NeRFocus: Bringing Lightweight Focus Control to Neural Radiance Fields

Originally published at https://www.unite.ai/nerfocus-bringing-lightweight-focus-control-to-neural-radiance-fields/

New research from China offers a method to achieve affordable control over depth of field effects for Neural Radiance Fields (NeRF), allowing the end user to rack focus and dynamically change the configuration of the virtual lens in the rendering space.

Titled *NeRFocus*, the technique implements a novel 'thin lens imaging' approach to focus traversal, and innovates *P-training*, a probabilistic training strategy that obviates the need for dedicated depth-of-field datasets, and simplifies a focus-enabled training workflow.

CLICK TO VIEW ANIMATED GIF

The paper is titled *NeRFocus: Neural Radiance Field for 3D Synthetic Defocus*, and comes from four researchers from the Shenzhen Graduate School at Peking University, and the Peng Cheng Laboratory at Shenzhen, a Guangdong Provincial Government-funded institute.

Addressing the Foveated Locus of Attention in NeRF

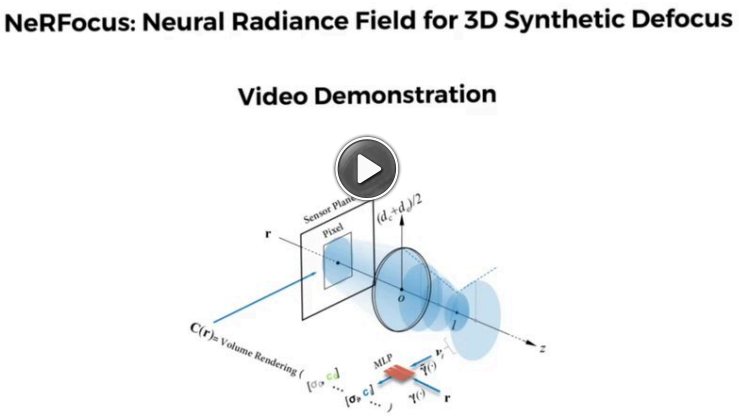
If NeRF is ever to take its place as a valid driving technology for virtual and augmented reality, it's going to need a lightweight method of allowing realistic foveated rendering, where the majority of rendering resources accrete around the user's gaze, rather than being indiscriminately distributed at lower resolution across the entire available visual space.

From the 2021 paper Foveated Neural Radiance Fields for Real-Time and Egocentric Virtual Reality, we see the attention locus in a novel foveated rendering scheme for NeRF. Source: https://arxiv.org/pdf/2103.16365.pdf

An essential part of the authenticity of future deployments of egocentric NeRF will be the system's ability to reflect the human eye's own capacity to switch focus across a receding plane of perspective (see first image above).

This gradient of focus is also a perceptual indicator of the scale of the scene; the view from a helicopter flying over a city will have zero navigable fields of focus, because the entire scene exists beyond the viewer's outermost focusing capacity, while scrutiny of a miniature or 'near field' scene will not only allow 'focus racking', but should, for realism's sake, contain a narrow depth of field by default.

Below is a video demonstrating the initial capabilities of NeRFocus, supplied to us by the paper's corresponding author:



CLICK TO PLAY VIDEO ON SITE

Beyond Restricted Focal Planes

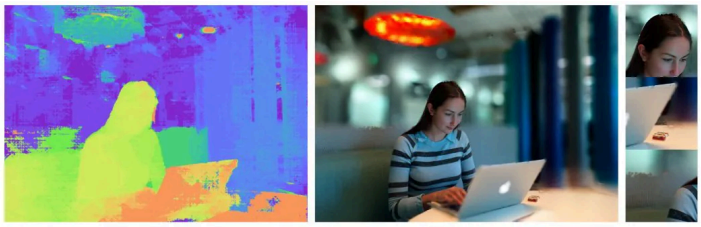
Aware of the requirements for focus control, a number of NeRF projects in recent years have made provision for it, though all the attempts to date are effectively sleight-of-hand workarounds of some kind, or else entail notable post-processing routines that make them unlikely contributions to the real-time environments ultimately envisaged for Neural Radiance Fields technologies.

Synthetic focal control in neural rendering frameworks has been attempted by various methods in the past 5-6 years – for instance, by using a segmentation network to fence off the foreground and background data, and then to generically defocus the background – a common solution for simple two-plane focus effects.

From the paper 'Automatic Portrait Segmentation for Image Stylization', a mundane, animation-style separation of focal planes. Source: https://jiaya.me/papers/po

Multiplane representations add a few virtual 'animation cels' to this paradigm, for instance by using depth estimation to cut the scene up into a choppy but manageable gradient of distinct focal planes, and then orchestrating depth-dependent kernels to synthesize blur.

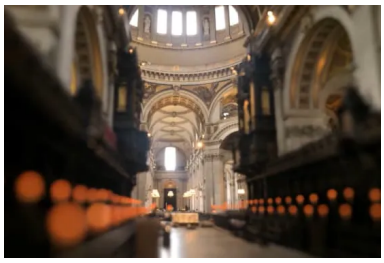
Additionally, and highly relevant to potential AR/VR environments, the disparity between the two viewpoints of a stereo camera setup can be utilized as a depth proxy – a method proposed by Google Research in 2015.



From the Google-led paper Fast Bilateral-Space Stereo for Synthetic Defocus, the difference between two viewpoints provides a depth map that can facilitate blur However, this approach is inauthentic in the situation envisaged above, where the photo is clearly taken with a 35-50mm (SLR standard) lens, but the extreme def of the background would only ever occur with a lens exceeding 200mm, which has the kind of highly constrained focal plane that produces narrow depth of field in human-sized environments. Source

Approaches of this nature tend to demonstrate edge artifacts, since they attempt to represent two distinct and edge-limited spheres of focus as a continual focal gradient.

In 2021 the RawNeRF initiative offered High Dynamic Range (HDR) functionality, with greater control over low-light situations, and an apparently impressive capacity to rack focus:



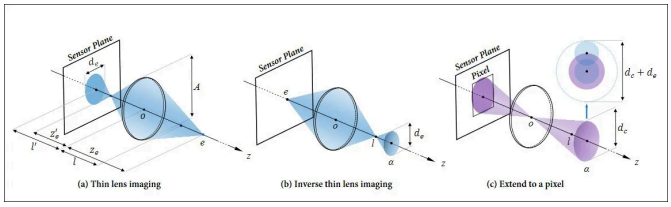
RawNeRF racks focus beautifully (if, in this case, inauthentically, due to unrealistic focal planes), but comes at a high computing cost. Source: https://bmild.github.io/rawnerf/

However, RawNeRF requires burdensome precomputation for its multiplane representations of the trained NeRF, resulting in a workflow that can't be easily adapted to lighter or lower-latency implementations of NeRF.

Modeling a Virtual Lens

NeRF itself is predicated on the pinhole imaging model, which renders the entire scene sharply in a manner similar to a default CGI scene (prior to the various approaches that render blur as a post-processing or innate effect based on depth of field).

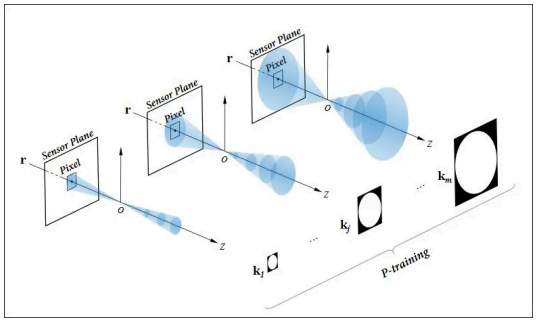
NeRFocus creates a virtual 'thin lens' (rather than a 'glassless' aperture) which calculates the beam path of each incoming pixel and renders it directly, effectively inverting the standard image capture process, which operates *post facto* on light input that has already been affected by the refractive properties of the lens design.



This model introduces a range of possibilities for content rendering inside the frustum (the largest circle of influence depicted in the image above).

Calculating the correct color and density for each multilayer perceptron (MLP) in this broader range of possibilities is an additional task. This has been solved before by applying supervised training to a high number of DLSR images, entailing the creation of additional datasets for a probabilistic training workflow – effectively involving the laborious preparation and storage of multiple possible computed resources that may or may not be needed.

NeRFocus overcomes this by *P-training*, where training datasets are generated based on basic blur operations. Thus, the model is formed with blur operations innate and navigable.



Aperture diameter is set to zero during training, and predefined probabilities used to choose a blur kernel at random. This obtained diameter is used to scale up each composite cone's diameters, letting the MLP accurately predict the radiance and density of the frustums (the wide circles in the above images, representing the maximum zone of transformation for each pixel)

The authors of the new paper observe that NeRFocus is potentially compatible with the HDR-driven approach of RawNeRF, which could potentially help in the rendering of certain challenging sections, such as defocused specular highlights, and many of the other computationally-intense effects which have challenged CGI workflows for thirty or more years.

The process does not entail additional requirements for time and/or parameters in comparison to prior approaches such as core NeRF and Mip-NeRF (and, presumably Mip-NeRF 360, though this is not addressed in the paper), and is applicable as a general extension to the central methodology of neural radiance fields.

First published 12th March 2022.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More