

- Supplemental Material -

Seeing through a Black Box: Toward High-Quality Terahertz Imaging via Subspace-and-Attention Guided Restoration

Wen-Tai Su¹, Yi-Chun Hung², Po-Jen Yu¹, Shang-Hua Yang¹[0000-0002-5528-9281], and Chia-Wen Lin¹[0000-0002-9097-2318]

¹ Dept. EE, National Tsing Hua University, Hsinchu 300044, Taiwan

² Dept. ECE, University of California, Los Angeles, USA

This supplementary material provides those details that could not be accommodated in the main manuscript due to the limited space. Sec. 1 provides the detailed settings for our in-house THz system. Sec. 3.1 provides the additional quantitative evaluation results. Sec. 3.2 provides the evaluation results testing on the real-world objects. Finally, Sec. 3.3 demonstrates some sample views of 3D tomography reconstruction from 30 views restored by our model as part of the supplementary material.

1 ASOPS THz System

Our in-house THz measurement system is an asynchronous optical sampling THz time-domain spectroscopy system (ASOPS THz-TDS), which is composed of two asynchronous femtosecond lasers whose central wavelength are located at 1550 nm with tens of mW level, a pair of THz photoconductive antenna (THz PCA) source and detector, a linear and rotation motorized stage, four plane-convex THz lens with 50 mm focal length, a transimpedance amplifier (TIA), and a unit of data acquisition (DAQ) and processing [2]. The repetition rates of the two asynchronous femtosecond lasers are 100 MHz and 100 MHz + 200 Hz, respectively. The sampling rate of DAQ is 20 MHz. With the configuration above, our ASOPS THz-TDS system delivers 0.1 ps temporal resolution and the THz frequency bandwidth of 5 THz. Additionally, our ASOPS THz-TDS system provides THz pulse signals with 41.7 dB dynamic range from 0.3 THz to 3 THz and 516 femtoseconds at full width at half maximum (FWHM). However, under the configuration above, the number of sampling points for a trace is approximately 100 K, consuming extremely large transmission bandwidth. To address this limitation, only the 100-ps segment of the THz pulse signal is extracted. With the extracted segment of 100 ps, the frequency resolution is 10 GHz.

2 About Channel Attention Module

To fuse multi-scale features from different spectral bands in the channel dimension, we incorporate the efficient channel attention mechanism proposed in [4] in the decoder path of SARNet. The block diagram of CAM as shown in Fig. S1.

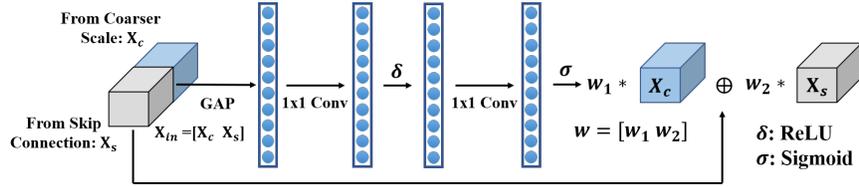


Fig. S1: Block diagram of Channel Attention Module (CAM).

3 Supplementary Results

3.1 Additional Image Restoration Experiments

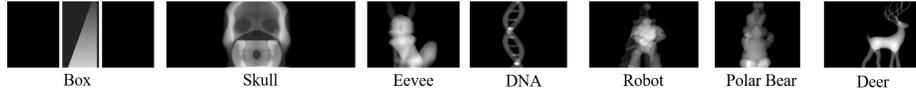
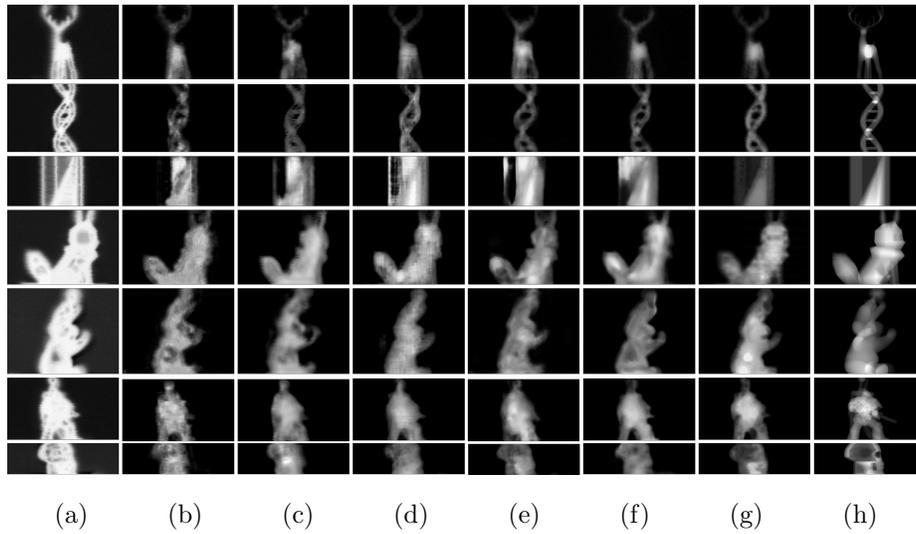


Fig. S2: Illustration of authentic measured THz data for the seven 3D-printed HIPS objects in our experiments.

Fig. S3: Qualitative comparison of THz image restoration results for seven objects from left to right: (a) Time-max, (b) DnCNN-S [6], (c) RED [3], (d) NBNNet [1], (e) U-Net_{base} [5], (f) U-Net_{MS}, (g) SARNNet, and (h) the ground-truth.

The seven sample objects are illustrated in Fig. S2, consisting of 60 projections per object and 420 2D THz images in total. Fig. S3 illustrates additional restored views of 7 sample objects. Both the quantitative and qualitative evaluations confirm a significant performance leap with **SARNet** over the competing methods. The raw image files of the above 2D restored images for the seven sample objects are zipped into a file (the file folder is 2D_restored) in the complementary material.

Table. T1 shows the model complexities of the different methods. When compared to the state-of-the-art method **NBNet** [1], and two U-Net-based baselines **U-Net_{base}** and **U-Net_{MS}**, our **SARNet** requires much fewer numbers of parameters and GFLOPs. The run-time with **SARNet** is also less than **NBNet**, but more than **U-Net_{base}** and **U-Net_{MS}**, though.

All the above comparisons demonstrate that **SARNet** not only achieves superior performance improvements, but also is computation and memory efficient.

3.2 Test on real-world objects

To verify the generality of our **SARNet** to real objects, we train it on 3D printed objects and then test the model on a real object containing 3 materials (including paper, polymer and metal) in Fig. S4. The result shows that our model can well restore real objects even if the training data do not contain these materials. It supports our aim that **SARNet** can effectively utilize the physics characteristics of both targeted objects and environments, which can easily adapt to real-world data.

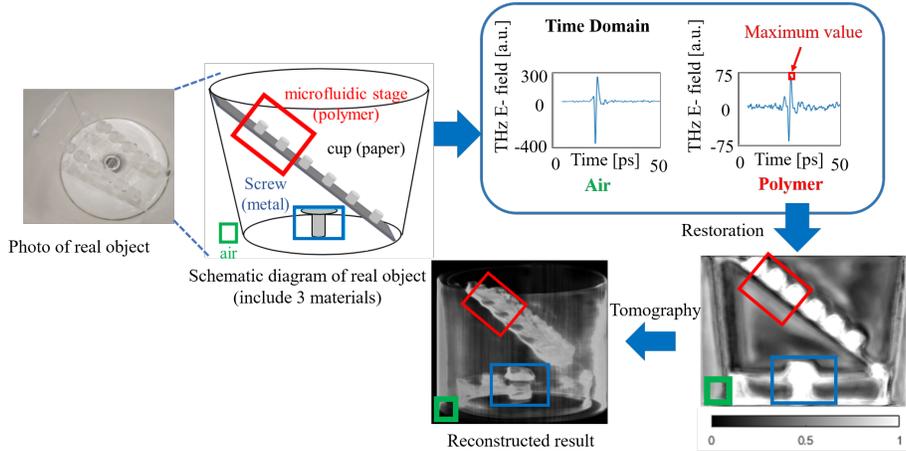


Fig. S4: Illustration of restoration result for real object.

3.3 3D Tomographic Reconstruction

Fig. S5 to Fig. S11 show additional snapshots of 3D tomographic reconstructions for **Deer**, **DNA**, **Box**, **Eevee**, **Bear**, **Robot**, and **Skull**, respectively.

To facilitate inspecting the results, the video files of the above 3D tomographic reconstruction experiments for the seven sample objects are zipped into a file (the file folder is tomography_result) in the complementary material. We also make the video files available at our project site³. Each video file compares the ground-truth and seven 3D tomographic reconstructions of each sample object reconstructed by seven different methods.

3.4 Limitations

SARNet uses multi-spectral THz data to extract geometric information. Depending on the selected THz frequency bands and their SNRs, the diffraction-limited system resolution can theoretically push down to 0.1mm. As water/metal are highly absorptive/reflective materials to the THz wave, our system is not applicable to the aqueous objects or objects hidden inside metallic packages.

Table T1: Comparison of the model complexity (the numbers of Parameters and GFLOPs, and run-time) with different methods. Run-time are measured with the Nvidia Titan 2080 Ti.

Method	Params (M)	GFLOPs	Run-time (ms)
DnCNN-S [6]	0.55	4.55	6
RED [3]	0.66	1.36	4
NBNet [1]	13.31	22.20	25
U-Net _{base} [5]	9.5	3.88	11
U-Net _{MS}	9.5	3.88	12
SARNet (Ours)	3.5	1.91	19

³ Please refer to https://github.com/wtnthu/THz_Tomography_eccv2022

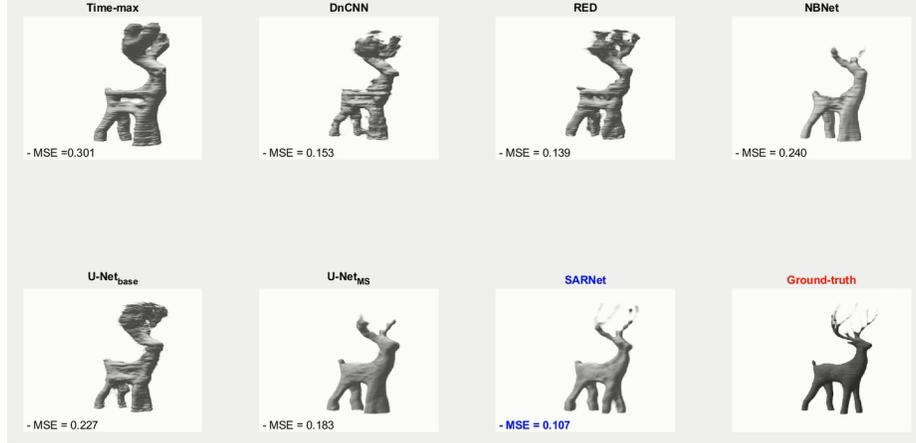


Fig. S5: Snapshots of 3D tomographic imaging results for **Deer** from left to right: (a) Time-max, (b) DnCNN-S [6], (c) RED [3], (d) NBNet [1], (e) U-Net_{base} [5], (f) U-Net_{MS}, (g) SARNet, and (h) the ground-truth.

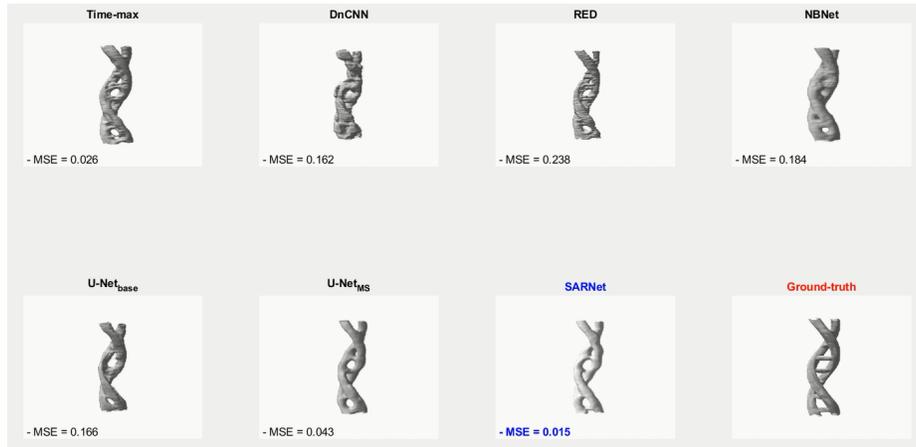


Fig. S6: Snapshots of 3D tomographic imaging results for **DNA** from left to right: (a) Time-max, (b) DnCNN-S [6], (c) RED [3], (d) NBNet [1], (e) U-Net_{base} [5], (f) U-Net_{MS}, (g) SARNet, and (h) the ground-truth.

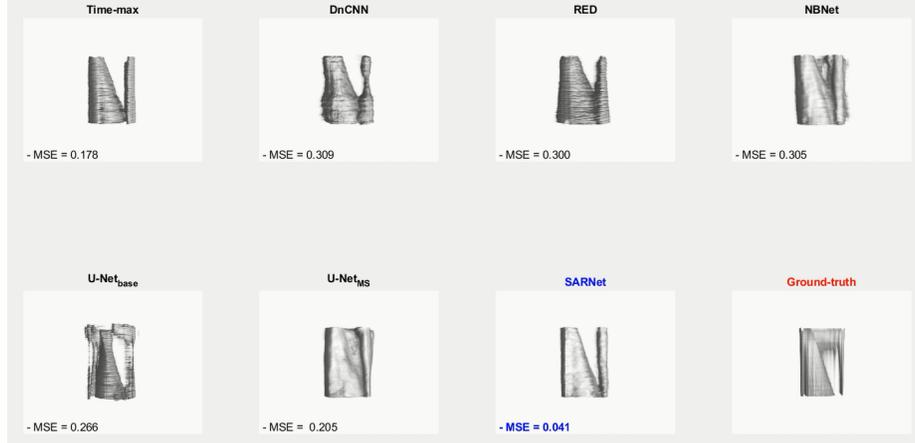


Fig. S7: Snapshots of 3D tomographic imaging results for **Box** from left to right: (a) Time-max, (b) DnCNN-S [6], (c) RED [3], (d) NBNet [1], (e) U-Net_{base} [5], (f) U-Net_{MS}, (g) SARNet, and (h) the ground-truth.

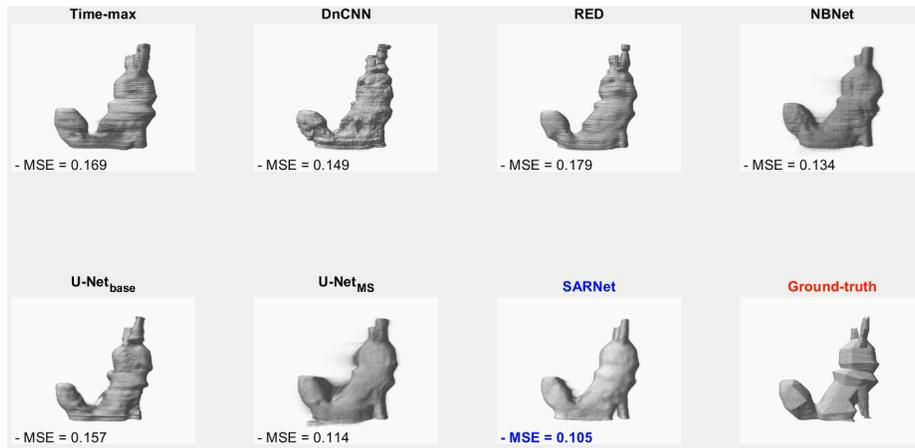


Fig. S8: Snapshots of 3D tomographic imaging results for **Eevee** from left to right: (a) Time-max, (b) DnCNN-S [6], (c) RED [3], (d) NBNet [1], (e) U-Net_{base} [5], (f) U-Net_{MS}, (g) SARNet, and (h) the ground-truth.

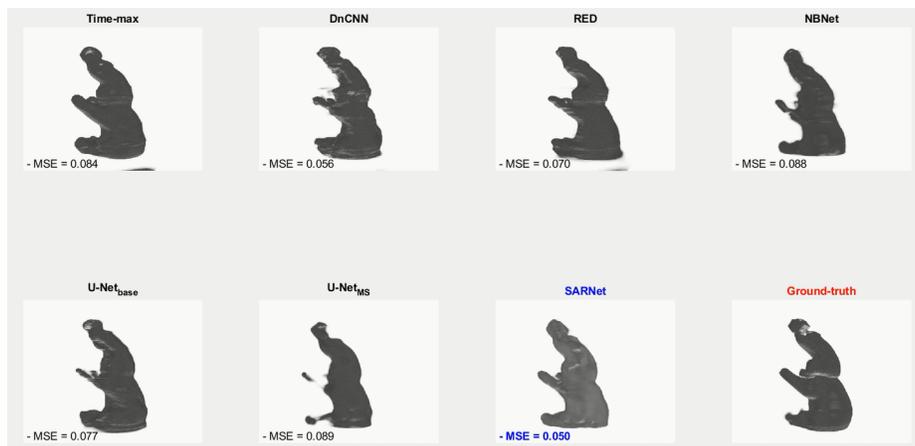


Fig. S9: Snapshots of 3D tomographic imaging results for **Bear** from left to right: (a) Time-max, (b) DnCNN-S [6], (c) RED [3], (d) NBNet [1], (e) U-Net_{base} [5], (f) U-Net_{MS}, (g) SARNet, and (h) the ground-truth.

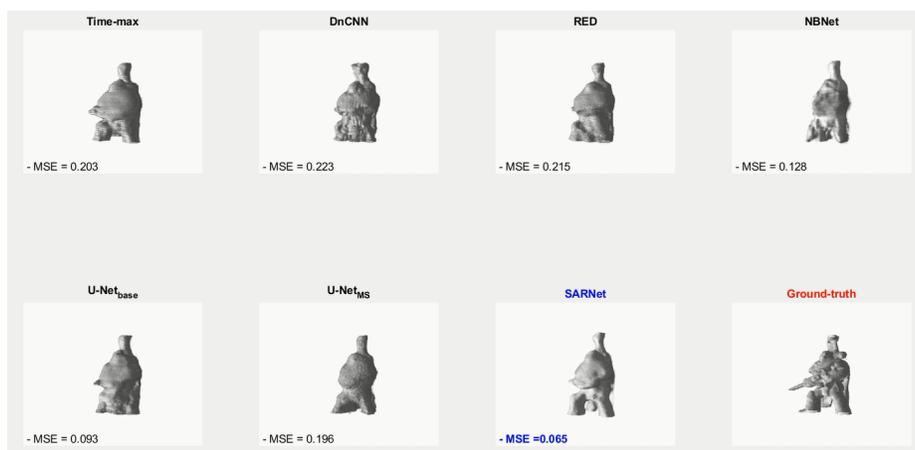


Fig. S10: Snapshots of 3D tomographic imaging results for **Robot** from left to right: (a) Time-max, (b) DnCNN-S [6], (c) RED [3], (d) NBNet [1], (e) U-Net_{base} [5], (f) U-Net_{MS}, (g) SARNet, and (h) the ground-truth.

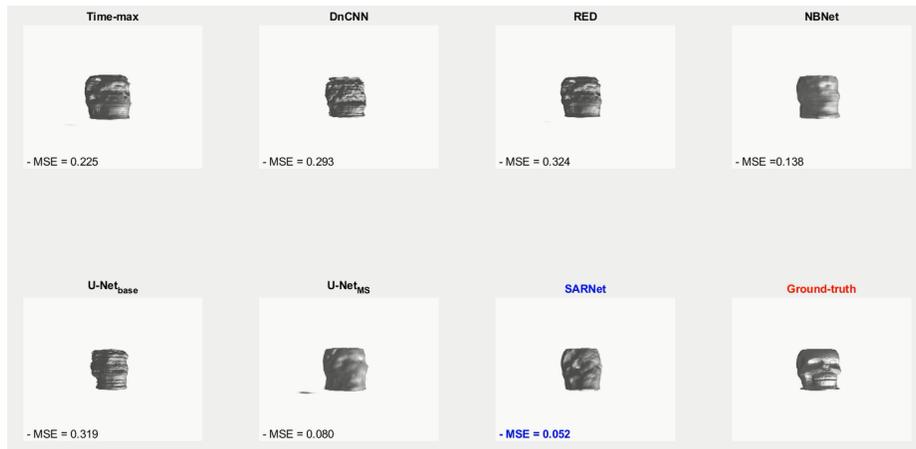


Fig. S11: Snapshots of 3D tomographic imaging results for **Skull** from left to right: (a) Time-max, (b) DnCNN-S [6], (c) RED [3], (d) NBNet [1], (e) U-Net_{base} [5], (f) U-Net_{MS}, (g) SARNet, and (h) the ground-truth.

References

1. Cheng, S., Wang, Y., Huang, H., Liu, D., Fan, H., Liu, S.: NNet: Noise basis learning for image denoising with subspace projection. In: Proc. IEEE/CVF Int. Conf. Comput. Vis. Pattern Recognit. pp. 4896–4906 (2021)
2. Janke, C., Först, M., Nagel, M., Kurz, H., Bartels, A.: Asynchronous optical sampling for high-speed characterization of integrated resonant terahertz sensors. *Optics Lett.* **30**(11), 1405–1407 (2005)
3. Mao, X., Shen, C., Yang, Y.B.: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. In: Proc. Adv. Neural Inf. Process. Syst. p. 2802–2810 (2016)
4. Qin, X., Wang, X., Bai, Y., Xie, X., Jia, H.: Ffa-net: Feature fusion attention network for single image dehazing. In: Proc. of the AAAI Conf. on Artificial Intelligence. vol. 34, pp. 11908–11915 (2020)
5. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proc. Int. Conf. Medical Image Comput. Comput.-Assisted Intervention. pp. 234–241 (2015)
6. Zhang, K., ana Y. Chen, W.Z., Meng, D., Zhang, L.: Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Trans. Image Process.* **26**(7), 3142–3155 (2017)