Instant Neural Radiance Fields

Thomas Müller*
NVIDIA
Zürich, Switzerland

Alex Evans*
NVIDIA
London, United Kingdom

Isaac Deutsch
NVIDIA
Zürich, Switzerland

András Bódis-Szomorú
NVIDIA
Zürich, Switzerland

Michael Shelley
NVIDIA
Munich, Germany

Christoph Schied
NVIDIA
Seattle, USA

Marco Foco
NVIDIA
Zürich, Switzerland

Alexander Keller
NVIDIA
Berlin, Germany

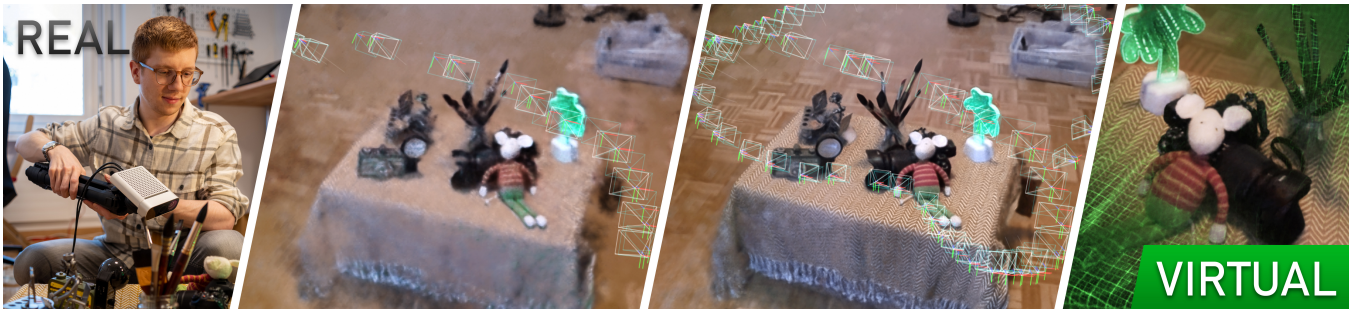

Figure 1: A static scene in the real world can be rapidly captured in 3D into a NeRF, in this example using a hand-held Microsoft Azure Kinect color & depth camera. A realtime SLAM implementation estimates camera poses, while instant training and rendering of the NeRF provides live feedback to guide the capture of areas that have not yet been covered. The whole process takes less than one minute.

ABSTRACT

We extend our instant NeRF implementation [Müller et al. 2022] to allow training from an incremental stream of images and camera poses, provided by a realtime Simultaneous Localization And Mapping (SLAM) system. Camera poses are refined end-to-end by back-propagating the gradients from NeRF training. Reconstruction quality is further improved by compensating for various camera properties, such as rolling shutter, non-linear lens distortion, and variable exposure typical of digital cameras.

Static scenes can be scanned, the NeRF model trained, and the reconstruction verified in an interactive fashion, in under a minute.

ACM Reference Format:
Thomas Müller, Alex Evans, Isaac Deutsch, András Bódis-Szomorú, Michael Shelley, Christoph Schied, Marco Foco, and Alexander Keller. 2022. Instant Neural Radiance Fields. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Real-Time Live! (SIGGRAPH '22 Real-Time Live!), August 07-11, 2022.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3532833.3538678

*Joint first authors

SIGGRAPH '22 Real-Time Live! , August 07-11, 2022, Vancouver, BC, Canada

1 INTRODUCTION

Photogrammetry allows real-world scenes to be reconstructed in 3D. Prior works which output surface representations [Izadi et al. 2011] struggle with the view-dependent appearance of real materials, and smooth out volumetric details such as hair or fur. These limitations have largely been overcome by neural radiance and density fields (NeRF) [Mildenhall et al. 2020], in which a small neural network configured as a *coordinate network* is trained to predict the view-dependent color and density at any point within a volume. Instant NeRF [Müller et al. 2022] can train NeRFs in seconds, which we extend to an end-to-end real-time 3D reconstruction pipeline.

2 METHOD

By implementing NeRF in terms of (i) optimized ray marching routines, (ii) fully fused neural networks [Müller et al. 2021], and (iii) a multiresolution hash encoding [Müller et al. 2022], we can train it 1000× faster than prior work, leading to a high quality reconstruction in a few seconds.

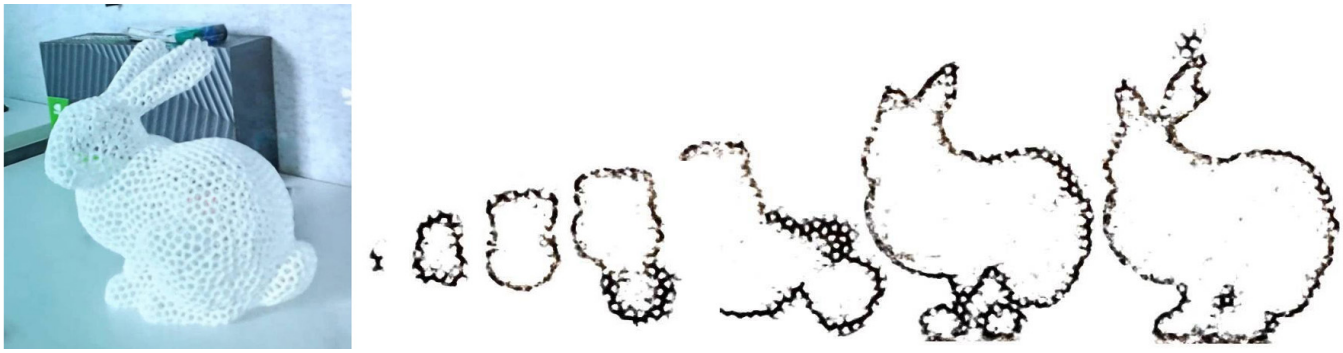

Figure 2: A rendering of the NeRF reconstruction of a 3D-printed bunny is shown on the left, and a number of slices through the NeRF density volume is shown on the right. Notice in the cross sections that even the opposite side of the bunny is reconstructed, despite that it was only visible through the front holes in the input.

Typically, a large number of images with associated camera parameters are processed all at once. In contrast, NeRF can also be trained incrementally, with new images being added to the training set as they arrive from a hand-held camera.

2.1 Realtime visual odometry

Initial camera poses from color and depth data are estimated by frame-to-model tracking, where the model is estimated using a surfel-based SLAM method inspired by [Keller et al. 2013]. Depth input improves the stability and performance of SLAM.

Other camera tracking, visual / inertial odometry or SLAM pipelines that provide approximate camera extrinsics and intrinsics could also be used.

2.2 Camera pose optimisation

As shown in Figure 1, the camera poses can be visualized during capture. The gradients from NeRF training are used to refine these poses end-to-end, similarly to [Lin et al. 2021]. Rather than reparameterizing the camera rotations using matrix logarithms or screw transforms, we follow an approach inspired by rigid body simulation, and treat the loss gradients at each point as if they were forces and torques acting on a mass centered at the camera. A variant of the Adam [Kingma and Ba 2014] optimizer, specialised for rotations, "nudges" the cameras into place, improving reconstruction quality. A weak L2 regularization (factor 10^{-6}) pulls the cameras back towards the initial poses, improving stability.

2.3 Exposure optimisation

To compensate for auto-exposure of the camera given in some capture scenarios, which has a detrimental effect on the reconstruction, we associate a trainable exposure compensation value $e_i \in \mathbb{R}^3$ with each training image, which scales the brightness of the i-th image by $\exp(e_i)$. Since $e_i \in \mathbb{R}^3$, this also corrects small whitepoint shifts.

3 RESULTS AND DISCUSSION

Figure 2 shows a 3D-printed bunny reconstructed using our incremental SLAM + NeRF pipeline, using data from a Microsoft Azure Kinect camera. The shape of the latticework is faithfully reconstructed, even in the partially occluded regions on the opposite side of the bunny.

All results were computed using a single RTX 3090 GPU, rendering at 30–60 frames per second at a resolution of 640×360–960×540, depending on scene, upsampled using NVIDIA DLSS to 1920×1080. The training images were captured over a period of around 20 seconds, and the NeRF allowed to train for another 20–40 seconds.

We have shown that the combination of interactive NeRF training from an incrementally delivered stream of training images, annotated with poses from a SLAM implementation, yields an end-to-end system that permits high quality lightfield capture in less than one minute.

REFERENCES

Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. 2011. KinectFusion: Real-Time 3D Reconstruction and Interaction Using a Moving Depth Camera. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) *(UIST '11)*. Association for Computing Machinery, New York, NY, USA, 559–568. https://doi.org/10.1145/2047196.2047270

Maik Keller, Damien Lefloch, Martin Lambers, Shahram Izadi, Tim Weyrich, and Andreas Kolb. 2013. Real-time 3d reconstruction in dynamic scenes using point-based fusion. In *2013 International Conference on 3D Vision-3DV 2013*. IEEE, 1–8.

Diederik P. Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980* (June 2014).

Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. 2021. BARF: Bundle-Adjusting Neural Radiance Fields. *CoRR* abs/2104.06405 (2021). arXiv:2104.06405 https://arxiv.org/abs/2104.06405

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *ECCV*.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding. *ACM Trans. Graph.* 41, 4, Article 102 (July 2022), 15 pages. https://doi.org/10.1145/3528223.3530127

Thomas Müller, Fabrice Rousselle, Jan Novák, and Alexander Keller. 2021. Real-time Neural Radiance Caching for Path Tracing. *ACM Trans. Graph.* 40, 4, Article 36 (Aug. 2021), 16 pages. https://doi.org/10.1145/3450626.3459812