УДК 004.896 + 007.52

A REVIEW OF NEURAL RADIANCE FIELD FOR 3D RECONSTRUCTION TASK

Mohrat M. (ITMO University)
Scientific supervisor – professor, doctor of technical sciences, Kolyubin S. A. (ITMO University)

Abstract. In this work, we provide an in-depth overview of the Neural Radiance Fields NeRF method, its theoretical principles, its key features, and advantages, along with an analysis of the recent results achieved in this field by analyzing three different algorithms in this field. We also explore potential applications of NeRF in various disciplines, including 3D reconstruction, virtual and augmented reality, autonomous navigation, and robotics.

Introduction. NeRF is a new novel view synthesis with implicit scene representation, which has recently become popular method for representing 3D geometry using deep neural networks. NeRF was firstly proposed in a paper by Mildenhall et al. [1], and has obtained a high interest due to its ability to produce high quality 3D reconstructions from a small number of RGB images with their positions. As a novel view synthesis and 3D reconstruction method, NeRF models find applications in robotics, urban mapping, autonomous navigation, virtual/augmented reality, and more.

Main part. The basic idea behind NeRF is to synthesize views of scenes by optimizing a continuous volumetric function of the scene using a set of input views. The algorithm presented in [1] represents the scenes using a multilayer perceptron MLP deep neural network, whose input is a continuous 5D coordinate namely spatial location (x, y, z) and viewing direction (θ, φ) and whose output is the volume density (color) and view-dependent emitted radiance (opacity) at that spatial location.

Technically speaking, the continuous function that represent the 3D scene is called the radiance field. This radiance field describes the appearance and geometry of the scene, and can be trained using a neural network. During training, the network takes in a 3D point in space and outputs the color and opacity of that point, as seen from a given camera position.

The neural network is trained using a volume rendering loss as it's differentiable. This loss compares the output of the network to the ground truth images. The network tries to minimize the difference between the rendered images and the ground truth images, by adjusting the parameters of the radiance field. Once the network is trained, it can be used to synthesize views of the scene from any viewpoint, including views that were not seen during training.

To have a 3D reconstruction based on the NeRF optimization method, a process of overfitting should be done over a set of RGB images with their positions. The input data for the algorithm are the color RGB images, as well as the corresponding camera positions. The output data is a 3D synthesized model with detailed geometry of the scene or object.

An additional input for the algorithm can be a sparse depth map, which can be generated by Structure from Motion (SfM) algorithms as presented in [2], where the authors used COLMAP [3] to extract a sparse point cloud, which was processed by a depth completion network to produce depth and uncertainty maps. In addition to the standard volumetric rendering loss, the authors introduced a depth loss based on predicted depth and uncertainty.

Another algorithm presented in [4] uses multi-view images in their NeRF model focused on depth reconstruction. In [4], COLMAP was used to extract sparse depth priors in the form of a point cloud. This was then fed into a pretrained (fine-tuned on the scene) monocular depth network to extract a depth map prior. This depth map prior was used to supervise volume sampling by only allowing sampling points at the appropriate depth.

Conclusion. To sum up, this work presents an overview of NeRF method as a promising way for 3D reconstruction that has shown significant potential in many applications by analyzing three state-of-the-art algorithms related to this method. This new trend shows the ability to handle complex geometry views, as well as its ability to produce highly realistic and detailed reconstructions, which makes it an important solution for tasks related to computer vision, 3D mapping and autonomous navigation.

References.

1. Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM 65.1 (2021): 99-106.
2. Roessle, Barbara, et al. "Dense depth priors for neural radiance fields from sparse input views." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
3. Schonberger, Johannes L., and Jan-Michael Frahm. "Structure-from-motion revisited." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
4. Wei, Yi, et al. "Nerfingmvs: Guided optimization of neural radiance fields for indoor multi-view stereo." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

Mohrat M. (author) signature

Kolyubin S. A. (scientific supervisor) signature