NeRFocus: Neural Radiance Field for 3D Synthetic Defocus

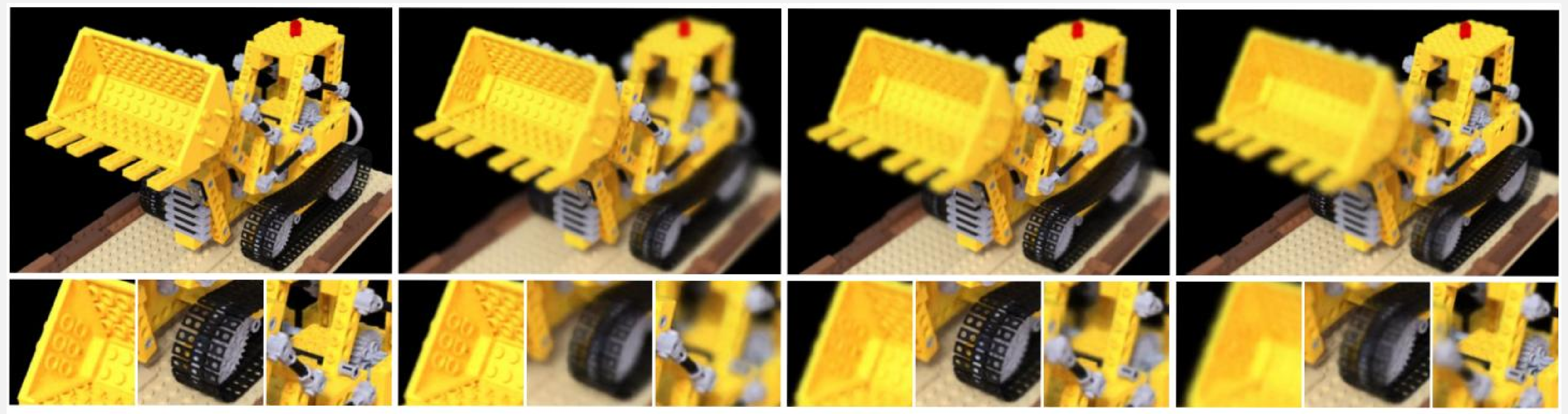
Yinhuai Wang†[1], Shuzhou Yang†[1,2], Yujie Hu[1], Xinhua Cheng[1], Jian Zhang*[1,2]

[1] Peking University Shenzhen Graduate School, [2] Peng Cheng Laboratory

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Task: Defocus Novel View Synthesis

- **Novel View Synthesis**: Predict the unseen novel view of a given scene.
- **Defocus Novel View Synthesis:** Simulate the effect of thin lens imaging, realize defocus effect in novel view (**New Task We Proposed**)



(a) A=0 (b) A=0. 06, l=3.4 (c) A=0.06, l=3.7 (d) A=0.06, l=4.0

Motivation

- **Defocus Effect:** Opportune defocus effect emphasizes the relationship between the subject and the background, and is widely used in photography.
- **Novel View Synthesis:** Low-cost novel view synthesis provides users with a better visual experience.
- **There has been no method to achieve defocus effect in novel view synthesis before!**

Other Information

- **Paper & Code**

PDF: https://arxiv.org/abs/2203.05189

Code: https://github.com/wyhuai/NeRFocus

- **Authors**

Yinhuai Wang†: yinhuai@stu.pku.edu.cn

Shuzhou Yang†: szyang@stu.pku.edu.cn

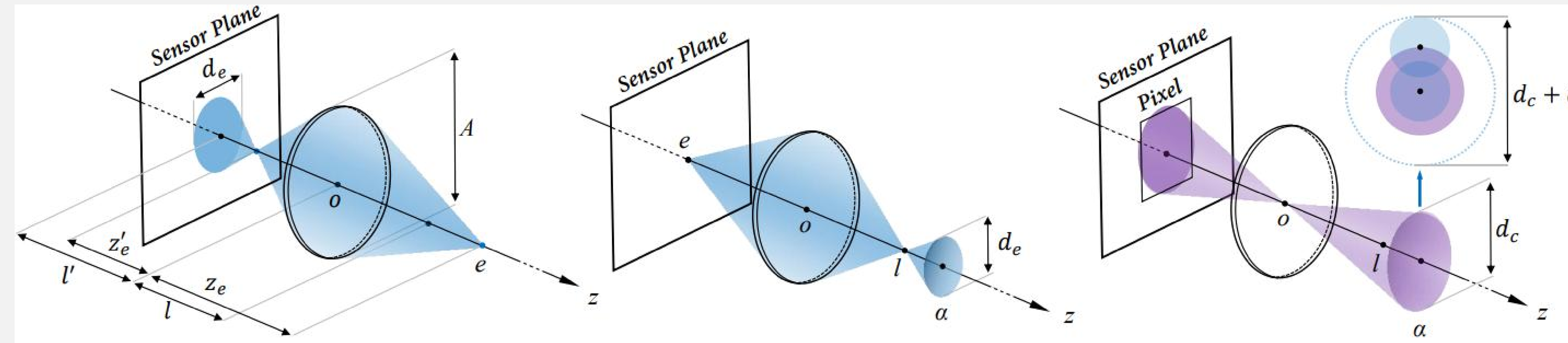
Jian Zhang*: zhangjian.sz@pku.edu.cn

- **Research Group**

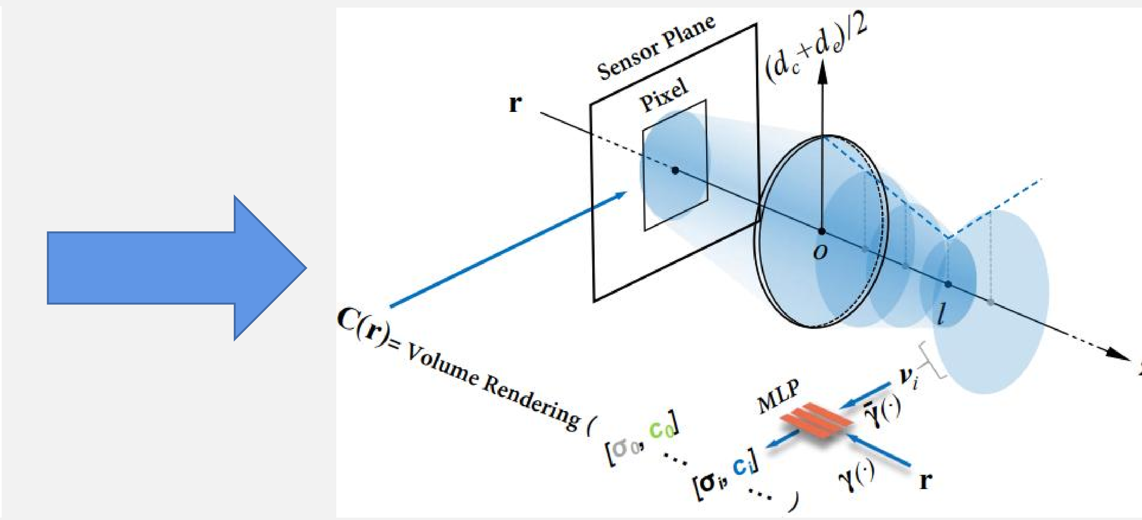
https://villa.jianzhang.tech/

VILLA Visual-Information Intelligent Learning LAB

Method: Approximate Thin Lens Imaging in NeRF & P-Training Strategy



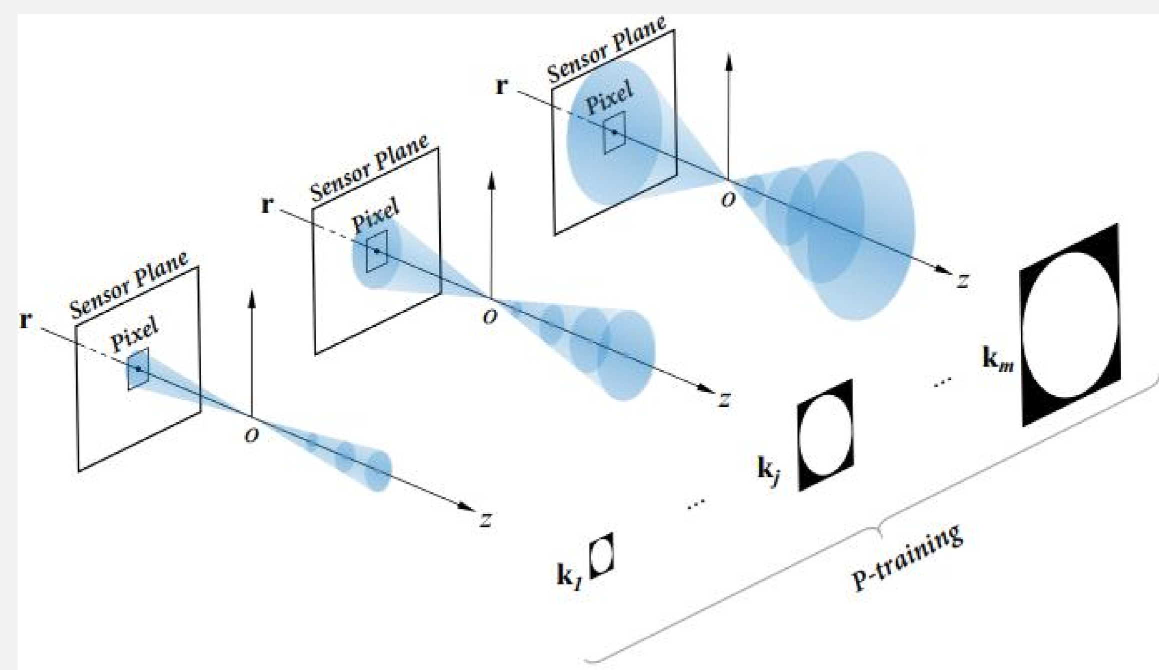
(a) Thin lens imaging (b) Inverse thin lens imaging (c) Extend to a pixel



- **Overview. (a)**: Defocus in thin lens imaging. If a point is away from the focal plane, its imaging on the sensor plane is a circle rather than a point, called the circle of confusion (CoC). **(b):** Inverse the thin lens imaging to form a bicone. α denotes the plane. de denotes the bicone's diameter on plane α. **(c)** Use a circle to represent the receptive field of a pixel. All the axis of the bicones belonging to this pixel forms a cone (purple region), with dc denoting its diameter on plane α.

$$l` = \frac{fl}{l-f} \qquad d_e = \left|\frac{A(l-z_\alpha)}{l}\right| \qquad d_c = \frac{d_0 z_\alpha}{l`} = \frac{d_0 z_\alpha (l-f)}{fl} \qquad d = d_c + d_e = \frac{d_0 z_\alpha (l-f)}{fl} + \left|\frac{A(l-z_\alpha)}{l}\right|$$

- **Rendering Process of a Pixel.** The blue dotted line denotes the radius. We divide the composite cone into frustums, and use the modified integrated positional encoding (IPE) to encode each frustum. Then use an MLP to predict the expected radiance and density of each frustum.



P-Training: The aperture diameter is set to zero during training, so de is equal to zero, the composite cone is degrade as a cone. For each training step, we use predefined probabilities to select a blur kernel kj , j ∈ {1, ..., m} randomly. Then, use the diameter of kj to scale up every composite cone's diameters and use kj to blur the original image as the rendering target. This training process will urge the MLP to correctly predict the radiance and density of frustums in different sizes.

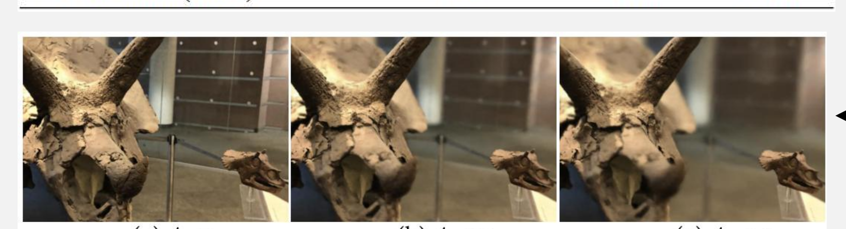
$$\mathbf{D}_j = \{\mathbf{y}_1 * \mathbf{k}_j, \ \ldots, \ \mathbf{y}_u * \mathbf{k}_j\} \ j \in \{1, \ \ldots, \ m\}$$

Experiments

Table 1. Quantitative comparisons on training time, model parameters, and PSNR. For NeRFocus, we set A=0 to calculate the PSNR by the original GT (large DOF). Though NeRFocus achieves defocus effects that mip-NeRF and NeRF cannot do, it neither incurs additional costs on computation and parameters nor sacrifices performance in rendering large DOF images.

Method	Time(hours)↓	Params↓	PSNR↑
NeRF (ECCV'20)	64.8	1192k	29.00
mip-NeRF (ICCV'21)	55.6	613k	28.83
NeRFocus (ours)	55.6	613k	28.26

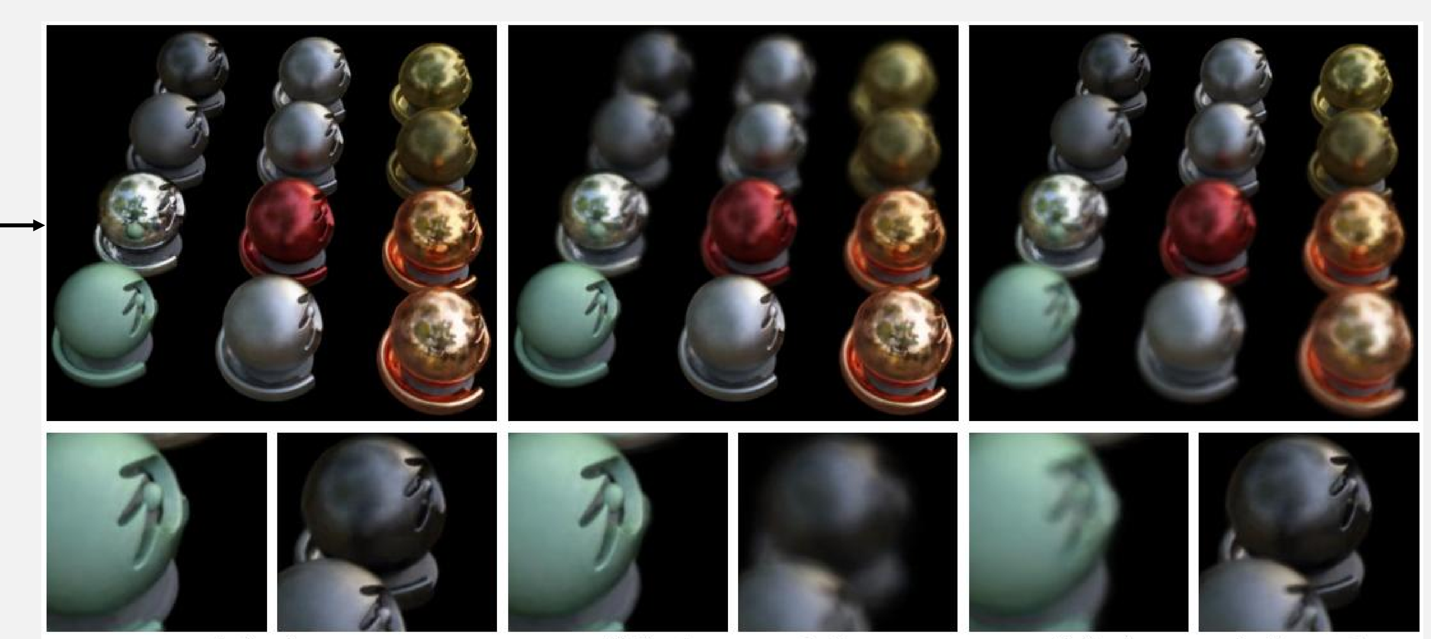
← Performance of Novel View Synthesis



(a) A=0 (b) A=0.1 (c) A=0.2

← Visualization of Defocus Effect (fix the focus distance)

Visualization of Defocus Effect →



(a) A=0 (b) A=0. 06, l=3.3 (c) A=0.06, l=4.6