

pixelSplat: 3D Gaussian Splats from Image Pairs for Scalable Generalizable 3D Reconstruction

Supplementary Material

7. Training Details

We train all methods using our data loaders, which gradually increase the distance between reference views during training. Specifically, over the first 150,000 training steps, we linearly increase the distance between reference views from 25 to 45.

7.1. Our Method

We train our method for 300,000 steps using a batch size of 7, which requires about 80 GB of VRAM on a single GPU. We use an MSE loss for the first 150,000 iterations and supplement it with an LPIPS loss with weight 0.05 starting at 150,000 steps. For each batch element (scene), we render 4 target views. To allow our method to produce gradients for multiple estimated depths during each forward pass, we place 3 Gaussians along each ray and determine their positions by independently sampling from the ray’s probability distribution 3 times. We then divide each Gaussian’s alpha value by 3 such that $\alpha = \frac{\phi_z}{3}$, which ensures that the Gaussians placed along any particular ray have a total opacity of roughly 1. For an overview of pixelSplat’s architecture, see Figure 8.

Depth regularization. To generate the point clouds shown in the main paper, we fine-tune our model for 50,000 steps with a depth regularization loss $\mathcal{L}_{\text{reg. depth}}$. To compute this loss, for each rendered view, we generate a corresponding depth map D . We use D to compute the loss as follows, where $D[\mathbf{u}_x, \mathbf{u}_y]$ represents indexing:

$$\begin{aligned} D_x^\Delta[\mathbf{u}] &= D[\mathbf{u}_{x-1}, \mathbf{u}_y] - 2D[\mathbf{u}_x, \mathbf{u}_y] + D[\mathbf{u}_{x+1}, \mathbf{u}_y] \\ D_y^\Delta[\mathbf{u}] &= D[\mathbf{u}_x, \mathbf{u}_{y-1}] - 2D[\mathbf{u}_x, \mathbf{u}_y] + D[\mathbf{u}_x, \mathbf{u}_{y+1}] \\ \mathbf{I}_x^\Delta &= \max(|\mathbf{I}[\mathbf{u}_{x-1}, \mathbf{u}_y] - \mathbf{I}[\mathbf{u}_x, \mathbf{u}_y]|, |\mathbf{I}[\mathbf{u}_{x+1}, \mathbf{u}_y] - \mathbf{I}[\mathbf{u}_x, \mathbf{u}_y]|) \\ \mathbf{I}_y^\Delta &= \max(|\mathbf{I}[\mathbf{u}_x, \mathbf{u}_{y-1}] - \mathbf{I}[\mathbf{u}_x, \mathbf{u}_y]|, |\mathbf{I}[\mathbf{u}_x, \mathbf{u}_{y+1}] - \mathbf{I}[\mathbf{u}_x, \mathbf{u}_y]|) \\ \mathcal{L}_{\text{reg. depth}}[\mathbf{u}] &= D_x^\Delta[\mathbf{u}] \exp\left(8 * \mathbf{I}_x^\Delta[\mathbf{u}]\right) + D_y^\Delta[\mathbf{u}] \exp\left(8 * \mathbf{I}_y^\Delta[\mathbf{u}]\right) \end{aligned} \quad (9)$$

Transformation of Gaussians into world space. Because our model predicts Gaussian parameters in camera space, these parameters must be transformed into world space before rendering. Given a 4×4 camera-to-world extrinsics matrix \mathbf{T} containing a 3×3 rotation \mathbf{R} and a 3×1 translation \mathbf{t} , we transform them as follows:

$$\begin{aligned} \boldsymbol{\mu}_{\text{world}} &= \mathbf{T}\boldsymbol{\mu}_{\text{cam.}} \\ \boldsymbol{\Sigma}_{\text{world}} &= \mathbf{R}\boldsymbol{\Sigma}_{\text{cam.}}\mathbf{R}^T \\ \alpha_{\text{world}} &= \alpha_{\text{cam.}} \\ \mathbf{s}_{\text{world}} &= \mathbf{D}\mathbf{R}\mathbf{s}_{\text{cam.}} \end{aligned} \quad (10)$$

Here, $\mathbf{D}\mathbf{R}$ is a block-diagonal matrix consisting of Wigner D matrices, which rotates the spherical harmonics coefficients $\mathbf{s}_{\text{cam.}}$. In practice, we use

Method	PSNR \uparrow	SSIM \uparrow	LPIPS \downarrow
Ours	26.09	0.863	0.136
Ours (3 Views)	28.31	0.908	0.100

Table 3. **Quantitative comparison of 2 and 3 view Real Estate 10k results.** Given a third reference view located halfway between the two existing reference views, a 3-view variant of pixelSplat produces slightly better results.

the e3nn library to compute these matrices and recompile the original 3D Gaussian splatting code base to follow e3nn’s conventions.

7.2. Method of Du et al.

We train the method of Du et al. [10] for 300,000 iterations using a total batch size of 32 spread across 4 GPUs, which requires around 44 GB of VRAM per GPU. We train using the authors’ default hyperparameters and enable the LPIPS loss after 150,000 iterations.

7.3. pixelNeRF

We train pixelNeRF [58] for 500,000 iterations using a batch size of 12, which requires about 20 GB of VRAM on a single GPU. We use the authors’ default hyperparameters for the NMR dataset, in which the first pooling layer of pixelNeRF’s ResNet is disabled to increase feature resolution. Following Du et al. [10], we set the near and far planes to be 0.1 and 10.0 respectively.

7.4. GPNR

We train GPNR [46] for 250,000 iterations using a batch size of 4098 spread across 6 GPUs, which requires about 67 GB of VRAM per GPU. We use the authors’ default hyperparameters but reduce the learning rate to $1 * 10^{-4}$, since several attempts at using the default learning rate of $3 * 10^{-4}$ yielded sudden training collapses on our dataset.

8. Using More Reference Views

While our epipolar encoder is primarily designed for novel view synthesis from pairs of images, it can be extended to an arbitrary number of views. Specifically, for a given pixel coordinate, epipolar samples can be taken from any number of images. The union of these samples can subsequently be used in place of a single epipolar line’s samples. To allow the epipolar transformer to distinguish between samples taken from different views, we add a learnable per-image embedding to each sample in this union. Figure 9 and Table 3 show 3-view results.

9. Limitations

Our model has several limitations. First, rather than fusing or de-duplicating Gaussians observed from both reference views, it simply outputs the union of the Gaussians predicted from each view. Second, it does not address generative modeling of unseen parts of the scene. Finally, when extended to many reference views, our epipolar attention mechanism becomes prohibitively expensive in terms of memory. Addressing these issues would be an exciting topic for future work.

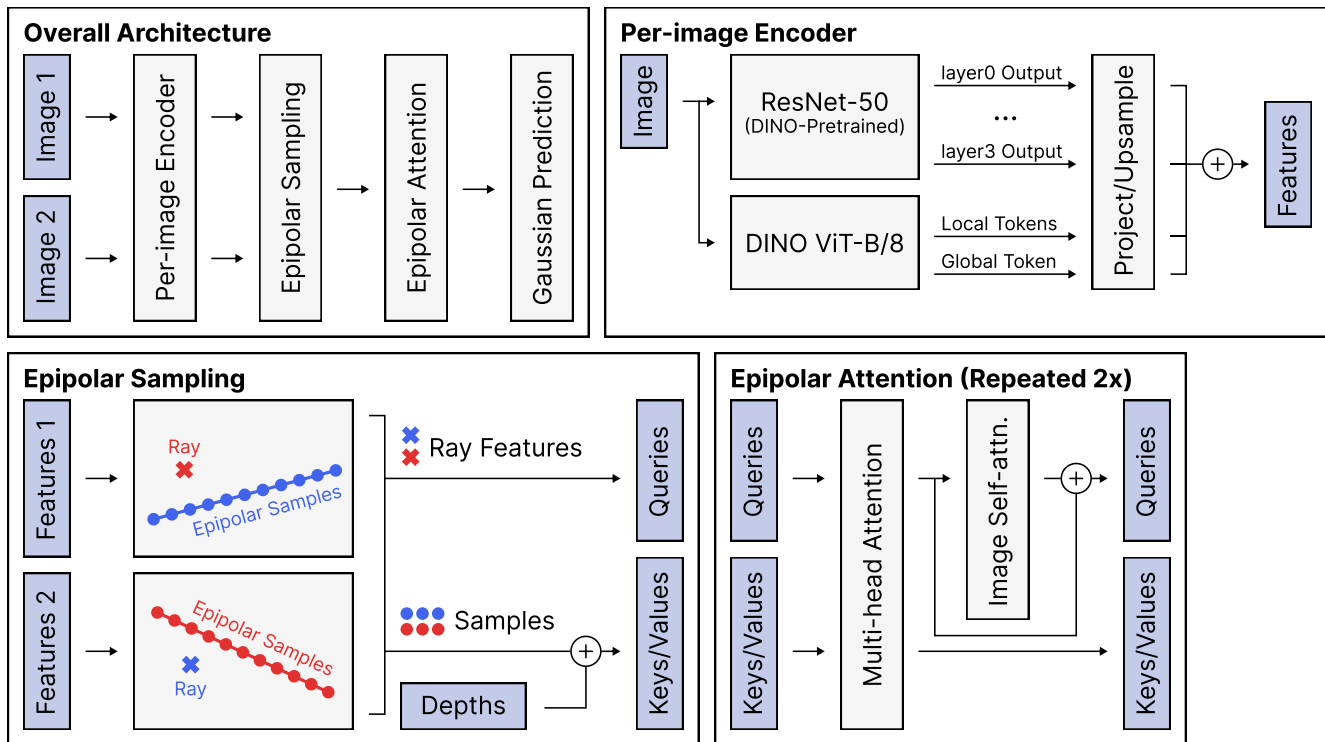


Figure 8. Architecture diagram.

10. Additional Results

We present additional results on the following pages.

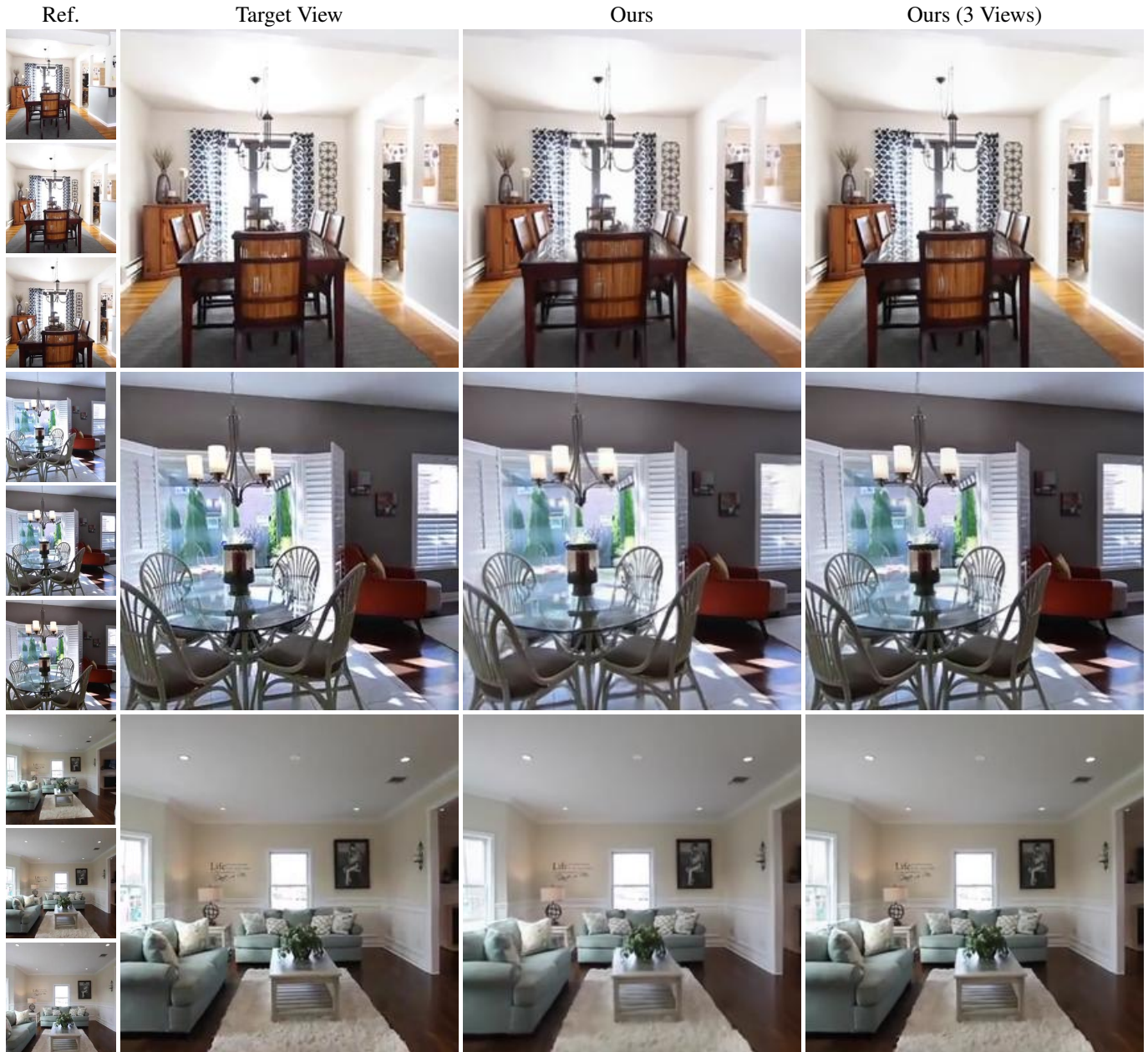


Figure 9. **Qualitative comparison of novel views given 2 and 3 reference views.** Our method can be extended to use an arbitrary, fixed number of reference views as input. Here, we compare the results from a 3-view model with those from the original 2-view model. The 2-view model uses the top and bottom reference views as input, while the 3-view model uses all three.

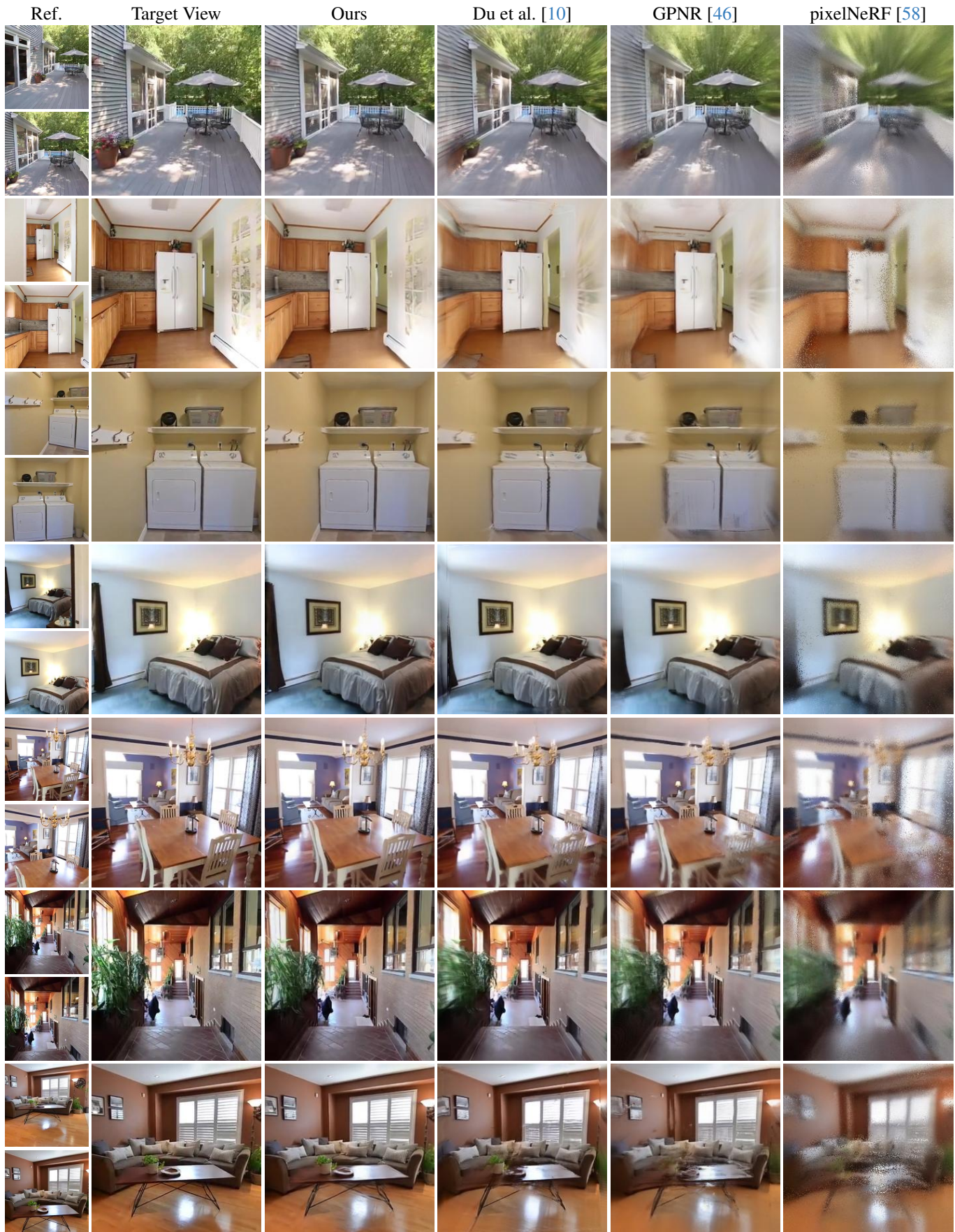


Figure 10. More results on the Real Estate 10k dataset.

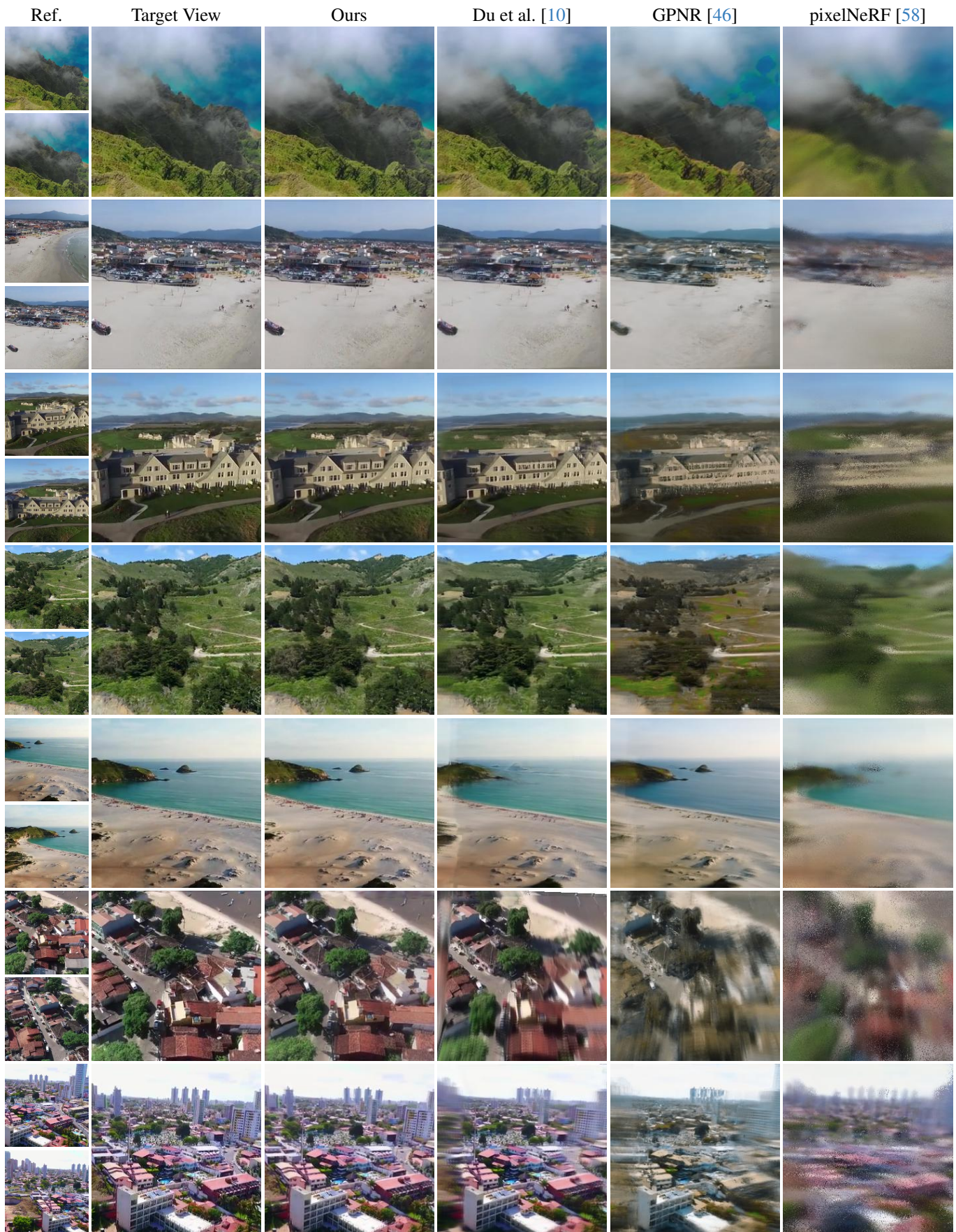


Figure 11. More results on the ACID dataset.

References

- [1] Kara-Ali Aliev, Artem Sevastopolsky, Maria Kolos, Dmitry Ulyanov, and Victor Lempitsky. Neural point-based graphics. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 696–712. Springer, 2020. [2](#)
- [2] Jonathan T. Barron, Ben Mildenhall, Matthew Tancik, Peter Hedman, Ricardo Martin-Brualla, and Pratul P. Srinivasan. Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. [2](#)
- [3] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. [6](#)
- [4] Eric R Chan, Koki Nagano, Matthew A Chan, Alexander W Bergman, Jeong Joon Park, Axel Levy, Miika Aittala, Shalini De Mello, Tero Karras, and Gordon Wetzstein. Generative novel view synthesis with 3d-aware diffusion models. *Proceedings of the International Conference on 3D Vision (3DV)*, 2023. [2](#)
- [5] Anpei Chen, Zexiang Xu, Fuqiang Zhao, Xiaoshuai Zhang, Fanbo Xiang, Jingyi Yu, and Hao Su. Mvsnerf: Fast generalizable radiance field reconstruction from multi-view stereo. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 14124–14133, 2021. [2](#)
- [6] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. *arXiv preprint arXiv:2203.09517*, 2022. [2](#)
- [7] Julian Chibane, Aayush Bansal, Verica Lazova, and Gerard Pons-Moll. Stereo radiance fields (srf): Learning view synthesis from sparse views of novel scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2021. [2](#)
- [8] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1):1–22, 1977. [2](#)
- [9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2021. [6](#)
- [10] Yilun Du, Cameron Smith, Ayush Tewari, and Vincent Sitzmann. Learning to render novel views from wide-baseline stereo pairs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. [1](#), [2](#), [5](#), [6](#), [7](#), [9](#), [12](#), [13](#)
- [11] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2014. [2](#), [3](#)
- [12] 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 Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5501–5510, 2022. [2](#)
- [13] Clément Godard, Oisín Mac Aodha, Michael Firman, and Gabriel J Brostow. Digging into self-supervised monocular depth estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. [2](#), [3](#)
- [14] Pengsheng Guo, Miguel Angel Bautista, Alex Colburn, Liang Yang, Daniel Ulbricht, Joshua M Susskind, and Qi Shan. Fast and explicit neural view synthesis. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 3791–3800, 2022. [2](#)
- [15] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. [6](#)
- [16] Yihui He, Rui Yan, Katerina Fragkiadaki, and Shoou-I Yu. Epipolar transformers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. [2](#), [3](#)
- [17] Chi Jin, Yuchen Zhang, Sivaraman Balakrishnan, Martin J Wainwright, and Michael I Jordan. Local maxima in the likelihood of gaussian mixture models: Structural results and algorithmic consequences. *Advances in Neural Information Processing Systems (NeurIPS)*, 29, 2016. [2](#)
- [18] Mohammad Mahdi Johari, Yann Lepoittevin, and François Fleuret. Geonerf: Generalizing nerf with geometry priors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18365–18375, 2022. [2](#)
- [19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics (ToG)*, 42(4): 1–14, 2023. [1](#), [2](#), [3](#), [4](#), [7](#)
- [20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. [6](#)
- [21] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *Proceedings of the International Conference on Learning Representations (ICLR)*, 2014. [5](#)
- [22] Georgios Kopanas, Julien Philip, Thomas Leimkühler, and George Drettakis. Point-based neural rendering with per-view optimization. In *Computer Graphics Forum*, pages 29–43. Wiley Online Library, 2021. [2](#)
- [23] Kai-En Lin, Lin Yen-Chen, Wei-Sheng Lai, Tsung-Yi Lin, Yi-Chang Shih, and Ravi Ramamoorthi. Vision transformer for nerf-based view synthesis from a single input image. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2022. [2](#)
- [24] Andrew Liu, Richard Tucker, Varun Jampani, Ameesh Makadia, Noah Snavely, and Angjoo Kanazawa. Infinite nature: Perpetual view generation of natural scenes from a single image. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. [6](#)
- [25] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. [2](#)
- [26] Yuan Liu, Sida Peng, Lingjie Liu, Qianqian Wang, Peng Wang, Christian Theobalt, Xiaowei Zhou, and Wenping Wang.

- Neural rays for occlusion-aware image-based rendering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7824–7833, 2022. 2
- [27] Stephen Lombardi, Tomas Simon, Jason Saragih, Gabriel Schwartz, Andreas Lehrmann, and Yaser Sheikh. Neural volumes: Learning dynamic renderable volumes from images. *ACM Trans. Graph.*, 2019. 2
- [28] Ricardo Martin-Brualla, Noha Radwan, Mehdi SM Sajjadi, Jonathan T Barron, Alexey Dosovitskiy, and Daniel Duckworth. Nerf in the wild: Neural radiance fields for unconstrained photo collections. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7210–7219, 2021. 2
- [29] 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 *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 405–421, 2020. 1, 2
- [30] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102:1–102:15, 2022. 2
- [31] Thu H Nguyen-Phuoc, Chuan Li, Stephen Balaban, and Yongliang Yang. Rendernet: A deep convolutional network for differentiable rendering from 3d shapes. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018. 2
- [32] Michael Niemeyer, Lars Mescheder, Michael Oechsle, and Andreas Geiger. Differentiable volumetric rendering: Learning implicit 3d representations without 3d supervision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3504–3515, 2020. 2
- [33] René Ranftl, Katrin Lasinger, David Hafner, Konrad Schindler, and Vladlen Koltun. Towards robust monocular depth estimation: Mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020. 2, 3
- [34] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12179–12188, 2021. 2
- [35] Jeremy Reizenstein, Roman Shapovalov, Philipp Henzler, Luca Sbordone, Patrick Labatut, and David Novotny. Common objects in 3d: Large-scale learning and evaluation of real-life 3d category reconstruction. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 10901–10911, 2021. 2
- [36] Gernot Riegler and Vladlen Koltun. Free view synthesis. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 2
- [37] Mehdi SM Sajjadi, Henning Meyer, Etienne Pot, Urs Bergmann, Klaus Greff, Noha Radwan, Suhani Vora, Mario Lucic, Daniel Duckworth, Alexey Dosovitskiy, et al. Scene representation transformer: Geometry-free novel view synthesis through set-latent scene representations. *arXiv preprint arXiv:2111.13152*, 2021. 1, 2
- [38] Kyle Sargent, Zizhang Li, Tanmay Shah, Charles Herrmann, Hong-Xing Yu, Yunzhi Zhang, Eric Ryan Chan, Dmitry Lagun, Li Fei-Fei, Deqing Sun, et al. Zeronvs: Zero-shot 360-degree view synthesis from a single real image. *arXiv preprint arXiv:2310.17994*, 2023. 2
- [39] Prafull Sharma, Ayush Tewari, Yilun Du, Sergey Zakharov, Rares Andrei Ambrus, Adrien Gaidon, William T Freeman, Fredo Durand, Joshua B Tenenbaum, and Vincent Sitzmann. Neural groundplans: Persistent neural scene representations from a single image. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2022. 2
- [40] Vincent Sitzmann, Justus Thies, Felix Heide, Matthias Nießner, Gordon Wetzstein, and Michael Zollhöfer. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1, 2
- [41] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems (NeurIPS)*, 32, 2019. 1, 2
- [42] Vincent Sitzmann, Julien N.P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 2
- [43] Vincent Sitzmann, Semon Rezkikov, Bill Freeman, Josh Tenenbaum, and Fredo Durand. Light field networks: Neural scene representations with single-evaluation rendering. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 1, 2
- [44] Cameron Smith, Yilun Du, Ayush Tewari, and Vincent Sitzmann. Flowcam: Training generalizable 3d radiance fields without camera poses via pixel-aligned scene flow. *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 8
- [45] Pratul P Srinivasan, Richard Tucker, Jonathan T Barron, Ravi Ramamoorthi, Ren Ng, and Noah Snavely. Pushing the boundaries of view extrapolation with multiplane images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 175–184, 2019. 2
- [46] Mohammed Suhail, Carlos Esteves, Leonid Sigal, and Ameesh Makadia. Generalizable patch-based neural rendering. In *Proceedings of the European Conference on Computer Vision (ECCV)*. Springer, 2022. 2, 6, 7, 9, 12, 13
- [47] Mohammed Suhail, Carlos Esteves, Leonid Sigal, and Ameesh Makadia. Light field neural rendering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8269–8279, 2022. 1, 2, 5
- [48] Stanislaw Szymanowicz, Christian Rupprecht, and Andrea Vedaldi. Viewset diffusion:(0-) image-conditioned 3d generative models from 2d data. *Proceedings of the International Conference on Computer Vision (ICCV)*, 2023. 8
- [49] Andrea Tagliasacchi and Ben Mildenhall. Volume rendering digest (for nerf). *arXiv preprint arXiv:2209.02417*, 2022. 2
- [50] Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Tretschk, W Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. Advances in neural rendering. In *Computer Graphics Forum*, pages 703–735. Wiley Online Library, 2022. 1, 2

- [51] Ayush Tewari, Tianwei Yin, George Cazenavette, Semon Rezhchikov, Joshua B Tenenbaum, Frédo Durand, William T Freeman, and Vincent Sitzmann. Diffusion with forward models: Solving stochastic inverse problems without direct supervision. *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. [2](#), [8](#)
- [52] Alex Trevithick and Bo Yang. Grf: Learning a general radiance field for 3d representation and rendering. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 15182–15192, 2021. [2](#)
- [53] Richard Tucker and Noah Snavely. Single-view view synthesis with multiplane images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 551–560, 2020. [2](#)
- [54] Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul Srinivasan, Howard Zhou, Jonathan T. Barron, Ricardo Martin-Brualla, Noah Snavely, and Thomas Funkhouser. Ibrnet: Learning multi-view image-based rendering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. [2](#)
- [55] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004. [6](#)
- [56] Olivia Wiles, Georgia Gkioxari, Richard Szeliski, and Justin Johnson. Synsin: End-to-end view synthesis from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7467–7477, 2020. [2](#)
- [57] Yiheng Xie, Towaki Takikawa, Shunsuke Saito, Or Litany, Shiqin Yan, Numair Khan, Federico Tombari, James Tompkin, Vincent Sitzmann, and Srinath Sridhar. Neural fields in visual computing and beyond. *Computer Graphics Forum*, 2022. [2](#)
- [58] Alex Yu, Vickie Ye, Matthew Tancik, and Angjoo Kanazawa. pixelNeRF: Neural radiance fields from one or few images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. [1](#), [2](#), [6](#), [7](#), [9](#), [12](#), [13](#)
- [59] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep networks as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. [6](#)
- [60] Yunzhi Zhang and Jiajun Wu. Video extrapolation in space and time. *arXiv e-prints*, pages arXiv–2205, 2022. [2](#)
- [61] Tinghui Zhou, Richard Tucker, John Flynn, Graham Fyffe, and Noah Snavely. Stereo magnification: Learning view synthesis using multiplane images. *ACM Trans. Graph. (Proc. SIGGRAPH)*, 37, 2018. [2](#), [6](#)