NeRFLiX: High-Quality Neural View Synthesis by Learning a Degradation-Driven Inter-viewpoint MiXer

Kun Zhou\textsuperscript{1,2}\textsuperscript{*}  Wenbo Li\textsuperscript{3}\textsuperscript{*}  Yi Wang\textsuperscript{4}
Tao Hu\textsuperscript{3}  Nianjuan Jiang\textsuperscript{2}  Xiaoguang Han\textsuperscript{1}  Jiangbo Lu\textsuperscript{2}\textsuperscript{†}
\textsuperscript{1}SSE, CUHK-Shenzhen,  \textsuperscript{2}SmartMore Corporation  \textsuperscript{3}CUHK  \textsuperscript{4}Shanghai AI Laboratory
kunzhou@link.cuhk.edu.cn,\{wenboli,taohu\}@cse.cuhk.edu.hk
hanxiaoguang@cuhk.edu.cn,\{jiangbo.lu,wygamle\}@gmail.com

Abstract

Neural radiance fields (NeRF) show great success in novel view synthesis. However, in real-world scenes, recovering high-quality details from the source images is still challenging for the existing NeRF-based approaches, due to the potential imperfect calibration information and scene representation inaccuracy. Even with high-quality training frames, the synthetic novel views produced by NeRF models still suffer from notable rendering artifacts, such as noise, blur, etc. Towards to improve the synthesis quality of NeRF-based approaches, we propose NeRFliX, a general NeRF-agnostic restorer paradigm by learning a degradation-driven inter-viewpoint mixer. Specially, we design a NeRF-style degradation modeling approach and construct large-scale training data, enabling the possibility of effectively removing NeRF-native rendering artifacts for existing deep neural networks. Moreover, beyond the degradation removal, we propose an inter-viewpoint aggregation framework that is able to fuse highly related high-quality training images, pushing the performance of cutting-edge NeRF models to entirely new levels and producing highly photo-realistic synthetic views.

1. Introduction

Neural radiance fields (NeRF) can generate photo-realistic images from new viewpoints, playing a heated role in novel view synthesis. In light of NeRF’s\textsuperscript{37} success, numerous approaches [2, 9, 11, 19, 34, 35, 38, 40, 41, 47, 53, 54, 59, 61, 65] along these lines have been proposed, contin-
ually raising the performance to greater levels. In fact, one prerequisite of NeRF is the precise camera settings of the taken photos for training [22, 32, 61]. However, accurately calibrating camera poses is exceedingly difficult in practice. Contrarily, the shape-radiance co-adaption issue [74] reveals that while the learned radiance fields can perfectly explain the training views with inaccurate geometry, they poorly generalize to unseen views. On the other hand, the capacity to represent sophisticated geometry, lighting, object materials, and other factors is constrained by the simplified scene representation of NeRF [19, 77, 78]. On the basis of such restrictions, advanced NeRF models may nonetheless result in notable artifacts (such as blur, noise, detail missing, and more), which we refer to as NeRF-style degradations in this article and are shown in Fig. 1.

To address the aforementioned limitations, numerous works have been proposed. For example, some studies, including [22, 59, 66, 70], jointly optimize camera parameters and neural radiance fields to refine camera poses as precisely as possible in order to address the camera calibration issue. Another line of works [19, 73, 77, 78] presents physical-aware models that simultaneously take into account the object materials and environment lighting, as opposed to using MLPs or neural voxels to implicitly encode both the geometry and appearance. To meet the demands for high-quality neural view synthesis, one has to carefully examine all of the elements when building complex inverse rendering systems. In addition to being challenging to optimize, they are also not scalable for rapid deployment with hard re-configurations in new environments. Regardless of the intricate physical-aware rendering models, is it possible to design a practical NeRF-agnostic restorer to directly enhance synthesized views from NeRFs?

In the low-level vision, it is critical to construct large-scale paired data to train a deep restorer for eliminating real-world artifacts [56, 72]. When it comes to NeRF-style degradations, there are two challenges: (1) sizable paired training data; (2) NeRF degradation analysis. First, it is impractical to gather large-scale training pairs (more specifically, raw outputs from well-trained NeRFs and corresponding ground truths). Second, the modeling of NeRF-style degradation has received little attention. Unlike real-world images that generally suffer from JPEG compression, sensor noise, and motion blur, the NeRF-style artifacts are complex and differ from the existing ones. As far as we know, no previous studies have ever investigated NeRF-style degradation removal which effectively leverages the ample research on image and video restoration.

In this work, we are motivated to have the first study on the feasibility of simulating large-scale NeRF-style paired data, opening the possibility of training a NeRF-agnostic restorer for improving the NeRF rendering frames. To this end, we present a novel degradation simulator for typical NeRF-style artifacts (e.g., rendering noise and blur) considering the NeRF mechanism. We review the overall NeRF rendering pipeline and discuss the typical NeRF-style degradation cases. Accordingly, we present three basic degradation types to simulate the real rendered artifacts of NeRF synthetic views and empirically evaluate the distribution similarity between real rendered photos and our simulated ones. The feasibility of developing NeRF-agnostic restoration models has been made possible by constructing a sizable dataset that covers a variety of NeRF-style degradations, over different scenes.

Next, we show the necessity of our simulated dataset and demonstrate that existing state-of-the-art image restoration frameworks can be used to eliminate NeRF visual artifacts. Furthermore, we notice, in a typical NeRF setup, neighboring high-quality views come for free, and they serve as potential reference bases for video-based restoration with a multi-frame aggregation and fusion module. However, this is not straightforward because NeRF input views are taken from a variety of very different angles and locations, making the estimation of correspondence quite challenging. To tackle this problem, we propose a degradation-driven viewpoint “mixer” that progressively aligns image contents at the pixel and patch levels. In order to maximize efficiency and improve performance, we also propose a fast view selection technique to only choose the most pertinent reference training views for aggregation, as opposed to using the entire NeRF input views.

In a nutshell, we present a NeRF-agnostic restorer (termed NeRFLiX) which learns a degradation-driven viewpoint mixer. As illustrated in Fig. 1, given NeRF synthetic frames with various rendering degradations, NeRFLiX successfully restores high-quality results. Our contributions are summarized as:

- **Universal enhancer for NeRF models.** NeRFLiX is powerful and adaptable, removing NeRF artifacts and restoring clearly details, pushing the performance of cutting-edge NeRF models to entirely new levels.

- **NeRF rendering degradation simulator.** We develop a NeRF-style degradation simulator (NDS), constructing massive amounts of paired data and aiding the training of deep neural networks to improve the quality of NeRF-rendered images.

- **Inter-viewpoint mixer.** Based on our constructed NDS, we further propose an inter-viewpoint baseline that is able to mix high-quality neighboring views for more effective restorations.

- **Training time acceleration.** We show how NeRFLiX makes it possible for NeRF models to produce even better results with a 50% reduction in training time.
2. Related Works

NeRF-based novel view synthesis. NeRF-based novel view synthesis has received a lot of attention recently and has been thoroughly investigated. For the first time, Mildenhall et al. [37] propose the neural radiance field to implicitly represent static 3D scenes and synthesize novel views from multiple posed images. Inspired by their successes, a lot of NeRF-based models [2, 8, 11, 13, 19–21, 23, 26, 33–35, 38, 41, 42, 44, 47, 51, 54, 62, 65, 73, 76] have been proposed. For example, point-NeRF [63] and DS-NeRF [14] incorporate sparse 3D point cloud and depth information for eliminating the geometry ambiguity for NeRFs, achieving more accurate/efficient 3D point sampling and better rendering quality. Plenoxels [16], TensorRF [7], DirectVoxGo [45], FastNeRF [17], Plenoctrees [67], KiloNeRF [43], and MobileNeRF [10], aim to use various advanced technologies to speed up the training or inference phases. Though these methods have achieved great progress, due to the potential issue of inaccurate camera poses, simplified pinhole camera models as well as scene representation inaccuracy, they still suffer from rendering artifacts of the predicted novel views.

Degradation simulation. Since no existing works have explored the NeRF-style degradation cases, we will overview the real-world image restoration works that are most related to ours. The previous image/video super-resolution approaches [15, 27, 28, 31, 55, 57, 68, 79, 79, 80] typically follow a fix image degradation type (e.g., blur, bicubic/bilinear down-sampling). Due to the large domain shift between the real-world and simulated degradations, the earlier image restoration methods [27, 29, 71, 79] generally fail to remove complex artifacts of the real-world images. In contrast, BSRGAN [72] design a practical degradation approach for real-world image super-resolution. In their degradation process, multiple degradations are considered and applied in random orders, largely covering the diversity of real-world degradations. Compared with the previous works, BSRGAN achieves much better results quantitatively and qualitatively. Real-ESRGAN [56] develops a second-order degradation process for real-world image super-resolution. In this work, we propose a NeRF-style degradation simulator and construct a large-scale training dataset for modeling the NeRF rendering artifacts.

Correspondence estimation. In the existing literature, the video restoration methods [3, 6, 48, 52, 69] aim to restore a high-quality frame from multiple low-quality frames. To achieve this goal, cross-frame correspondence estimation is essential to effectively aggregate the informative temporal contents. Some works [5, 6, 64, 69] explore building pixel-level correspondences through optical-flow estimation and perform frame-warping for multi-frame compensation. Another line of works [49, 55, 80] tries to use deformable convolution networks (DCNs [12]) for adaptive correspondence estimation and aggregation. More recently, transformer-based video restoration models [4, 30] implement spatial-temporal aggregation through an attention mechanism and achieve promising performance. However, it is still challenging to perform accurate correspondence estimation between frames captured with very distinctive viewpoints.

3. Preliminaries

Figure 2. A general illustration of NeRF-based novel view synthesis pipeline. Three main steps are involved: (1) ray shooting, (2) ray marching, and (3) radiance accumulation.

In this section, we review the general pipeline of NeRF-based novel view synthesis and discuss potential rendering artifacts. As shown in Fig. 2, three main steps are involved in the rendering: (1) Ray Shooting. To render the color of a target pixel in a particular view, NeRF utilizes the camera’s calibrated parameters \( \pi \) to generate a ray \( r(o, \mathbf{d}) \) through this pixel, where \( o, \mathbf{d} \) are the camera center and the ray direction. (2) Ray Marching. A set of 3D points are sampled along the chosen ray as it moves across the 3D scene represented by neural radiance fields. The NeRF models encode a 3D scene and predict the colors and densities of these points. (3) Radiance Accumulation. The pixel color is extracted by integrating the predicted radiance features of the sampled 3D points.

Discussion. We can see that establishing a relationship between 2D photos and the 3D scene requires camera calibration. Unfortunately, it is very challenging to precisely calibrate the camera poses, leading to noisy 3D sampling. Meanwhile, some previous works [22, 59, 66, 70] also raise other concerns, including the non-linear pinhole camera model [22] and shape-radiance ambiguity [74]. Because of these inherent limitations, as discussed in Section 1, NeRF models still synthesize unsatisfied novel test views.

4. Methodology

Overview. In this work, we present NeRFLiX, a general NeRF-agnostic restorer which employs a degradation-driven inter-viewpoint mixer to enhance novel view images rendered by NeRF models. It is made up of two essential components: a NeRF-style degradation simulator (NDS) and an inter-viewpoint mixer (IVM). As seen in Fig. 3(a), during the training phase, we employ the proposed NDS
to create large-scale paired training data, which are subsequently used to train an IVM for improving a NeRF-rendered view using two corresponding reference pictures (reference views). In the inference stage, as illustrated in Fig. 3(b), IVM is adopted to enhance a rendered view by fusing useful information from the selected most relevant reference views.

4.1. NeRF-Style Degradation Simulator (NDS)

Due to the difficulties in gathering well-posed scenes under various environments and training NeRF models for each scene, it is infeasible to directly collect large amounts of paired NeRF data for artifact removal. To address this challenge, motivated by BSRGAN [72], we design a general NeRF degradation simulator to produce a sizable training dataset that is visually and statistically comparable to NeRF-rendered images (views).

To begin with, we collect raw data from LLFF-T\(^1\) and Vimeo90K [64] where the adjacent frames are treated as raw sequences. Each raw sequence consists of three images \(\{I^{gt}, I_1^{gt}, I_2^{gt}\}\): a target view \(I^{gt}\) and its two reference views \(\{I_1^{gt}, I_2^{gt}\}\). To construct the paired data from a raw sequence, we use the proposed NDS to degrade \(I^{gt}\) and obtain a simulated degraded view \(I\), as shown in Fig. 3(a).

The degradation pipeline is illustrated in Fig 4. We design three types of degradation for a target view \(I^{gt}\): splatted Gaussian noise (SGN), re-positioning (Re-Pos.), and anisotropic blur (A-Blur). It should be noted that there may be other models for such a simulation, and we only utilize this route to evaluate and justify the feasibility of our idea.

**Splatted Gaussian noise.** Although additive Gaussian noise is frequently employed in image/video denoising, NeRF rendering noise clearly differs. Rays that hit a 3D point will be re-projected within a nearby 2D area because of noisy camera parameters. As a result, the NeRF-style noise is dispersed over a 2D space. This observation led us

\(^{1}\)the training parts of LLFF [36].
4.2. Inter-viewpoint Mixer (IVM)

Problem formulation. Given a degraded view \( I \) produced by our NDS or NeRF models, we aim to extract useful information from its two high-quality reference views \( \{ I_1^r, I_2^r \} \), and restore an enhanced version \( \hat{I} \). To this end, we develop a view selection strategy to choose two reference views \( \{ I_1^r, I_2^r \} \) from the input views that are most overlapped with the rendered view \( I \). Specifically, we formulate the view selection problem based on the pinhole camera model. An arbitrary 3D scene can be roughly approximated as a bounding sphere in Fig. 7, and cameras are placed around it to take pictures. When camera-emitted rays hit the sphere, there are a set of intersections. We refer to the 3D point sets as \( \Phi_i = \{ p_{i0}, p_{i1}, \ldots, p_{iM_i} \} \) and \( \Phi_j = \{ p_{j0}, p_{j1}, \ldots, p_{jM_j} \} \) for the \( i \)-th and \( j \)-th cameras. For \( m_i \)-th intersection \( p_{m_i}^i \in \Phi_i \) of view \( i \), we search its nearest point in view \( j \) with the L2 distance

\[
p_{m_i}^{i \rightarrow j} = \arg \min_{p \in \Phi_j} \| p - p_{m_i}^i \|_2^2.
\]

Then the matching cost from the \( i \)-th view to the \( j \)-th view is calculated by

\[
C_{i \rightarrow j} = \sum_{m_i = 0}^{M_i} \| p_{m_i}^i - p_{m_i}^{i \rightarrow j} \|_2^2.
\]

IVM architecture. For multi-frame processing, existing techniques either use optical flow \([5, 52, 69]\) or deformable convolutions \([12, 30, 55]\) to realize the correspondence estimation and aggregation for consistent displacements. In contrast, NeRF rendered and input views come from very different angles and locations, making it challenging to perform precise inter-viewpoint aggregation.

To address this problem, we propose IVM, a hybrid recurrent inter-viewpoint “mixer” that progressively fuses pixel-wise and patch-wise contents from two high-quality reference views, achieving more effective inter-viewpoint aggregation. There are three modules i.e., feature extraction, hybrid inter-viewpoint aggregation and reconstruction, as shown in Fig. 6. Two convolutional encoders are used in the feature extraction stage to process the degraded view \( I \) and two high-quality reference views \( \{ I_1^r, I_2^r \} \), respectively. We then use inter-viewpoint window-based attention modules and deformable convolutions to achieve recurrent patch-wise and pixel-wise aggregation. Finally, the enhanced view \( \hat{I} \) is generated using the reconstruction module under the supervision

\[
\text{Loss} = |\hat{I} - I^{gt}|, \text{ where } \hat{I} = f(I, I_1^r, I_2^r; \theta),
\]

where \( \theta \) is the learnable parameters of IVM. The framework architecture is given in our supplementary materials.

4.3. View Selection

In the inference stage, for a NeRF-rendered view \( I \), our IVM produces an enhanced version by aggregating contents from two neighboring high-quality views. But, multiple input views are available and only a part of them are largely overlapped with \( I \). In general, only the most pertinent input views are useful for the inter-viewpoint aggregation.

To this end, we develop a view selection strategy to choose two reference views \( \{ I_1^r, I_2^r \} \) from the input views that are most overlapped with the rendered view \( I \). Specifically, we formulate the view selection problem based on the pinhole camera model. An arbitrary 3D scene can be roughly approximated as a bounding sphere in Fig. 7, and cameras are placed around it to take pictures. When camera-emitted rays hit the sphere, there are a set of intersections. We refer to the 3D point sets as \( \Phi_i = \{ p_{i0}, p_{i1}, \ldots, p_{iM_i} \} \) and \( \Phi_j = \{ p_{j0}, p_{j1}, \ldots, p_{jM_j} \} \) for the \( i \)-th and \( j \)-th cameras. For \( m_i \)-th intersection \( p_{m_i}^i \in \Phi_i \) of view \( i \), we search its nearest point in view \( j \) with the L2 distance

\[
p_{m_i}^{i \rightarrow j} = \arg \min_{p \in \Phi_j} \| p - p_{m_i}^i \|_2^2.
\]

Then the matching cost from the \( i \)-th view to the \( j \)-th view is calculated by

\[
C_{i \rightarrow j} = \sum_{m_i = 0}^{M_i} \| p_{m_i}^i - p_{m_i}^{i \rightarrow j} \|_2^2.
\]

We finally obtain the mutual matching cost between views \( i \) and \( j \) as

\[
C_{i \leftrightarrow j} = C_{i \rightarrow j} + C_{j \rightarrow i}.
\]

In this regard, two reference views \( \{ I_1^r, I_2^r \} \) are selected at the least mutual matching costs for enhancing the NeRF-rendered view \( I \). Note that we also adopt this strategy to decide the two reference views for the LLFF-T \([36]\) data during the training phase.
5. Experiments

5.1. Implementation Details

We train the IVM for 300K iterations. The batch size is 16 and the patch size is 128. We adopt random cropping, vertical or horizontal flipping, and rotation augmentations. Apart from the inherent viewpoint changes over \{I, I_1, I_2\}, random offsets (±5 pixels) are globally applied to the two reference views \(I_1, I_2\) to model more complex motion. We adopt an Adam [23] optimizer and a Cosine annealing strategy to decay the learning rate from 5 × 10^-4 to 0. We train a single IVM on the LLFF-T and Vimeo datasets and test it on all benchmarks (including user-captured scenes).

5.2. Datasets and Metrics

We conduct the experiments on three widely used datasets, including LLFF [36], Tanks and Temples [25], and Noisy LLFF Synthetic.

**LLFF [36].** LLFF is a real-world dataset, where 8 different scenes have 20 to 62 images. Following the commonly used protocols [1, 7, 16, 39, 59], we adopt 1008 × 756 resolution for LLFF-P1 and 504 × 376 resolution for LLFF-P2.

**Tanks and Temples [25].** It contains 5 scenes captured by inward-facing cameras. There are 152-384 images in the 1920 × 1080 resolution. It should be noted that the viewpoints of different frames are significantly larger than LLFF.

**Noisy LLFF Synthetic [37].** There are 8 virtual scenes, each of which has 400 images with a size of 800 × 800. To simulate noisy in-the-wild calibration, we ad hoc apply camera jittering (random rotation and translation are employed) to the precise camera poses.

**Metrics.** Following previous NeRF methods, we adopt PSNR (↑)/SSIM (↑)/LPIPS (↓) for evaluation.

5.3. Improvement over SOTA NeRF Models

We demonstrate the effectiveness of our approach by showing that it consistently improves the performance of cutting-edge NeRF approaches across various datasets.

**LLFF.** In order to fully verify the generalization ability of our NeRFliX, we investigate six representative models, including NeRF [37], TensoRF [7], Plenoxels [16], NeRF-mm [59], NLF [1], and RegNeRF [39]. Using rendered images of NeRF models as inputs to our model, we aim to further improve the synthesis quality. The quantitative results are provided in Table 1. We find that under both of the two protocols, our method raises NeRF model performance entirely to new levels. For example, NeRFliX improves NeRF by 0.40 dB/0.025/0.054 in terms of PSNR/SSIM/LPIPS.

**Tanks and Temples.** Due to large variations of camera viewpoints, even advanced NeRF models, e.g., TensoRF [7] and DIVeR [60], show obviously inferior rendering quality on this dataset. As illustrated in Table 2a, we show that our NeRFliX can still boost the performance of these models by a large margin, especially TensoRF [7] achieves 0.51 dB/0.010/0.022 improvements on PSNR/SSIM/LPIPS.
**Noisy LLFF Synthetic.** Apart from in-the-wild benchmarks above, we also demonstrate the enhancement capability of our model on noisy LLFF Synthetic. From the results shown in Table 2b, we see that our NeRFLiX yields substantial improvements for two SOTA NeRF models.

**Qualitative results.** In Fig. 8, we provide some visual examples for qualitative assessment. It is obvious that our NeRFLiX restores clearer image details while removing the majority of NeRF-style artifacts in the rendered images, clearly manifesting the effectiveness of our method. More results are provided in our supplementary materials.

5.4. Training Acceleration for NeRF Models

In this section, we show how our approach makes it possible for NeRF models to produce better results even with a 50% reduction in training time. To be more precise, we use NeRFLiX to improve the rendered images of two SOTA NeRF models after training them with half the training period specified in the publications. The enhanced results outperform the counterparts with full-time training, as shown in Table 2c. Notably, NeRFLiX has reduced the training period for Plenoxels [16] from 24 minutes to 10 minutes while also improving the quality of the rendered images.

5.5. Ablation Study

In this section, we conduct comprehensive experiments on LLFF [36] under the LLFF-P1 protocol to analyze each of our designs. We use TensoRF [7] as our baseline.\(^2\)

5.5.1 NeRF-Style Degradation Simulator

**Simulation quality.** We first examine the simulation quality of the proposed NeRF-style degradation simulator. To this end, we analyze the distribution of our degraded images, BSR [72] degraded images and NeRF rendered images on LLFF [36]. We use t-SNE [50] to visualize deep image features (by Inception-v3 [46]) and results are shown in Fig. 9. Our simulated data is statistically much closer to the real rendered images than BSR. This conclusion is also supported by Table 3b, which demonstrates that our NDS significantly surpasses BSR and yields 0.6-1.0dB improvements when used for learning NeRF degradations.

**Degradation type.** We also evaluate the detailed contribution of each data degradation. We use simulated data to train our models by gradually including four types of degradation, as illustrated in Table 4. From quantitative comparisons on LLFF [36], we observe that all employed degradations are beneficial to our system.

5.5.2 Inter-viewpoint Mixer

**View selection strategy.** We develop a view selection strategy to make full use of high-quality reference views. As shown in Fig. 10, our system can identify the most relevant views for quality enhancement when compared to random selection. Also, the quantitative results in Table 5 sug-
Table 4. Influences of different degradations used in our NeRF-style degradation simulator. “SGN” and “RA” are shorted for splatted Gaussian noise and region-adaptive schemes and “A-Blur” refers to anisotropic Gaussian blur.

Table 5. Ablation studies of our view selection strategy.

Table 6. Ablation studies of hybrid inter-viewpoint aggregation module. The running time is tested with an input size of $256 \times 256$.

**6. Conclusion**

We presented NeRFLiX, a general NeRF-agnostic restoration paradigm for high-quality neural view synthesis. We systematically analyzed the NeRF rendering pipeline and introduced the concept of NeRF-style degradations. Towards to eliminate NeRF-style artifacts, we presented a novel NeRF-style degradation simulator and constructed a large-scale simulated dataset. Benefiting from our simulated dataset, we demonstrated how SOTA deep neural networks could be trained for NeRF artifact removal. To further restore missing details of NeRF rendered frames, we proposed an inter-viewpoint mixer that is capable of aggregating multi-view frames captured from free viewpoints. Additionally, we developed a view selection scheme for choosing the most pertinent reference frames, largely alleviating the computing burden while achieving superior results. Extensive experiments have verified the effectiveness of our NeRFLiX. Code will be made publicly available.

**Acknowledgments.** This work is partially supported by Shenzhen Science and Technology Program KQTD20210811090149095, the Basic Research Project No. HZQB-KCZYX-2021067 of Hetao Shenzhen-HK S&T Cooperation Zone. It was also supported in part by NSFC-62172348 and Shenzhen General Project with No. JCYJ20220530143604010, the National Key R&D Program of China with grant No. 2018YFB1800800, by Shenzhen Outstanding Talents Training Fund 202002, by Guangdong Research Projects No. 2017ZT07X152 and No. 2019CX01X104, by the Guangdong Provincial Key Laboratory of Future Networks of Intelligence (Grant No. 2022B1212010001), and by Shenzhen Key Laboratory of Big Data and Artificial Intelligence (Grant No. ZDSYS201707251409055).
References


Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Ben Mildenhall, Mehdi S. M. Sajjadi, Andreas Geiger, and Noha Radwan. 


Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. 

Daniel Rebain, Wei Jiang, Soroosh Yazdani, Ke Li, Kwang Moo Yi, and Andrea Tagliasacchi. 

Christian Reiser, Songyou Peng, Yiyi Liao, and Andreas Geiger. 

Mohammed Suhail, Carlos Esteves, Leonid Sigal, and Ameesh Makadia. 

Cheng Sun, Min Sun, and Hwann-Tzong Chen. 

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 


Zachary Teed and Jia Deng. 

Yapeng Tian, Yulun Zhang, Yun Fu, and Chenliang Xu. 

Laurens Van der Maaten and Geoffrey Hinton. 


