w/o prior volume)	14.8 (±2.7)	0.12 (±0.10)	18.4 (±3.8)	0.17 (±0.10)
NeRP (w/ prior volume)	22.5 (±3.2)	0.29 (±0.07)	23.5 (±3.5)	0.30 (±0.06)
X2CT-GAN	20.7 (±2.4)	0.57 (±0.07)	21.8 (±2.5)	0.72 (±0.08)
Ours	**23.2** (±**2.8**)	**0.79** (±**0.09**)	**25.8** (±**3.2**)	**0.85** (±**0.10**)
	4 Projections		8 Projections	
NeRP (w/o prior volume)	19.9 (±2.6)	0.21 (±0.04)	20.0 (±2.5)	0.23 (±0.05)
NeRP (w/ prior volume)	24.2 (±2.7)	0.32 (±0.05)	24.9 (±4.9)	0.34 (±0.08)
Ours	**28.2** (±**3.5**)	**0.89** (±**0.10**)	**30.1** (±**3.9**)	**0.92** (±**0.11**)

improvements compared to X2CT-GAN [30] and have similar constraints to feed-forward methods. Additionally, no public implementation is available. [26] uses a flow-based generative model, but the results are of lower quality compared to GANs and similar to X2CT-GAN [30].

3D Reconstruction. To evaluate our method's performance with biplanar projections, we focused on positioning imaging for radiotherapy. Figure 3 compares our reconstruction with those of the baselines from biplanar projections. Our method achieves better fitting of the patient structure, including bones, tissues, and air separations, almost matching the real CT volume. X2CT-GAN [30] produced realistic structures, but failed to match the actual structures as it does not enforce consistency with the projections. In some clinical procedures, an earlier CT volume of the patient may be available and can be used as an additional input for NeRP [23]. Without a previous CT volume, NeRP lacks the necessary prior to accurately solve the ill-posed problem. Even when initialised with a previous CT volume, NeRP often fails to converge to the correct volume and introduces many artifacts when few projections are used. In contrast, our method is more versatile and produces better results.

We used quantitative metrics (PSNR and SSIM) to evaluate reconstruction error and human perception, respectively. Table 1 shows these metrics for our method and baselines with 1 to 8 cone beam projections. Deviation from projections, as in X2CT-GAN, leads to inaccurate reconstruction. However, relying solely on projection consistency is inadequate for this ill-posed problem. NeRP matches projections but cannot reconstruct the volume correctly. Our approach balances between instant and iterative methods by providing a reconstruction in 25 s with 100 optimization steps, while ensuring maximal consistency. In contrast, NeRP requires 7 min, and X2CT-GAN produces structures instantly but unmatching. Clinical CBCT acquisition and reconstruction by FDK [3] take about 1–2 min and 10 s respectively. Our approach significantly reduces clin-

ical time and radiation dose by using instant biplanar projections, making it promising for fast 3D visualization towards complex positioning.

4 Conclusion, Limitations, and Future Work

We proposed a new unsupervised method for 3D reconstruction from biplanar X-rays using a deep generative model to learn the structure manifold and retrieve the maximum a posteriori volume with the projections, leading to state-of-the-art reconstruction. Our approach is fast, robust, and applicable to various human body parts, making it suitable for many clinical applications, including positioning and visualization with reduced radiation.

Future hardware improvements may increase resolution, and our approach could benefit from other generative models like latent diffusion models. This approach may provide coarse reconstructions for patients with rare abnormalities, as most learning methods, but a larger dataset or developing a prior including tissue abnormalities could improve robustness.

References

1. Beichel, R.R., et al.: Data from QIN-HEADNECK (2015)
2. Ellis, S., et al.: Evaluation of 3D GANs for Lung Tissue Modelling in Pulmonary CT. arXiv (2022)
3. Feldkamp, L.A., Davis, L.C., Kress, J.W.: Practical cone-beam algorithm. J. Opt. Soc. Am. A-Opt. Image Sci. Vis. (1984)
4. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Med. Image Anal. **12**(6), 731–741 (2008)
5. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11) (2020)
6. Grossberg, A., et al.: Anderson Cancer Center Head and Neck Quantitative Imaging Working Group. HNSCC (2020)
7. Henzler, P., Rasche, V., Ropinski, T., Ritschel, T.: Single-image tomography: 3D volumes from 2D cranial X-rays. In: Computer Graphics Forum (2018)
8. Herman, G.T.: Fundamentals of Computerized Tomography: Image Reconstruction from Projections. Springer, London (2009). https://doi.org/10.1007/978-1-84628-723-7
9. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: GANs trained by a two time-scale update rule converge to a local nash equilibrium. In: NeurIPS (2017)
10. Hong, S., et al.: 3D-StyleGAN: a style-based generative adversarial network for generative modeling of three-dimensional medical images. In: Engelhardt, S., et al. (eds.) DGM4MICCAI/DALI -2021. LNCS, vol. 13003, pp. 24–34. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-88210-5_3
11. Hubbell, J.H.: Tables of X-Ray Mass Attenuation Coefficients 1 keV to 20 MeV for Elements Z=1 to 92 and 48 Additional Substance of Dosimetric Interest. NISTIR 5632 (1995)

12. Jiang, Y.: MFCT-GAN: multi-information network to reconstruct CT volumes for security screening. J. Intell. Manuf. Spec. Equip. **3**(1), 17–30 (2022)
13. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 694–711. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_43
14. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data. In: NeurIPS (2020)
15. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of stylegan. In: CVPR (2020)
16. Kinahan, P., Muzi, M., Bialecki, B., Coombs, L.: Data from the ACRIN 6685 Trial HNSCC-FDG-PET/CT (2020)
17. Kwan, J.Y.Y., et al.: Data from Radiomic Biomarkers to Refine Risk Models for Distant Metastasis in Oropharyngeal Carcinoma (2019)
18. Marinescu, R.V., Moyer, D., Golland, P.: Bayesian Image Reconstruction Using Deep Generative Models. arXiv (2020)
19. Mescheder, L., Geiger, A., Nowozin, S.: Which training methods for GANs do actually converge? In: ICML (2018)
20. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: representing scenes as neural radiance fields for view synthesis. Commun. ACM **65**(1) (2021)
21. Peng, C., Liao, H., Wong, G., Luo, J., Zhou, S.K., Chellappa, R.: XraySyn: realistic view synthesis from a single radiograph through CT priors. In: AAAI (2021)
22. 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
23. Shen, L., Pauly, J., Xing, L.: NeRP: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. IEEE Trans. Neural Netw. (2022)
24. Shen, L., Zhao, W., Capaldi, D., Pauly, J., Xing, L.: A geometry-informed deep learning framework for ultra-sparse 3D tomographic image reconstruction. Comput. Biol. Med. **148**, 105710 (2022)
25. Shen, L., Zhao, W., Xing, L.: Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. Nature **3**(11), 880–888 (2019)
26. Shibata, H., et al.: On the simulation of ultra-sparse-view and ultra-low-dose computed tomography with maximum a posteriori reconstruction using a progressive flow-based deep generative model. Tomography **8**(5), 2129–2152 (2022)
27. Unberath, M., et al.: DeepDRR – a catalyst for machine learning in fluoroscopy-guided procedures. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11073, pp. 98–106. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00937-3_12
28. Vallières, M., et al.: Data from Head-Neck-PET-CT (2020)
29. Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers (2003)
30. Ying, X., Guo, H., Ma, K., Wu, J., Weng, Z., Zheng, Y.: X2CT-GAN: reconstructing CT from biplanar X-rays with generative adversarial networks. In: CVPR (2019)

31. Zha, R., Zhang, Y., Li, H.: NAF: neural attenuation fields for sparse-view CBCT reconstruction. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13436, pp. 442–452. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_42
32. Zuley, M.L., et al.: The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma Collection TCGA-HNSC) (2015)