C^{2}RV: Cross-Regional and Cross-View Learning for Sparse-View CBCT Reconstruction — Supplementary Material

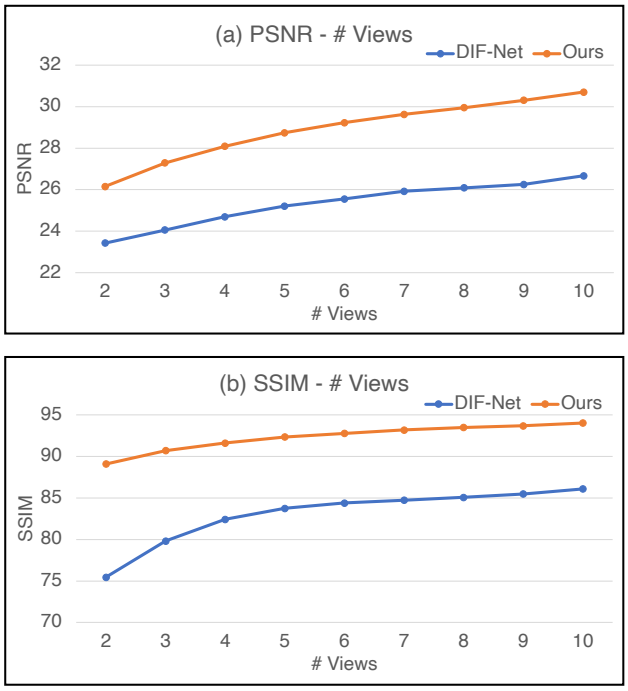

Figure 8. Reconstruction performance with different numbers of projection views. We compare our proposed C^{2}RV and previous state-of-the-art DIF-Net [6] on N-view reconstruction of chest CT with the resolution of 256^3. PSNR (dB) and SSIM ($\times 10^{-2}$) are evaluated and reported in (a) and (b), respectively.

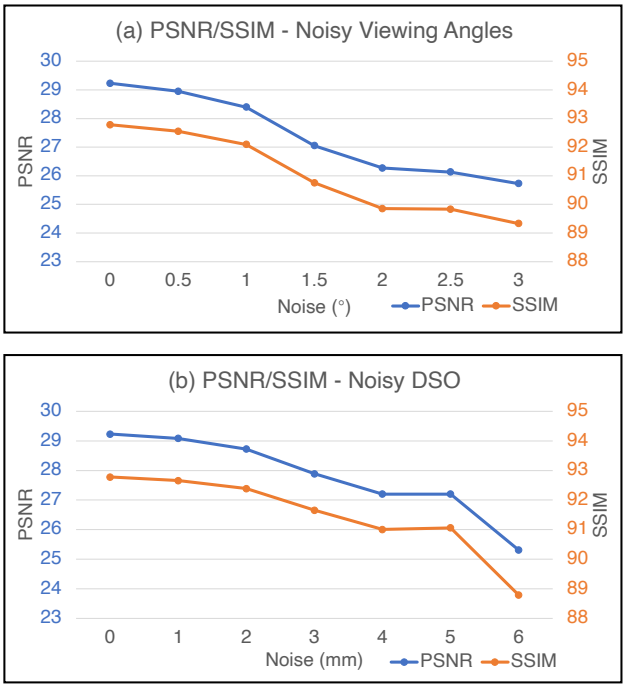

Figure 9. Reconstruction performance with noisy scanning parameters, including viewing angles and the distance of source to origin (DSO) in (a) and (b), respectively. Experiments are conducted on 6-view chest CT reconstruction with the resolution of 256^3. PSNR (dB) and SSIM ($\times 10^{-2}$) are reported.

A. Additional Experiments

The Number of Views. In Figure 8, we compare the proposed C^{2}RV with previous state-of-the-art DIF-Net [6] on sparse-view chest CT reconstruction from different numbers of views (*i.e.*, 2-10). With only 2 views as the input, our C^{2}RV can perform better than DIF-Net with 8 views in terms of PSNR and SSIM.

Noisy Scanning Parameters. In Figure 9, we evaluate the reconstruction performance of C^{2}RV with different noisy scanning parameters. A detailed description of the experimental setup is given in Sec. 5.2. Our C^{2}RV can be robust to slight shifts in scanning parameters. However, larger shifts will cause performance collapse as the back-projection (Eqn. 3) and pixel-aligned querying (Eqn. 1) are highly dependent on the precise measurement process.

Comparison with The Resolution of 128^3. In Table 6, we compare different reconstruction methods with the reconstruction resolution of 128^3. To evaluate the performance, we resample ground-truth CT with double voxel spacing.

Limited-Angle Reconstruction. In Table 7, we compare our C^{2}RV and DIF-Net [6] on limited-angle reconstruction, where view angles are selected in the range of 144°, 108°, and 72°. 6-view and 8-view reconstruction experiments are included to demonstrate the consistently superior performance of our proposed C^{2}RV.

Results on Dental CBCT. In Table 8, we additionally conduct experiments on a dental CBCT dataset, ToothFairy [1], consisting of 343 training, 25 validation, and 75 testing cases. Visualization examples are shown in Figure 10.

Table 6. Comparison of different methods on two CT datasets (*i.e.*, chest and knee) with various numbers of projection views. The resolution of the reconstructed CT is 128^3. The reconstruction results are evaluated with PSNR (dB) and SSIM ($\times 10^{-2}$), where higher PSNR/SSIM indicate better performance. The best values are **bolded** and the second-best values are underlined.

Method	Type	LUNA16 [8] (Chest CT)						Lin *et al.* [6] (Knee CBCT)					
		6-View		8-View		10-View		6-View		8-View		10-View	
FDK [3]	Self-Supervised	16.35	37.36	17.74	40.39	18.61	42.86	19.57	38.53	21.27	42.58	22.68	46.76
SART [2]		20.45	65.87	20.76	69.10	20.91	71.56	25.43	80.75	26.56	85.24	27.55	88.42
NAF [11]		19.13	50.95	21.12	61.23	23.02	68.06	20.82	53.58	23.01	63.86	24.98	73.12
NeRP [9]		24.31	75.80	26.66	83.08	26.93	83.66	25.10	72.28	26.57	77.52	27.23	81.92
FBPConvNet [4]	Data-Driven: Denoising	26.75	80.48	27.45	81.33	28.21	83.27	26.41	84.38	27.01	86.60	27.82	88.48
FreeSeed [7]		27.31	82.25	27.85	82.85	28.29	83.74	27.92	87.21	28.73	88.89	30.03	90.36
BBDM [5]		26.16	81.39	26.93	81.74	27.55	83.01	27.86	87.05	28.80	88.75	30.11	90.43
PixelNeRF [10]	Data-Driven: INR-based	25.91	82.30	26.40	84.29	26.84	85.87	27.22	89.67	28.13	91.19	28.79	92.28
DIF-Net [6]		26.93	87.50	27.55	88.29	28.23	89.66	28.27	91.07	29.60	92.23	30.72	94.17
C^2RV (*ours*)		**30.59**	**94.78**	**31.35**	**95.50**	**32.03**	**95.92**	**30.91**	**95.21**	**31.89**	**95.99**	**32.73**	**96.58**

Table 7. Limited-angle reconstruction with different numbers of projection views on chest CT. PSNR (dB) and SSIM ($\times 10^{-2}$) are reported, and the reconstruction resolution is 256^3.

Method	180°		144°		108°		72°	
6-View								
DIF-Net [6]	25.6	84.4	25.5	84.3	25.3	83.0	24.0	78.7
C^2RV (*ours*)	**29.2**	**92.8**	**29.2**	**92.8**	**28.6**	**92.0**	**27.5**	**90.8**
8-View								
DIF-Net [6]	26.1	85.1	26.0	85.0	25.6	84.0	24.2	79.8
C^2RV (*ours*)	**30.0**	**93.5**	**29.9**	**93.4**	**28.9**	**92.5**	**27.7**	**91.0**

Table 8. Experiments on a dental CBCT dataset (ToothFairy [1]). PSNR/SSIM (dB/$\times 10^{-2}$) are evaluated, and the reconstruction resolution is 256^3. The best values are **bolded**.

Method	6-View		8-View		10-View	
FDK [3]	14.20	28.35	14.77	29.66	15.15	31.75
NeRP [9]	21.77	72.06	24.18	78.83	25.99	82.08
FreeSeed [7]	26.35	78.98	27.08	81.38	27.63	84.40
BBDM [5]	26.29	78.57	27.28	80.33	28.00	83.96
DIF-Net [6]	25.78	83.62	26.29	84.81	26.90	86.42
C^2RV (*ours*)	**28.85**	**92.81**	**29.30**	**93.03**	**30.04**	**93.38**

Additional Visualization. In Figure 10, we visualize more examples reconstructed (6-view) by different methods on 3 datasets: LUNA16 [8], Lin *et al.* [6], and ToothFairy [1].

B. Processing Efficiency

Model Parameters and Reconstruction Time. In Table 9, we compare the number of network parameters and reconstruction speed with different numbers of projection views. Although C^2RV requires a longer time for reconstruction than DIF-Net [6], the reconstruction quality of C^2RV is much better, and the reconstruction time is lower than half a minute, which is acceptable in clinical practice.

Table 9. The number of model parameters (Param.) and reconstruction time of different methods are compared on chest CT. The reconstruction resolution is 256^3. Reconstruction (Recon.) performance (PSNR/SSIM: dB/$\times 10^{-2}$) is reported for better comparison.

Method	Param. (M)	6-View		8-View	
		Recon.	Time (s)	Recon.	Time (s)
FDK [3]	-	14.4/30.5	0.3	15.3/31.9	0.3
NeRP [9]	0.7	23.6/74.5	937.5	25.8/80.7	1361.1
FBPConvNet [4]	34.6	24.4/77.6	1.7	24.9/78.9	1.7
FreeSeed [7]	8.7	25.6/77.4	3.7	26.9/78.9	3.7
PixelNeRF [10]	24.7	24.7/78.7	40.4	25.0/80.6	57.6
DIF-Net [6]	31.1	25.6/84.4	1.1	26.1/85.1	1.4
C^2RV (*ours*)	50.8	29.2/92.8	23.8	30.0/93.5	31.3

Table 10. Reconstruction time (s) *vs*. reconstruction resolution of our proposed C^2RV on chest CT dataset.

Resolution	256^3	192^3	128^3	96^3	64^3
6-view	23.8	9.3	3.0	0.8	<0.1
8-view	31.3	12.8	3.9	1.1	<0.1
10-view	39.3	15.7	4.9	1.5	<0.1

Speed *vs*. Resolution. In Table 10, we report the reconstruction speed of our proposed C^2RV under different experimental settings, including the number of views and reconstruction resolution.

C. Data Preprocessing

We follow [6] to generate projections by DRRs using TIGRE[1]. For knee CT, the projection configuration is the same as in [6]. For chest CT, we set the distance of source to detector (DSD) and the distance of source to origin (DSO) as 1200 mm and 800 mm, respectively. Configuration files will be made public along with the code.

[1] https://github.com/CERN/TIGRE

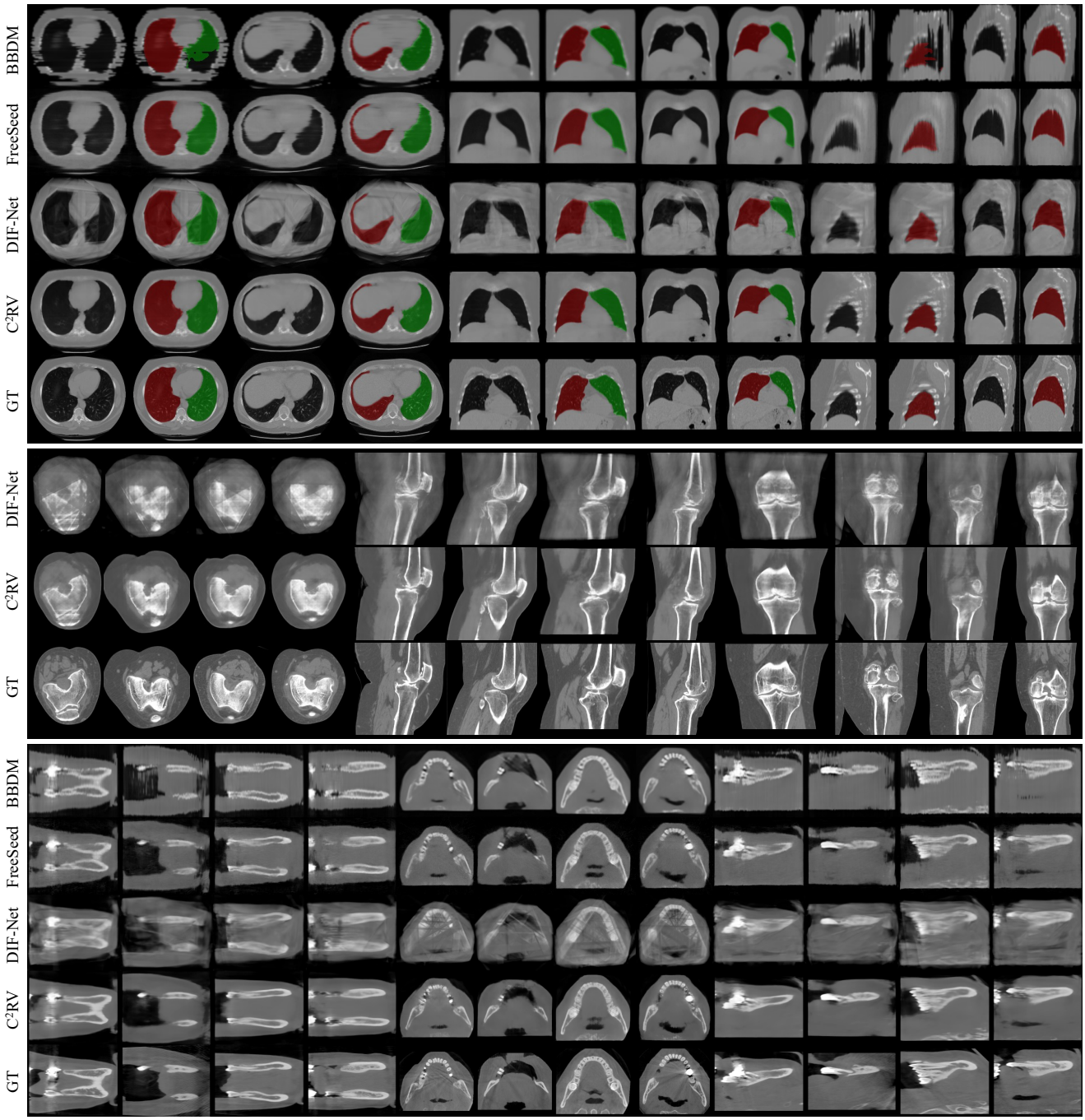

Figure 10. Visualization of 6-view reconstruction on three different datasets. From top to bottom: LUNA16 [8] (chest CT), Lin *et al*. [6] (knee CBCT), and ToothFairy [1] (dental CBCT).

References

[1] Deep segmentation of the mandibular canal: a new 3d annotated dataset of cbct volumes. *IEEE Access*, 2022. 1, 2, 3

[2] Anders H Andersen and Avinash C Kak. Simultaneous algebraic reconstruction technique (sart): a superior implementation of the art algorithm. *Ultrasonic imaging*, 6(1):81–94, 1984. 2

[3] Lee A Feldkamp, Lloyd C Davis, and James W Kress. Practical cone-beam algorithm. *Josa a*, 1(6):612–619, 1984. 2

[4] Kyong Hwan Jin, Michael T McCann, Emmanuel Froustey, and Michael Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, 2017. 2

[5] Bo Li, Kaitao Xue, Bin Liu, and Yu-Kun Lai. Bbdm: Image-

to-image translation with brownian bridge diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1952–1961, 2023. 2

[6] Yiqun Lin, Zhongjin Luo, Wei Zhao, and Xiaomeng Li. Learning deep intensity field for extremely sparse-view cbct reconstruction. *arXiv preprint arXiv:2303.06681*, 2023. 1, 2, 3

[7] Chenglong Ma, Zilong Li, Junping Zhang, Yi Zhang, and Hongming Shan. Freeseed: Frequency-band-aware and self-guided network for sparse-view ct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 250–259. Springer, 2023. 2

[8] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas De Bel, Moira SN Berens, Cas Van Den Bogaard, Piergiorgio Cerello, Hao Chen, Qi Dou, Maria Evelina Fantacci, Bram Geurts, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images: the luna16 challenge. *Medical image analysis*, 42:1–13, 2017. 2, 3

[9] Liyue Shen, John Pauly, and Lei Xing. Nerp: implicit neural representation learning with prior embedding for sparsely sampled image reconstruction. *IEEE Transactions on Neural Networks and Learning Systems*, 2022. 2

[10] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4578–4587, 2021. 2

[11] Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: Neural attenuation fields for sparse-view cbct reconstruction. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part VI*, pages 442–452. Springer, 2022. 2