Neural Implicit Surface Reconstruction from Noisy Camera Observations

Sarthak Gupta[1, 2], Patrik Huber [2]

[1] Indian Institute of Technology Roorkee, Roorkee, Uttarakhand, India - 247667

[2] University of York, Deramore Lane, Heslington, York, YO10 5GH, United Kingdom

mrsarthakgupta@gmail.com, patrik.huber@york.ac.uk

Abstract

Representing 3D objects and scenes with neural radiance fields has become very popular over the last years. Recently, surface-based representations have been proposed, that allow to reconstruct 3D objects from simple photographs. However, most current techniques require an accurate camera calibration, i.e. camera parameters corresponding to each image, which is often a difficult task to do in real-life situations. To this end, we propose a method for learning 3D surfaces from noisy camera parameters. We show that we can learn camera parameters together with learning the surface representation, and demonstrate good quality 3D surface reconstruction even with noisy camera observations.

Introduction

Representing 3D objects and scenes with neural networks has gained significant traction recently. In [1], *NeRF* is proposed, Neural Radiance Fields, where a neural network is used for volumetric representation of the scene. However, a volumetric representation is not the best representation in many cases; many objects like faces are better represented using surfaces. For this, [2], called *NeuS*, proposed to use neural implicit surfaces together with volume rendering for multi-view reconstruction. [3] addressed another shortcoming of NeRF method to work on image data when camera calibration data is absent. It jointly estimates the scene representation and optimises for the camera parameters. We propose to marry the benefits of each of these approaches: We propose a method to learn a neural implicit surface based representation of objects from noisy camera observations. We show that the classical NeuS method fails to learn an object completely if camera parameters are not precise, whereas our approach succeeds.

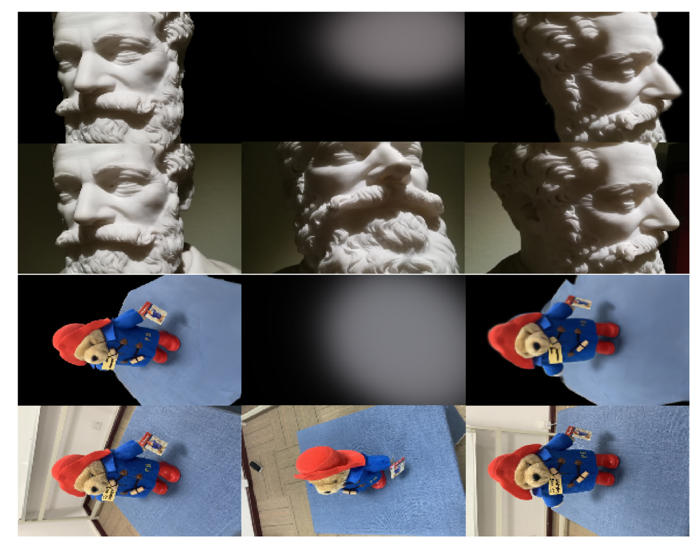

Figure 1: *(Left)*: [2] with ground truth camera parameters. *(Centre)*: [2] with noisy camera parameters. *(Right)*: Our approach with noisy camera parameters. The lower images represent the actual image of the object while the upper image is the rendered image of the reconstructed surface.

Important Result

Our method outperforms the existing method by being able to learn an equal or more accurate object representation, even in the presence of significant noise in the camera parameters.

Methodology

We begin with [3], where a NeRF are made learnable to converge to values that result in desirable reconstructions. While [3] works with unknown camera parameters for only forward-facing input images with rotational and translational perturbations of up to ±20°, our approach works successfully on images from 360° view angles. The latter results in a much harder problem that is prone to local optima that do not produce good results when learning camera parameters from scratch. The implicit surface based network of [2], *NeuS*, consists of two MLPs to encode a signed distance function (SDF) and colour, respectively. We add two modules that make both the extrinsic and intrinsic camera parameters learnable.

Methodology(continued)

The extrinsic parameters are expressed as a 4×4 camera-to-world space transformation matrix with $T_{wc} = [R|t]$ where $R \in SO(3)$ and $t \in \mathbb{R}^3$ denote the camera rotation and translation, respectively. We take the sensor's centre as the principal point, and we assume that the same camera takes all input images. Thus, estimating the camera intrinsics simplifies to finding the focal length.

The network, including the camera parameters, is trained on images of a scene using a weighted linear combination of Eikonal-loss [4], colour-loss, and mask-loss masking out the background.

Results

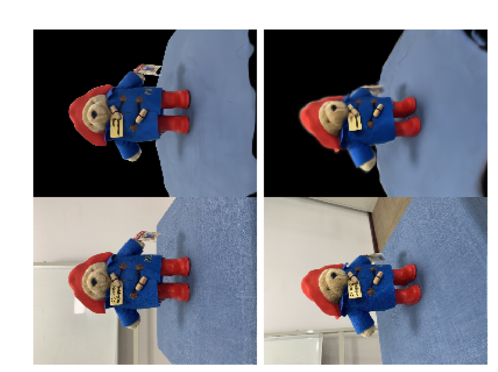

Figure 2: *(left)*: Vanilla NeuS with ground-truth camera parameters (*baseline-gt*). *(right)*: Our method, with very noisy initialisation (*learnable-noisy*). The top shows the neural rendering, the bottom shows the ground truth.

In noisy camera parameter setting, our method produces reconstructions visually indistinguishable from baseline NeuS and produces a final reconstruction loss and Peak Signal-to-Noise Ratio (PSNR) similar to the baseline with an equal number of iterations.

Conclusion

While comparing our work to the current 3D surface reconstruction state-of-the-art in a noisy parameter setting, we have found that the existing method fails as soon as camera parameters are not accurate. Our method outperforms the existing method by being able to learn an equal or more accurate object representation, even in the presence of significant noise in the camera parameters. Our work broadens the use cases of neural implicit surface based object reconstruction, by removing the need for accurate camera calibration information, and increasing the robustness to errors.

However, there is still much scope for work in this field, as reconstructing the surface from completely unknown camera parameters is still an open problem for 360° view angles. To tackle this, we would like to investigate an approach to learning surface reconstruction from unknown camera parameters. Furthermore, we plan to apply our method to a multi-view reconstruction benchmark, where 3D shape accuracy is evaluated.

References

[1] 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 *ECCV (1)*, LNCS. Springer, 2020.

[2] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *NeurIPS*, 2021.

[3] Zirui Wang, Shangzhe Wu, Weidi Xie, Min Chen, and Victor Adrian Prisacariu. Nerf--: Neural radiance fields without known camera parameters. *CoRR*, abs/2102.07064, 2021.

[4] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *ICML*, 2020.