Deferred Neural Rendering for View Extrapolation

Tobias Bertel, University of Bath
Yusuke Tomoto, Fyusion Inc.
Srinivas Rao, Fyusion Inc.
Rodrigo Ortiz-Cayon, Fyusion Inc.
Stefan Holzer, Fyusion Inc.
Christian Richardt, University of Bath

Motivation

Neural scene representations have shown great potential for high-quality view synthesis of casually captured real-world environments.

Training is usually done on a dense training corpus and the models are usually only used for interpolation.

Problem

1. The need for a dense training corpus can make the capturing procedure tedious and time-consuming.

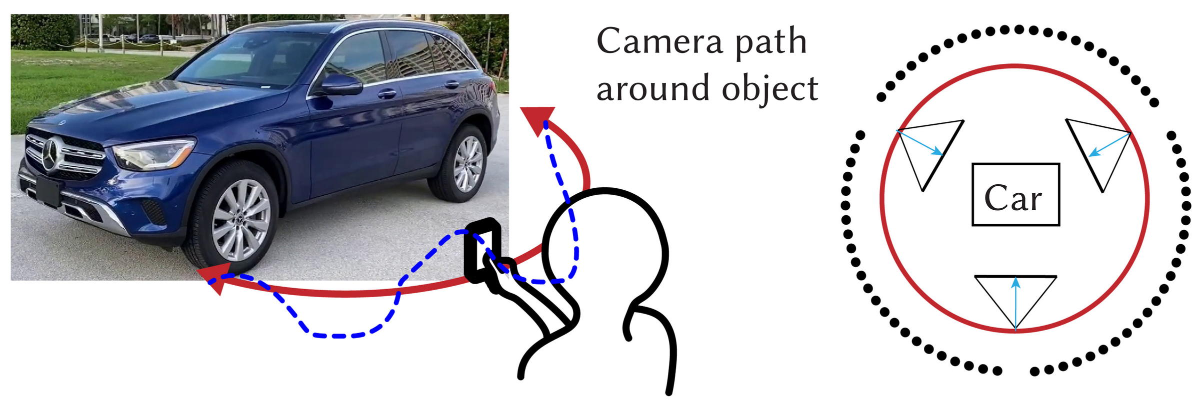
2. Loss functions are applied to individual input viewpoints leading to artefacts when trained on a sparse training corpus.

3. State-of-the-art in terms of visual quality, i.e. volumetric approaches like NeRF, are not suitable for interactive applications.

Our Approach

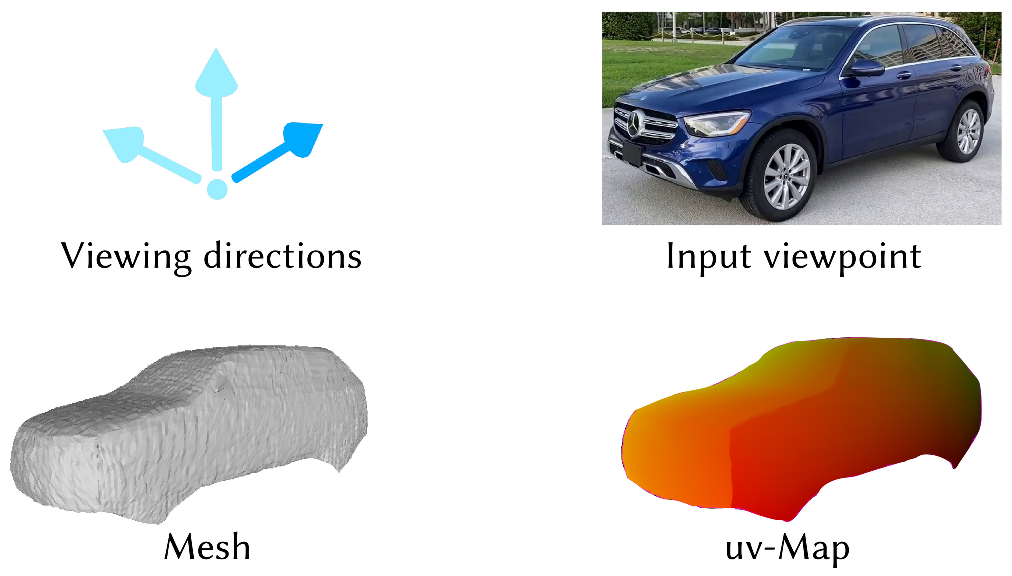
We present extensions to Deferred Neural Rendering (DNR) a method based on a Generative Adversarial Network (GAN) with the goal to extrapolate a casually captured and sparse training corpus.

Capture

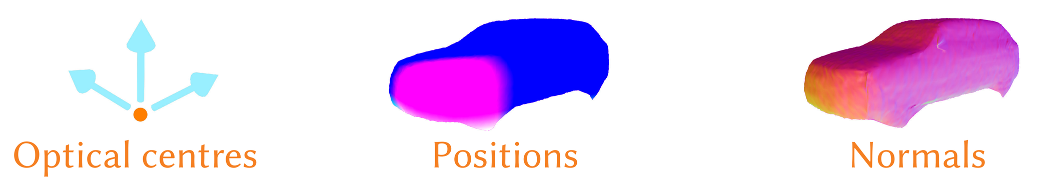


We use an iPhone X to capture a video while walking around a shiny object. We use an internal pipeline to obtain proxy geometry and uv map of the car.

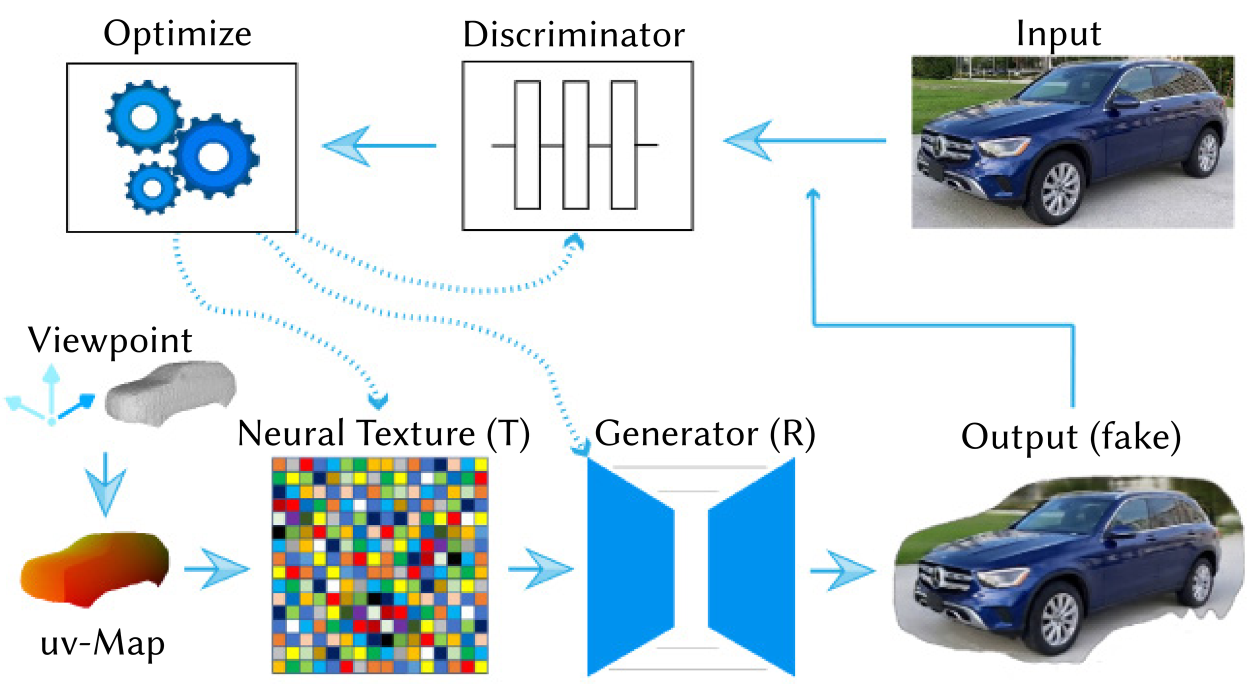
Dataset



The baseline dataset consists of: a set of posed viewpoints, a proxy geometry, and a corresponding uv map.



We propose a number of extensions to improve the model performance, i.e. adding guides to the generator input and injecting noise to the viewing direction during training.



Deferred Neural Rendering [Thies et al. 2019]

Architecture (GAN)

Generator/Renderer: 5-layer U-Net with skip connections.
Input: view-dependent neural features and background image.

Discriminator: 3-layered patch-GAN.
Input: generated or rendered (fake) image.

$$\mathcal{T}^*, \mathcal{R}^* = \underset{\mathcal{T},\mathcal{R}}{\arg\min} \sum_{d \in \mathcal{D}} \mathcal{L}(A(d) \mid F_d(\mathcal{T}), G_d(\mathcal{R})).$$

Neural texture can be seen as a learned surface light field.

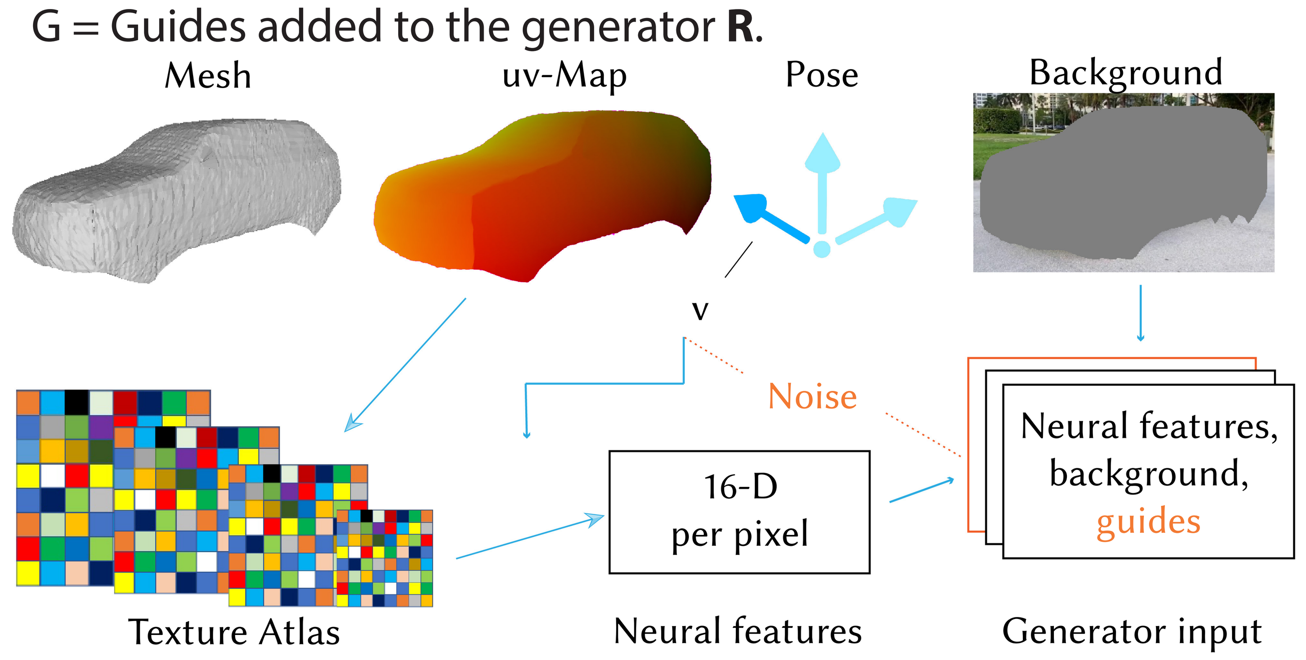
Objective function

L = loss (BCE + SSIM for generator, BCE for discriminator),

A = Augmentation operator when fetching data items **d**,

F = view-dependent lookup into neural texture **T**,

G = Guides added to the generator **R**.



Extensions

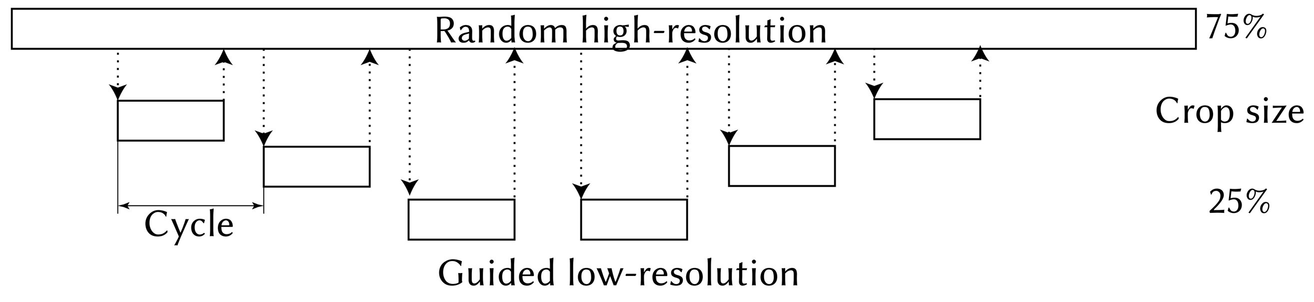
1. Augmentation strategy that guides to poorly inferred image regions during training.

2. Several guides are added to the generator input, i.e. optical centres of the viewpoints, screen space positions and normals.

3. We add noise to the viewing direction **v** and add noise to the additional guides.

4. Multi-stage training to speed up convergence

Training



Training is split into several stages following a pyramidal scheme.

The model is initialised with high-resolution crops that are randomly chosen as demonstrated in the baseline over 10 epochs.

The resolution is successively reduced and we apply our extensions in a cyclic manner.

The guided augmentation is updated at the beginning of each cycle. For every dataset item **d,** a list of crops is created that is sorted according to the prediction or reconstruction error.

Within each cycle, noise decays in the first 33-50% and we switch back to the baseline augmentation in the last 10%.

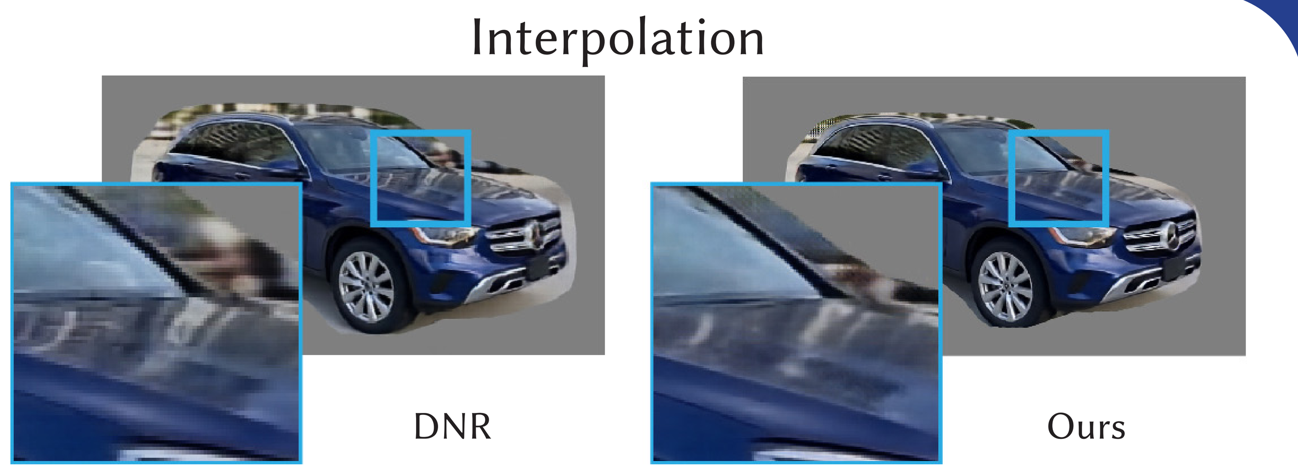
The process bottoms up after 3-5 levels of refinement and ends baseline augmentation for the last 10% of training epochs.

Testing

Full resolution output images, used for guided augmentation and showing final results, are obtained by tesing the model without using any injected noise or crop augmentation.

Results

Extrapolation: Our extensions enable smoother extrapolated viewpoints than DNR. Deviating from the training corpus still leads to disturbing artefacts increasing the farther away we go.



Interpolation: The reflection in our method is slightly blurred compared to the baseline because of the injected noise during training. Guided augmentation increases the grey area of the inpainted background.

We provide results for camera path stabilisation in the submission video.

Limitations

The viewpoint generation is based on whole images or subsets (crops) of it. A strategy motivated by a camera model seems reasonable for a generalised image formation.

Imperfect proxies cause various artefacts: (1) the proxy is too big and the generator must remove geometry and inpaint background. (2) the proxy is too small and the generator inpaints object-appearance in the background.

Extrapolated results show flickering that increase the farther viewpoints are away from the training corpus.

Appearance cannot be edited directly which is a severe limitation of the current representation.

References

P. Hedman, Julien Philipp, True Price, Jan-Michael Frahm, George Drettakis, and Gabriel Brostow. 2018. Deep Blending for Free-viewpoint Image-based Rendering. SIGGRAPH.

Abe Davis, Marc Levoy, and Fredo Durand. 2012. Unstructured Light Fields. CGF.

Philipp Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. CVPR.

Jeong Joon Park, Aleksander Holynski, and Steve Seitz. 2020. Seeing the World in a Bag of Chips. CVPR.

Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. 2020. Neural Sparse Voxel Fields. arXiv:2007.11571

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

Tobias Ritschel, Carsten Dachsbacher, Thorsten Grosch, and Jan Kautz. 2012. The State of the Art in Interactive Global Illumination. CGF.

Justus Thies, Michael Zollhöfer, and Matthias Niessner. 2019. Deferred Neural Render-ing: Image Synthesis Using Neural Textures. SIGGRAPH.

Acknowledgements

We thank Justus Thies for fruitful discussions and sharing code for DNR. This work was supported by EU Horizon 2020 MSCA grant FIRE (665992), the EPSRC Centre for Doctoral Training in Digital Entertainment (EP/L016540/1), RCUK grant CAMERA (EP/M023281/1) and a Rabin Ezra Scholarship.