DINER: Disorder-Invariant Implicit Neural Representation: Supplementary Material

Shaowen Xie[1,†], Hao Zhu[1,†], Zhen Liu[1,†], Qi Zhang[2], You Zhou[1], Xun Cao[1], Zhan Ma[1]
[1] School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China
[2] AI Lab, Tencent Company, Shenzhen 518054, China
[†] Equal contribution. Corresponding author: caoxun@nju.edu.cn

1. 2D Image Fitting

1.1. Comparisons of Learned INRs

Fig. 1 shows comparisons of learned INRs in the DINER with MLP and SIREN backbones on more 2D images.

1.2. Comparisons with the SOTAs

Fig. 2 provides more qualitative results on the image fitting task. All 5 methods (DINER with MLP and SIREN backbones, and [3–5]) share the same network size as described in the manuscript.

2. Neural Representation for 3D video

Fig. 3 shows comparisons of the neural video representation task on the 'ShakeNDry' data.

3. Lensless Imaging Experiments

We provide the comparison on reconstructed amplitude and phase images of animal skin section [7] by DINER, DNF [8] and PhysenNet [6]. Fig. 4 provides qualitative comparisons. The proposed DINER could provide more clear images than DNF and PhysenNet. Fig. 5 shows the PSNR of reconstructed measurements over training time. The proposed DINER has $100\times$ advantages on the convergence speed over the PhysenNet and DNF, respectively.

4. Comparisons of Refractive Index recovery

Please refer the attached video for visual comparison.

5. Comparisons of Novel View Synthesis

After the CVPR submission, we additionally applied the DINER to the novel view synthesis (NVS). We use the harmonic coefficients representations [1] to convert the continuous radiance field signal to a discrete signal. As a result, the problem of interpolating hash-key is avoided. We compare the DINER-based NVS with the NeRF [2] and the Plenoxels [1] (Fig. 6). The proposed DINER outperforms them.

References

[1] Sara Fridovich-Keil, Alex Yu, Matthew Tancik, Qinhong Chen, Benjamin Recht, and Angjoo Kanazawa. Plenoxels: Radiance fields without neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5501–5510, 2022. 1

[2] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages 405–421. Springer, 2020. 1

[3] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (TOG)*, 41(4):102:1–102:15, 2022. 1

[4] Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020. 1

[5] Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems*, 33:7537–7547, 2020. 1

[6] Fei Wang, Yaoming Bian, Haichao Wang, Meng Lyu, Giancarlo Pedrini, Wolfgang Osten, George Barbastathis, and Guohai Situ. Phase imaging with an untrained neural network. *Light: Science & Applications*, 9(1):1–7, 2020. 1

[7] You Zhou, Xia Hua, Weizhi Song, and Xun Cao. A novel denoising reconstruction algorithm for multi-height lensless microscopy. In *Biophotonics Congress: Biomedical Optics 2020 (Translational, Microscopy, OCT, OTS, BRAIN)*, page MTh4A.5. Optica Publishing Group, 2020. 1

[8] Hao Zhu, Zhen Liu, You Zhou, Zhan Ma, and Xun Cao. DNF: Diffractive neural field for lensless microscopic imaging. *Optics Express*, 30(11):18168–18178, 2022. 1

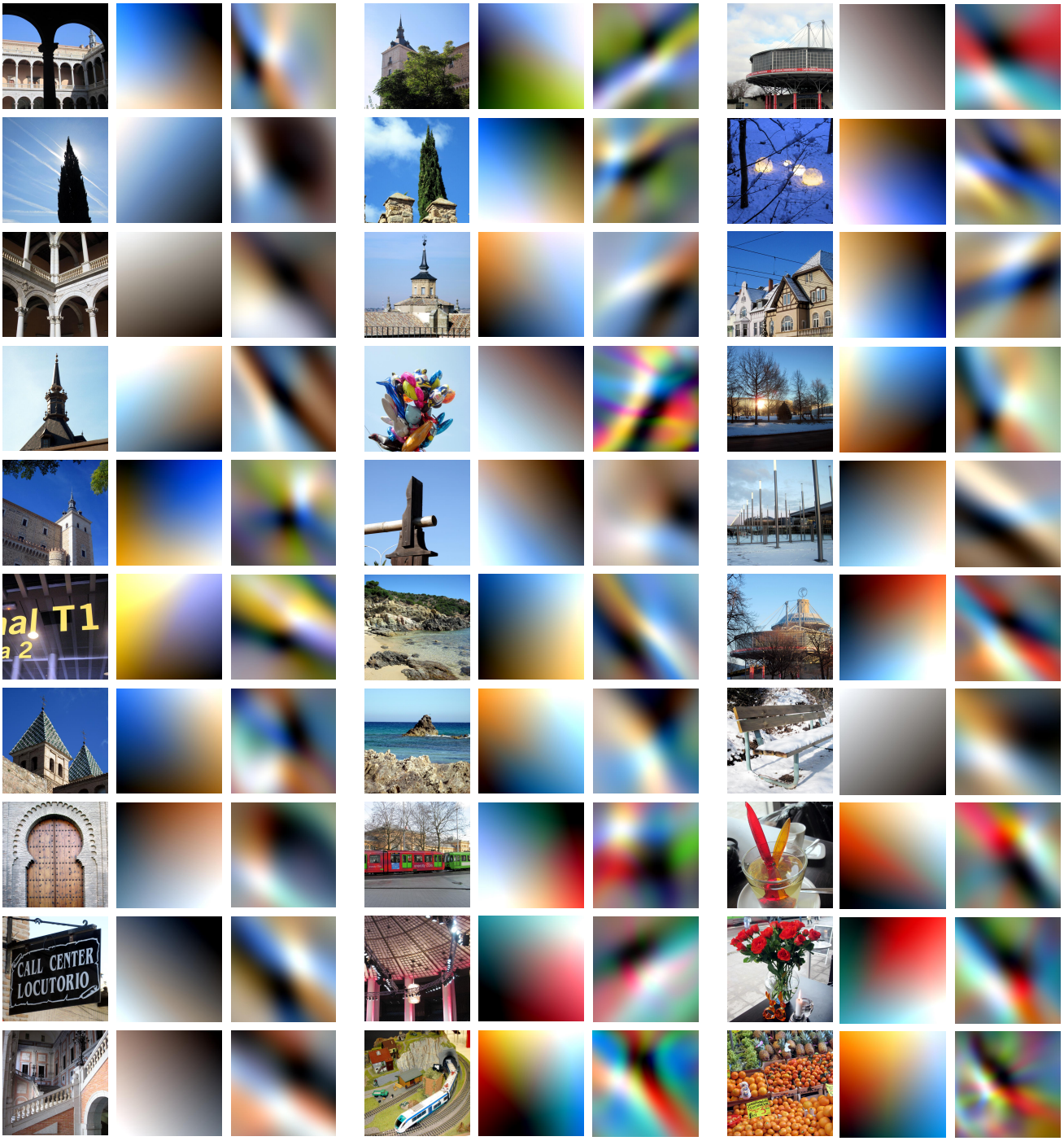

Figure 1. Comparisons of learned INRs in the DINER with MLP and SIREN backbones, respectively.

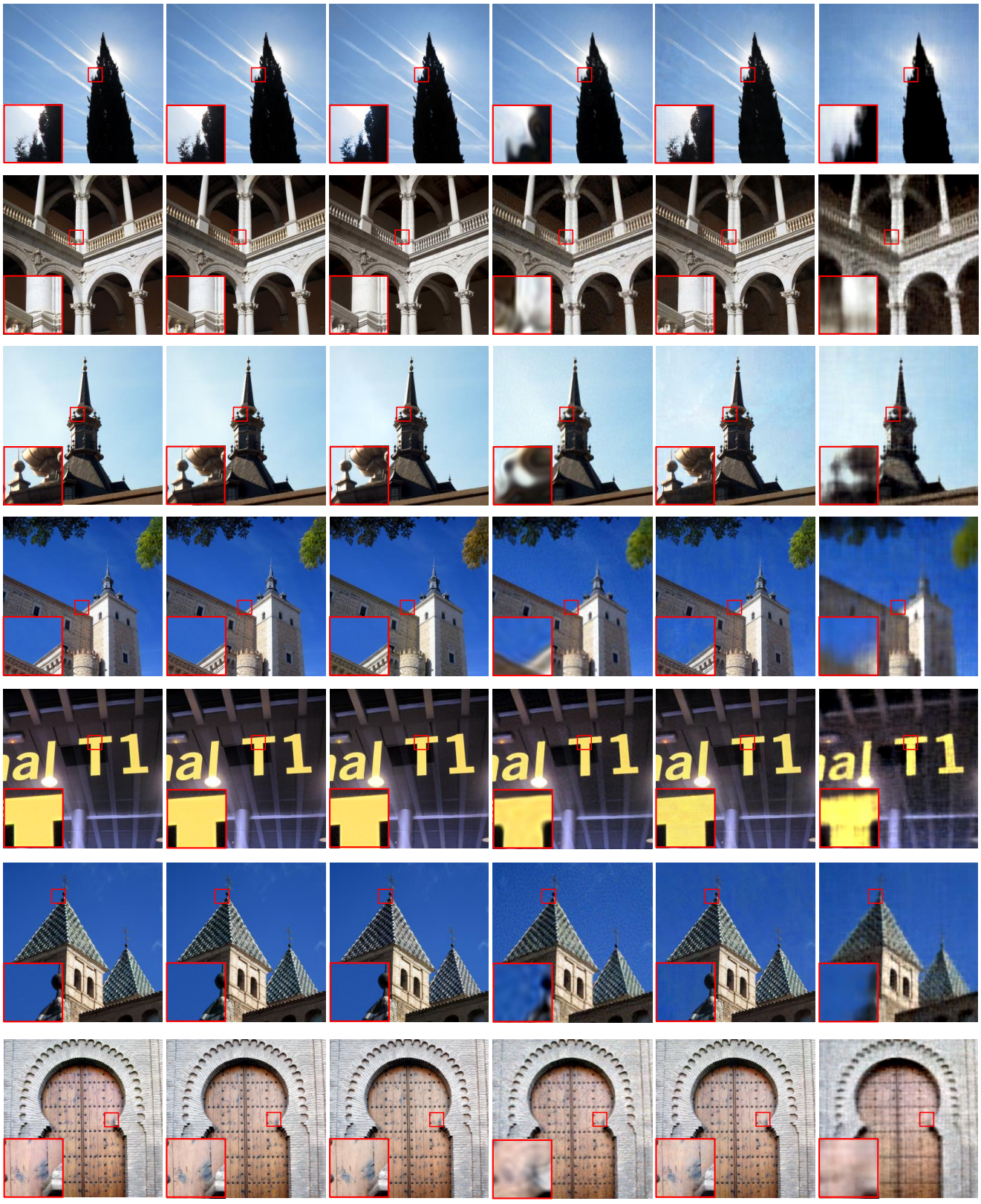

Figure 2. Comparisons of various methods on 2D image fitting after 3000 epochs.

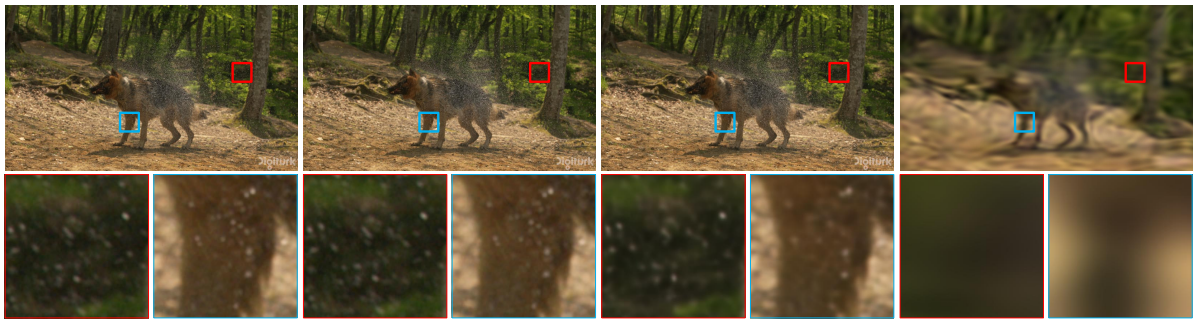

Figure 3. Qualitative comparisons of various methods on video representation after 500 epochs using the 'ShakeNDry' data.

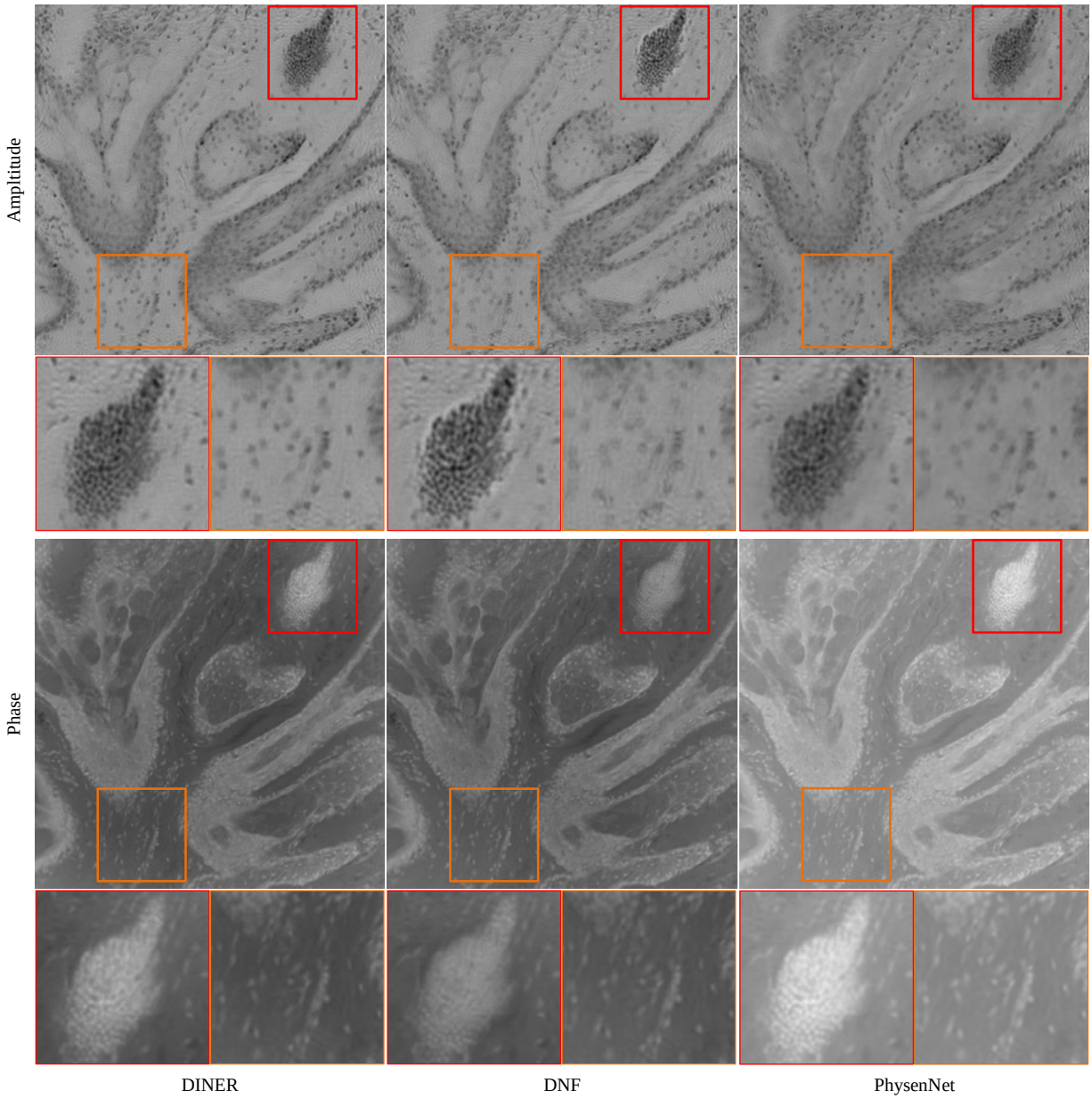

Figure 4. Comparisons on reconstructed complex field of animal skin section.

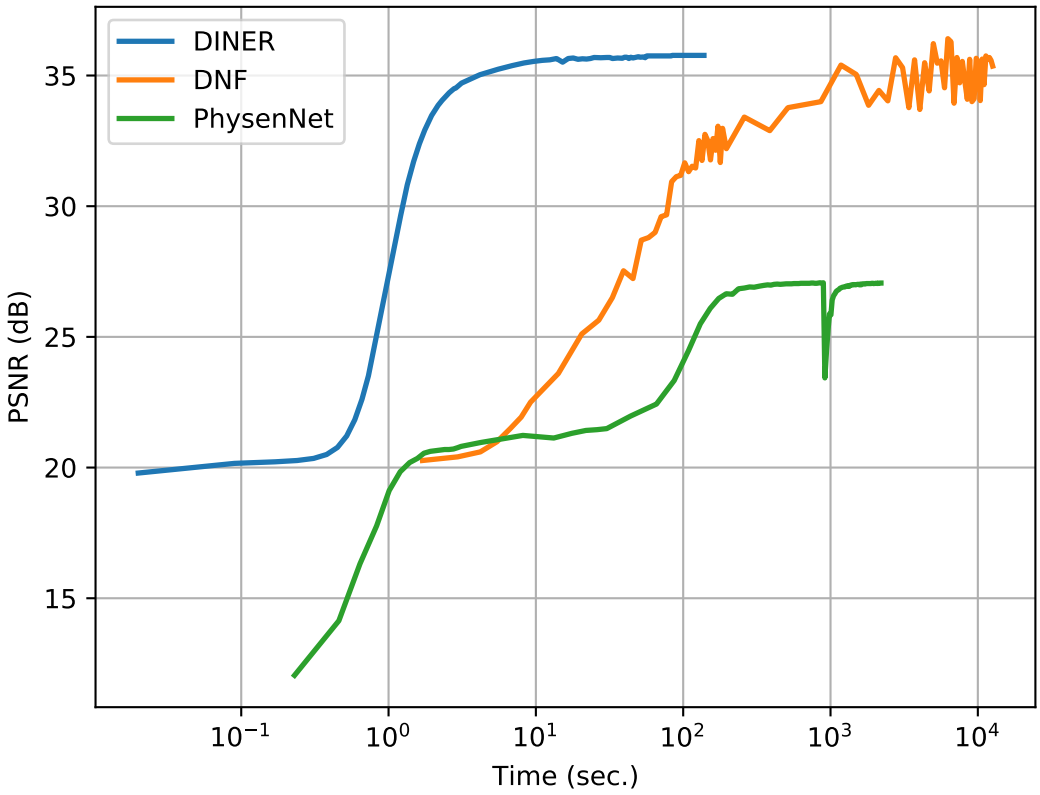

Figure 5. PSNR of reconstructed measurements of over training time on lensless imaging.

Figure 6. Results on novel view synthesis using A100 card.