

Joint Deblurring and 3D Reconstruction for Macrophotography

Yifan Zhao¹ and Liangchen Li¹ and Yuqi Zhou¹ and Kai Wang² and Yan Liang¹ and Juyong Zhang¹ [†]

¹University of Science and Technology of China

²China Unicom

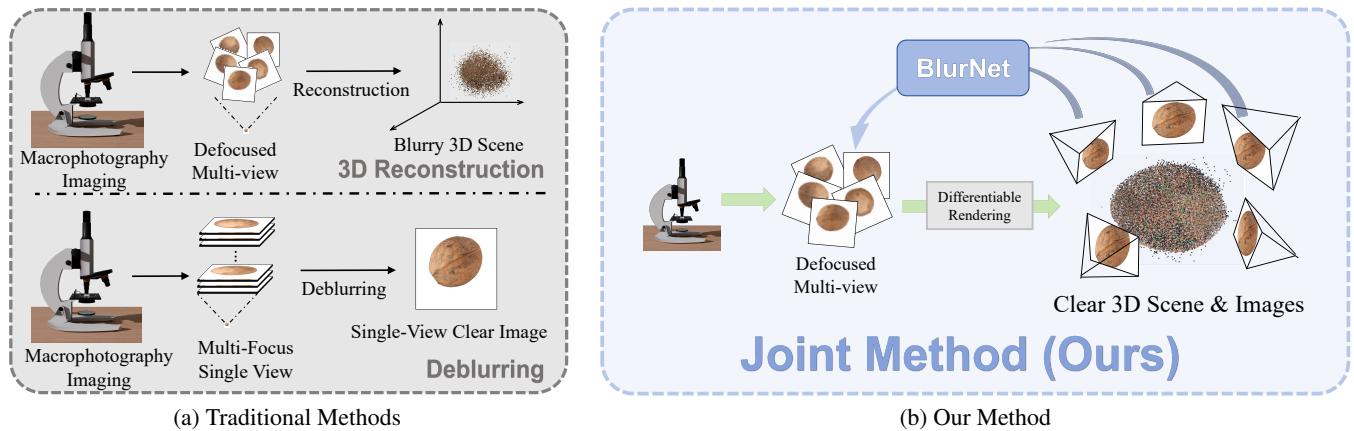


Figure 1: Defocus blur is a long-standing problem in macrophotography. Traditional multi-focus image fusion deblurring methods usually require the acquisition of a large number of input images. Without deblurring, 3D reconstruction of tiny objects can only produce blurred 3D scenes. To address this problem, we propose a method for combining deblurring and 3D reconstruction of macrophotography, which can produce clear 3D scenes and images.

Abstract

Macro lens has the advantages of high resolution and large magnification, and 3D modeling of small and detailed objects can provide richer information. However, defocus blur in macrophotography is a long-standing problem that heavily hinders the clear imaging of the captured objects and high-quality 3D reconstruction of them. Traditional image deblurring methods require a large number of images and annotations, and there is currently no multi-view 3D reconstruction method for macrophotography. In this work, we propose a joint deblurring and 3D reconstruction method for macrophotography. Starting from multi-view blurry images captured, we jointly optimize the clear 3D model of the object and the defocus blur kernel of each pixel. The entire framework adopts a differentiable rendering method to self-supervise the optimization of the 3D model and the defocus blur kernel. Extensive experiments show that from a small number of multi-view images, our proposed method can not only achieve high-quality image deblurring but also recover high-fidelity 3D appearance.

CCS Concepts

- Computing methodologies → Computer graphics; Image processing; 3D imaging;

1. Introduction

Macrophotography [MF21, CSE*22, Mas08] is a photographic technique that uses a specially designed macro lens or even a

low-power microscope to photograph tiny objects, such as insects and mineral particles, at a sub-centimeter size. Unlike microscopy which typically deals with microscopic specimens at micrometer scale, macrophotography focuses on slightly larger subjects (millimeter to centimeter scale) and employs different optical systems. This technique enables researchers and artists to ob-

[†] Corresponding Author

serve and document the subtle world that is difficult to observe with the naked eye but visible in the macro world, creating possibilities for high-fidelity 3D reconstruction of these small objects. Applications of macrophotography span diverse fields such as entomological documentation [Bis17, MMS^{*}22], precision industrial inspection [RMLGARG15, JLH^{*}17], and artistic expression [OP02, PFX^{*}24].

As illustrated in Fig. 2, macrophotography operates in a unique optical regime that distinguishes it from other imaging process. In terms of optical characteristics, macro lenses typically have large focal lengths and operate at close working distances, creating fundamental differences in image formation. Compared to normal photography, macrophotography works with a much smaller subject-to-camera distance and have a much smaller depth of field, magnifying subjects to near life-size or greater on the image sensor. In contrast to microscopy, which utilizes compound optical systems with transmitted light and fixed working distances, macrophotography employs specialized lenses with reflected light and variable working distances. While microscopes achieve very high magnifications (typically 40-1000 \times), macrophotography operates in a lower magnification range (1-5 \times) but offers greater flexibility in composition, lighting control, and subject placement. These distinct optical properties create a specific imaging environment with its own set of technical challenges.

The principal challenge in macrophotography arises from its extremely shallow depth of field. Due to the close focusing distances and relatively large apertures needed for adequate light gathering, even tiny variations in subject depth result in significant focus disparities. This creates a severe technical limitation: it becomes nearly impossible to capture an entire three-dimensional object in sharp focus with a single exposure. This limitation, known as defocus blur, manifests much more prominently in macrophotography than in normal photography and follows different optical patterns than in microscopy. When attempting to create 3D reconstructions from macrophotography images, this defocus blur causes significant image distortion and detail loss, degrading quality and hindering high-fidelity results. Consequently, specialized deblurring techniques designed specifically for macrophotography's unique optical characteristics become essential.

As shown in Fig. 1, a large number of works have been proposed to address the defocus blur problem [ZZY^{*}24, JLD^{*}20, GTS20, ZKC^{*}20, WXY^{*}23, ZWS^{*}23, KSN^{*}21, LHCT20], but they usually require large number of captured images and are only applicable to deblurring in 2D image space. They cannot aggregate information from adjacent views and ensure multi-view consistency. On the other hand, existing 3D scene deblurring methods do not consider the difference between macrophotography images and natural images. To the best of our knowledge, no one has considered the problem of restoring the clear 3D appearance of detailed tiny objects from blurred inputs. Compared with natural scenes, macro scenes are more sensitive to parameters, have shallower depth of field, and larger defocus blur, making reconstructing macro scenes more challenging.

In this work, we propose the first joint self-supervised optimization method that can simultaneously obtain high-quality 3D scene and clear images for subcentimeter tiny objects. We observe that

the essence of imaging is the projection of a clear 3D scene onto the imaging plane, and the depth variation during the projection process produces defocus blur. Therefore, our key idea is to accurately simulate the imaging process of blur caused by depth variation in the macro scene in the reconstructed 3D scene space. As long as we get a blurred image consistent with the input along this process, we can complete a high-fidelity 3D reconstruction of the macro scene through the latest differentiable rendering methods [KKLD23]. This 3D to 2D deblurring method achieves clear 3D scene for multi-view rendering while maintaining the geometric features of the image.

To simulate the blurred imaging process, we study the characteristics of macrophotography. Our key observation shows that the defocus blur in macrophotography images is much more sensitive to depth changes than natural images. To this end, we carefully design a blur module. This module contains multiple networks to fully extract information such as depth in 3D scenes and generate defocus maps. These processes take into account the characteristics of macrophotography and effectively avoid unacceptable results caused by small errors in parameter estimation. Based on the defocus map, we assign a blurred convolution kernel to each pixel to restore a clear 2D macrophotography image. In addition, we introduce the latest differentiable rendering method [KKLD23] into our method as a reliable 3D representation to estimate depth, maintain multi-view consistency and restore the 3D appearance of small objects. Combining it with the blur module achieves the simulation of the blurred imaging process, so that we only use the blurred image as input and obtain clear 3D reconstruction results through self-supervised optimization.

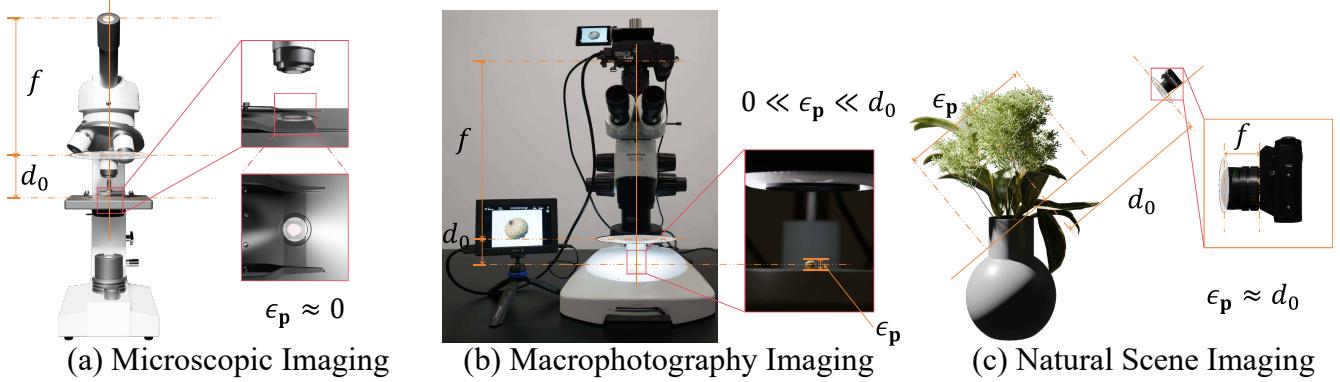
We conduct extensive experiments on both synthetic and real datasets of macrophotography defocus blur. Results show that our method outperforms existing single image deblurring methods (i.e., combined with 3D reconstruction methods) and other existing 3D scene deblurring methods in terms of deblurred reconstruction, as shown in Sec. 4. In summary, our contributions can be summarized into three parts:

- We propose the first joint self-supervised deblurring and 3D reconstruction method for macrophotography, recovering the sharp 3D appearance of subcentimeter-sized small objects from blurry inputs without requiring additional supervision.
- We formulate the deblurring problem within a multi-view 3D reconstruction framework, leveraging 3D constraints across different viewpoints to overcome the ill-posed problem of single-image deblurring and enable joint optimization for both clear 3D scenes and accurate blur modeling.
- We design BlurNet to extract depth and other features from 3D scenes to guide reconstruction, effectively incorporating optical priors of macrophotography to enhance deblurring performance in macro scenes.

2. Related Work

2.1. Focusing Techniques for Macrophotography Devices

In order to handle the defocus blur in macrophotography images, various specialized capturing devices have been designed to mitigate this effect [XTM^{*}17, PPB^{*}19, LMZ^{*}21, TPSY22,



Imaging Scene	Microscopic Scenes	Macrophotography Scenes	Natural Scene
Object Size	Micrometer scale	Sub-centimeter scale	Centimeter to meter scale
Complexity of 3D Information	2D plane	Rich 3D details	Rich 3D details
DoF(Depth of Field)	Shallow DoF	Shallow DoF	Wide DoF

Figure 2: Comparison of imaging scenes and their characteristics. In microscopic imaging, imaged objects are often slides with no 3D reconstruction possibilities. In macrophotography imaging, the imaged objects are with 3D structure, but small and have a shallow depth of field, resulting in defocus blur. In natural scene imaging, objects are distributed over a wide range of depths. During real macrophotography image capture, one of our measurements is $f + d_0 \approx 480\text{mm}$ and $\epsilon_p \approx 3\text{mm}$. This critical difference makes macro scenes 3D reconstructable but extremely sensitive to depth variations compared to other scenes, which the proposed method aims to address.

YRH*18, LWMC19]. Auto-focusing microscopes [XTM*17, PPB*19, LMZ*21] attempt to detect the depth of objects in the field of view and drive the focusing device to adjust the optical system to complete the focusing process [ZY21]. Microlens-assisted microscopes [TPSY22] replace a single lens in a microscope with a microlens array, thereby expanding the field of view. Multi-focus microscopes [LWMC19] can set multiple focal points at the same time, but at the cost of a complex system and reduced resolution and speed. However, defocus blur is caused by the inherent optical effects of the lens and is difficult to eliminate, so it is necessary to introduce an image defocus deblurring method in the post-processing of macrophotography.

2.2. Defocus Deblurring for Macrophotography

In macrophotography images, defocus blur is usually directly modeled using a complex network to map the blurred image to the sharp image. Due to the high sensitivity of these images to parameters, end-to-end methods have higher robustness and higher performance. Based on this, many works [YBMA*18, Gu23, ZZY*24, WH23] have studied the removal of defocus blur in macrophotography images. For example, early methods [YBMA*18] attempted to use deep neural networks to evaluate the focus quality of macrophotography images. Later, a coarse-to-fine deblurring method [WH23] based on U-Net [RFB15] was shown to be effective. The single focus estimation method KDAF [Gu23] introduced the concept of blur kernel to predict the focal length of macrophotography images. Recently, MPT-Catablur [ZZY*24] introduced an attention mechanism and used a pyramid structure to achieve image deblurring. These methods usually require training a powerful and large restoration model, which contains a huge number of training

pairs and a large neural network, consuming a lot of training time and memory. However, due to the difficulty in obtaining real training pairs on a large scale, such a model is difficult to adapt to every type of degradation. The 2D image spatial deblurring method is not easy to apply to 3D scenes due to the inevitable lack of multi-view consistency.

2.3. 3D Reconstruction from Defocus Blur

Differentiable rendering techniques offer more possibilities for defocus deblurring. 3D Gaussian Splatting (3DGS) [KKLD23] and Neural Radiance Field (NeRF) [MST*20] are effective methods for creating high-fidelity 3D scenes from 2D images. These techniques introduce multi-view consistent 3D spatial information into the image, enabling deblurring methods to take it into account. Many deblurring methods [PTZ*24, LLS*24, WCC24, CR23, MLL*22] based on NeRF [MST*20] and 3DGS [PTZ*24, LLS*24, WCC24, CR23, MLL*22] attempt to reconstruct sharp 3D scenes from multi-view images with defocus blur. For example, Deblur-NeRF [MLL*22] and PDRF [CR23] use 2D sparse pixel-level kernels to predict blurred images. With the development of 3DGS, 3DGS-based deblurring methods have also emerged. Deblurring-3DGS [LLS*24] optimizes the spatial properties of Gaussian primitives to obtain latent sharp 3D scenes, BAGS [PTZ*24] uses the efficient rasterizer of 3DGS to estimate the blur kernel pixel by pixel, and DOF-GS [WCC24] uses the physical model of the camera for controllable depth of field rendering. Although these 3D deblurring methods successfully recover sharp images from blurry inputs, there is currently no work that attempts to extend the applicability of deblurring methods to macrophotography images. Therefore, we propose a joint deblur-

ring and 3D reconstruction method for these tiny objects, which is able to recover sharp 3D appearance and images from a small number of multi-view blurred macrophotography images, considering the differences between macro and natural scenes.

3. Method

Our goal is to recover the geometry and appearance of a macro scene \mathcal{S} from multi-view blurred images $\{\mathbf{I}_m\}_{m=1}^M$ to produce sharp images. The pipeline of our method is shown in Fig. 3. To achieve this goal, we design an end-to-end optimization method, which explicitly models the defocus blur kernel in the micro-scene and utilizes a small network to predict it, jointly optimize the network and the 3D scene, and finally obtain a sharp 3D scene without complex network architecture and supervision.

In Sec. 3.1, we analyze the modeling of the defocus blur kernel in macrophotography and how to build a network to predict it. Based on this, we propose a joint approach for defocus deblurring and 3D reconstruction in Sec. 3.2. Sec. 3.3 details the optimization techniques of our approach.

3.1. Defocus Blur Kernel in Macrophotography

In many scenarios, the impact of defocus blur on imaging results cannot be ignored, which can be measured by the Circle of Confusion (CoC). Assuming that the camera uses an ideal thin lens, the CoC of a pixel \mathbf{p} with a depth of $d_{\mathbf{p}}$ refers to the circular area in the imaging plane onto which the pixel is projected by the lens. According to the imaging law in geometric optics [LHW05, Hec17], the radius $\sigma_{\mathbf{p}}$ of pixel \mathbf{p} satisfies

$$\sigma_{\mathbf{p}} = \frac{1}{2} f A \frac{|d_{\mathbf{p}} - d_0|}{d_{\mathbf{p}}(d_0 - f)}, \quad (1)$$

where f is the focal length of the camera, A is the F-number of the aperture, and d_0 is the depth of focal plane.

According to Eq.(1), large focal lengths and apertures of macro lenses and small depth of the object drastically increase blur-depth sensitivity. The equivalent parameters of cameras with macro lenses exceed normal cameras by orders of magnitude. In addition, the depth of the object in macro scenes is much smaller than that in natural scenes. These factors yield a critically shallow depth of field. It fundamentally prevents full focus coverage and causes unavoidable defocus blur regardless of focal plane adjustments. More details about CoC in macrophotography can be found in the Supplementary Material.

Given that the defocus blur in macrophotography images exhibits strong isotropy and smoothness [QWJ21], we consider the CoC of the pixel \mathbf{p} as an isotropic 2D Gaussian blur kernel [SXJ15, XQJ17, LLCL19] $\mathbf{G}_{\mathbf{p}}$ with variance $\sigma_{\mathbf{p}}$ in practice. Therefore, consider a defocused macrophotography image $\tilde{\mathbf{I}}$, and its corresponding clear image \mathbf{I} , the color of pixel \mathbf{p} in two images is related by the following formula:

$$\tilde{\mathbf{I}}[\mathbf{p}] = \sum_i \sum_j \mathbf{G}_{\mathbf{p}}[i, j] \mathbf{I}[\mathbf{p} + (i, j)], \quad (2)$$

where $\mathbf{G}_{\mathbf{p}}$ is the blur kernel of pixel \mathbf{p} , satisfying

$$\mathbf{G}_{\mathbf{p}}[i, j] = \frac{\exp(-(i^2 + j^2)/2\sigma_{\mathbf{p}}^2)}{\sum_{i,j=-K}^K \exp(-(i^2 + j^2)/2\sigma_{\mathbf{p}}^2)}. \quad (3)$$

As shown in Fig. 2, our key observation is that in macro scenes, while the possibility of 3D reconstruction is preserved compared to microscopic scenes, imaged objects are often closer to the camera focal plane than in natural scenes, and the depth $d_{\mathbf{p}}$ of each pixel \mathbf{p} varies only within a very small range around the focal plane depth d_0 . Even small depth perturbations can lead to visible defocus blur. Due to the sensitivity of $\sigma_{\mathbf{p}}$ to small errors, we choose to extract depth and other features through several carefully designed networks and predict $\sigma_{\mathbf{p}}$ from them.

Denote $\epsilon_{\mathbf{p}}$ as the depth of pixel \mathbf{p} relative to the focal plane, i.e. $d_{\mathbf{p}} - d_0 = \epsilon_{\mathbf{p}}$. Then Eq. (1) can be written as

$$\sigma_{\mathbf{p}} = \frac{1}{2} f A \frac{|\epsilon_{\mathbf{p}}|}{(d_0 + \epsilon_{\mathbf{p}})(d_0 - f)}. \quad (4)$$

Given that in macrophotography the depth of the object $d_{\mathbf{p}}$ is typically very close to the focal plane depth d_0 , the relative depth $|\epsilon_{\mathbf{p}}| = |d_{\mathbf{p}} - d_0|$ is much smaller than d_0 itself (i.e., $|\epsilon_{\mathbf{p}}| \ll d_0$). This crucial condition allows for the approximation $(d_0 + \epsilon_{\mathbf{p}}) \approx d_0$ in the denominator of Eq. (4). Consequently, Eq. (4) simplifies to:

$$\sigma_{\mathbf{p}} \approx \frac{f A}{2} \frac{|\epsilon_{\mathbf{p}}|}{d_0(d_0 - f)}. \quad (5)$$

By defining a term $\alpha = \frac{f A}{2d_0(d_0 - f)}$, which groups the camera's optical parameters (f, A) and the focal plane depth (d_0) for the current view, we arrive at the factored form:

$$\sigma_{\mathbf{p}} = \alpha \cdot |\epsilon_{\mathbf{p}}|. \quad (6)$$

Eq. (6) shows that $\sigma_{\mathbf{p}}$ can be divided into two parts α and $\epsilon_{\mathbf{p}}$, where $\epsilon_{\mathbf{p}}$ is relative to $d_{\mathbf{p}}$ but α is not. Therefore, we design two networks for each part. Formally:

$$\sigma_{\mathbf{p}} = A \odot E_{\mathbf{p}}, \quad (7)$$

where E is a feature map relative to depth and A is the feature independent to depth, both of which are output by a carefully designed network, and \odot represents element-wise multiplication. The network architecture is detailed in Sec. 3.2.

3.2. Joint Method for Deblurring and Reconstruction

As shown in Fig. 3, the input of our method is a set of multi-view blurry macrophotography images $\{\mathbf{I}_m\}_{m=1}^M$. Through the latest 3D reconstruction algorithm [KKLD23, SF16, SZPF16], we get a roughly initialized 3D scene, which can achieve differentiable rendering of novel view images. Then, our method generates correct blurred images by assigning blur kernels to the rendered images through our carefully designed blur kernel network (called BlurNet). Through BlurNet, we can obtain blurred images $\tilde{\mathbf{I}}_{\text{out}}$ and clear images \mathbf{I}_{out} .

Our optimization strategy is to refine the parameters of both the 3D scene \mathcal{S} which implicitly determines \mathbf{I}_{out} and the depth map \mathbf{D} rendered from the scene, and BlurNet. This joint optimization is

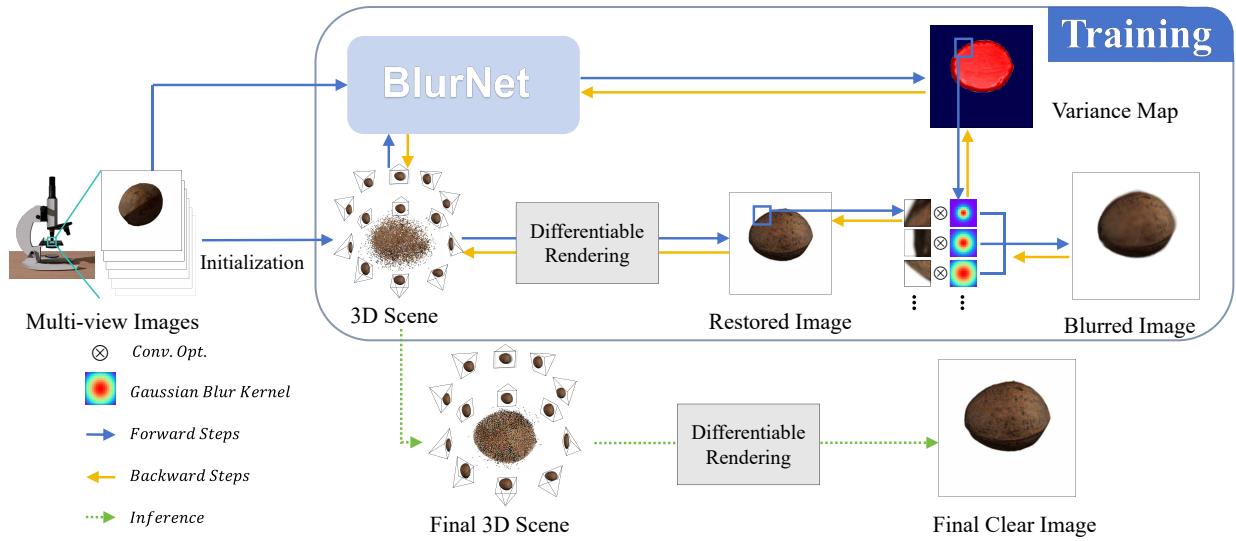


Figure 3: An overview of our joint deblurring and 3D reconstruction method. First, we initialize a differentiable coarse 3D scene from multi-view input blurred images. Afterwards, BlurNet leverages information from the input images to generate a variance map (i.e. defocus map) for each view. The variance map is then convolved with the rendered image pixel-wise. The output is a correctly blurred image, supervised by the blurred input images during training. Note that this process ultimately produces a sharp scene.

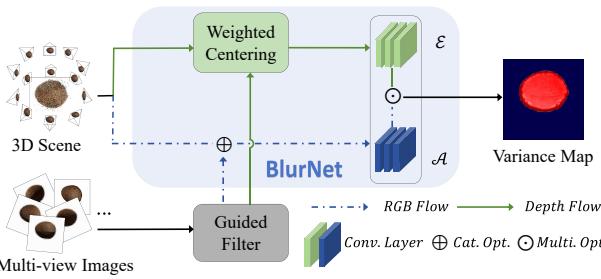


Figure 4: The architecture of BlurNet. At the beginning of training, the clarity mask is generated from the input multi-view images via guided filters. The depth map is weighted centered using the clarity mask to generate a relative depth map for the depth CNN E . The rendered image is concatenated with the clarity mask as input to the RGB CNN A . The outputs of both CNNs are multiplied to generate a variance map.

supervised by minimizing the discrepancy between the input blurry image \mathbf{I} and the simulated blurred image $\tilde{\mathbf{I}}_{\text{out}}$. The ultimate goal is to recover a high-fidelity clear 3D scene \mathcal{S} and a BlurNet capable of accurately modeling the specific defocus characteristics in macrophotography. This process ensures that the information about the defocus blur is effectively captured by BlurNet, while the 3D scene representation converges towards a sharp depiction of the object.

The key component in our method is BlurNet, whose structure is shown in Fig. 4. As described in Eq. (7), we design two networks to predict σ_p , one of which is related to the depth of the pixel and the other is independent of it. Using the depth map \mathbf{D} and the rendered RGB image \mathbf{I}_{out} output by the 3D scene representation as input of the network, Eq. (7) can be written as:

$$\sigma_p = A(\mathbf{I}_{\text{out}}) \odot E_p(\mathbf{D}). \quad (8)$$

For $E(\cdot)$ in Eq. (8), the original depth map \mathbf{D} is not enough for extracting depth features. We use the clarity mask \mathbf{M} as a weight to weighted center the depth map \mathbf{D} . In this way, we can accurately locate the in-focus area in each input image \mathbf{I} , and then calculate the average depth of the in-focus area as d_0 in Eq. (1), and finally obtain a relative depth map $\hat{\mathbf{D}}$. It can be written as:

$$\hat{\mathbf{D}}_p = \left| \mathbf{D}_p - \sum_{x \in \mathcal{P}} \mathbf{D}_x \mathbf{M}_x \right|, \quad (9)$$

where \mathcal{P} is the set that contains all pixels in the image \mathbf{I} .

While the global depth map from 3DGs may exhibit inconsistencies, it preserves local geometric fidelity, with neighboring pixels maintaining relative depth relationships critical for blur estimation. Therefore, we design a lightweight CNN encoder $E(\cdot)$ operating in these local neighborhoods, learning to extract and refine depth-aware features based on local geometric fidelity, and utilizing the positional encoding technique [MST*20, TSM*20] to provide a higher dimensional input to the CNN. The structure of the CNN is:

$$E(\mathbf{D}) = E(P_L(\hat{\mathbf{D}})), \quad (10)$$

where $P_L(\cdot)$ is a positional encoding function with frequency L and E is a small CNN.

For $A(\cdot)$ in Eq. (8), we also apply a small CNN to extract RGB features, and we find \mathbf{M} is helpful in feature extraction. Therefore,

$$A(\mathbf{I}_{\text{out}}) = \mathcal{A}(\mathbf{I}_{\text{out}} \oplus \mathbf{M}), \quad (11)$$

where \mathcal{A} is another small CNN.

BlurNet effectively extracts depth and other necessary features from the 3D scene and original multi-view images, ensuring the robustness of our method. Through BlurNet, we can get a variance map σ as the input of Eq. (3), and then we can apply Eq. (3) and Eq. (2) to get the blur kernel of all pixels and blurred image $\tilde{\mathbf{I}}_{\text{out}}$.

3.3. Representation and Optimization

3.3.1. 3D scene representation.

Our method uses 3DGS [KKLD23] as the basic representation of macrophotography scenes. In 3DGS, the scene is represented by a set of Gaussian points $\{\mathcal{G}_n | n = 1, \dots, N\}$, each of which contains multiple attributes, including center x_n , covariance matrix Σ_n , opacity α_n , and color \mathbf{c}_n . With carefully designed tiling, 3DGS provides sufficient rendering speed to develop new methods for clearer multi-view high-fidelity rendering. However, practical applications of 3DGS in macrophotography can be challenging. Despite its impressive performance on curated datasets, 3DGS requires high quality images input to work well. Without this, 3DGS often generates undesired Gaussians to overfit observation noise, leading to worse renderings [PTZ^{*}24].

3.3.2. Clarity mask.

The clarity mask \mathbf{M} accurately locates the focus area in our method. In the preprocessing, an effective guided filter-based algorithm [QLZY19] is introduced to obtain the clarity mask \mathbf{M} for each image input \mathbf{I} .

Specifically, at the beginning of training, we use a mean filter to extract the high-frequency information retained in the image to form a coarse clarity mask. Then, with the original image as a guide, we apply a guided filter to the coarse clarity mask to enhance the high-frequency information in the coarse clarity mask to form a refined clarity mask

$$\mathbf{M} = \text{GF}(\mathbf{I}, |\mathbf{I} - \text{MF}(\mathbf{I})|), \quad (12)$$

where MF is the mean filter operator and GF is the guided filter operator. This operation only needs to be performed once at the beginning of training, so it does not slow down training. Fig. 5 illustrates how clarity masks accurately identifies high-frequency, in-focus regions across blurry input. For more details of the guided filter-based algorithm, please refer to the Supplementary Materials.



Figure 5: Visualization of computed clarity masks. Comparison between defocused input and their corresponding clarity masks demonstrates how our method identifies high-frequency, in-focus regions (bright areas in mask) to estimate the focal plane depth.

3.3.3. Multi-stage training.

Our training is divided into three stages: initialization training, low-scale training, and high-scale training. The purpose of initialization training (first 3,000 iterations) is to create a good initial representation of the 3D scene \mathcal{S} . This stage follows the standard 3DGS training procedure. The training loss \mathcal{L}_{pre} during this stage, aims to reconstruct a plausible 3D scene by directly comparing the rendered image \mathbf{I}_{out} with the input blurry image \mathbf{I} :

$$\mathcal{L}_{\text{pre}} = (1 - \lambda) \|\mathbf{I} - \mathbf{I}_{\text{out}}\| + \lambda \mathcal{L}_{\text{D-SSIM}}(\mathbf{I}, \mathbf{I}_{\text{out}}), \quad (13)$$

where $\mathcal{L}_{\text{D-SSIM}}$ represents the structural similarity loss. This initial supervision, while using blurry targets, helps establish a foundational geometry.

In the subsequent low-scale training stage (iterations 3,000 to 15,000), we downsample the training images \mathbf{I} (e.g., with a scaling factor $s = 2.0$) and introduce BlurNet for joint optimization with the 3D scene. This is followed by the high-scale training stage (iterations 15,000 to 30,000) using the original image resolution ($s = 1.0$), where a new hidden layer may be added to BlurNet to enhance its fitting capability. During both low-scale and high-scale stages, the primary training loss \mathcal{L} is employed. This loss supervises the simulated blurred image \mathbf{I}_{out} generated by convolving the rendered sharp \mathbf{I}_{out} with predicted kernels by BlurNet to match the input blurry image \mathbf{I} , while also applying a total variation regularization to the rendered sharp image \mathbf{I}_{out} to encourage smoothness:

$$\mathcal{L} = (1 - \lambda) \|\mathbf{I} - \tilde{\mathbf{I}}_{\text{out}}\| + \lambda \mathcal{L}_{\text{D-SSIM}}(\mathbf{I}, \tilde{\mathbf{I}}_{\text{out}}) + \mathcal{L}_{\text{TV}}(\mathbf{I}_{\text{out}}). \quad (14)$$

By replicating the physical blurring process in rendering with priors from macrophotography, the end-to-end framework inherently constrains the solution space. This structural regularization enables effective supervision.

This multi-stage training strategy has several advantages: any degree of blurry pixels can be effectively modeled in multi-scale training, without multi-scale training, some kinds of blur kernels cannot be modeled; low-scale BlurNet provides an effective warm-up for high-scale training; and since half of the iterations are performed at low scales, training time is reduced by about half compared to using only high-scale inputs.

4. Experiments

4.1. Implementation Details

4.1.1. Training Settings.

We use the Adam optimizer [KB15] and set the learning rate of the two CNNs \mathcal{A} and \mathcal{E} to 0.001. For 2D Gaussian kernel, we set the pixel kernel size of each pixel to $2K + 1 = 13$, i.e. $K = 6$, and the other hyperparameters are the same as those in 3D-GS. \mathcal{A} and \mathcal{E} are both 2-layer CNNs with 32 hidden units in each layer and ReLU as the activation function. Only the first layer is applied in the low-scale training stage, while the second is added during high-scale training. The frequency L in positional encoding $P_L(\cdot)$ is 7. The total number of training iterations is 30,000, of which the first 3,000 iterations are pre-training, 3,000 to 15,000 iterations are low-scale training, and 15,000 to 30,000 iterations are high-scale training. All experiments are performed on an NVIDIA RTX 4090 GPU.

4.1.2. Datasets.

We use a camera with a macro lens to take multi-view images of 5 common subcentimeter tiny objects as a real dataset for evaluation, each of which contains about 200 images. The macro lens we use is an Olympus M.Zuiko Digital ED 60mm F2.8 Macro, paired with a Panasonic Lumix DC-GH5 camera. We use a turnable device to capture multi-view images, which enables us to change object pose easily while these objects are too small to change poses. Considering the difficulty for COLMAP [SF16, SZPF16] to estimate camera poses under extreme camera parameters and critical defocus blur,

Methods Metrics	3DGS [KKLD23]			Deblur-3DGS [LLS*24]			BAGS [PTZ*24]			Restormer+3DGS [ZAK*22]			INIKNet+3DGS [QWJ23]			Ours		
	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS	PSNR	SSIM	LPIPS
Micro-Chair	27.37	0.926	0.074	27.29	0.926	0.075	26.12	0.918	0.082	28.42	0.941	0.056	29.55	0.947	0.054	30.61	0.962	0.038
Micro-Ficus	24.00	0.862	0.132	24.00	0.863	0.132	24.63	0.858	0.128	24.66	0.887	0.105	24.68	0.882	0.109	25.71	0.902	0.088
Grain	33.40	0.925	0.119	33.42	0.925	0.119	33.04	0.925	0.123	34.19	0.937	0.098	34.92	0.938	0.099	34.77	0.942	0.086
Micro-Lego	24.49	0.841	0.176	24.50	0.842	0.176	24.32	0.832	0.182	25.96	0.876	0.139	26.58	0.880	0.138	26.33	0.893	0.119
Locust	37.60	0.978	0.036	37.58	0.978	0.037	35.57	0.968	0.051	38.86	0.983	0.026	38.64	0.982	0.029	37.92	0.981	0.023
Seed	38.73	0.956	0.090	38.56	0.953	0.092	35.88	0.942	0.110	38.19	0.958	0.078	38.62	0.955	0.085	40.26	0.974	0.056
Micro-Ship	25.79	0.818	0.221	25.78	0.816	0.222	25.00	0.801	0.240	26.20	0.828	0.212	26.33	0.831	0.207	26.57	0.842	0.171
Spider	37.50	0.965	0.058	37.51	0.965	0.058	35.67	0.954	0.071	38.41	0.975	0.041	37.74	0.971	0.048	37.86	0.972	0.038
Average	31.11	0.909	0.113	31.08	0.909	0.113	30.03	0.900	0.123	31.86	0.923	0.094	32.13	0.923	0.096	32.50	0.934	0.077

Table 1: Quantitative comparisons on the synthetic dataset. All metrics are averaged over the entire test set. We color code the **best** and **second best**.

we use a more professional software Agisoft Metashape [Agi16] to calculate the camera poses of blurry images and reference images in these real scenes, and then convert them to the form that COLMAP can read to check correctness. In addition, we selected 4 objects from the NeRF dataset [MST*20] and selected 4 common small objects, which were carefully processed into macro-scale models using Blender [Com18] as the synthetic dataset, and rendered 100 multi-view images for each object with reference to real-world camera parameters. It contains both blurry images and clear images from the same view, which is convenient for calculating quantitative results. More dataset settings can be found in the Supplementary Material.

4.2. Comparisons

Due to significant differences between macrophotography and natural imaging, existing natural deblurring methods [ZZY*24, ZSL*22] are not suitable for macrophotography deblurring. To the best of our knowledge, no studies have attempted to reconstruct tiny objects from multi-view blurred macrophotography inputs. We compare with several potentially effective methods: (1) 3DGS-based deblurring methods including BAGS [PTZ*24], Deblurring-3DGS [LLS*24], and naive 3DGS [KKLD23]; (2) image-space deblurring methods that perform well in macro scenes, combined with 3DGS. We evaluated LaKDNet + 3DGS [RBpS*23], Restormer-TLC + 3DGS [CCCL22], Restormer + 3DGS [ZAK*22], and INIKNet + 3DGS [QWJ23], and selected the two best-performing methods [ZAK*22, QWJ23] for comparison. Following 3DGS, we use PSNR, SSIM, and LPIPS as evaluation metrics for novel view synthesis quality.

Tab. 1 shows the quantitative comparison results. As shown in Tab. 1, our method performs better than other advanced deblurring and reconstruction methods, especially achieving significant improvements on SSIM and LPIPS. This shows that our method maintains good geometric characteristics during deblurring and reconstruction, keeping multi-view consistency. Note that BAGS and Deblurring-3DGS perform slightly worse than naive 3DGS. One possible reason is that their approaches introduce depth but handle it improperly. The importance and sensitivity of depth features in macrophotography lead to their performance degradation. Fig. 7 and Fig. 6 show the qualitative comparison results on synthetic and captured real datasets. The images rendered by our method have sharp edges and rich details, which are closest to reality. Other methods suffer from artifacts near object boundaries and

blurred textures. For more reconstruction results, please see our demo video.

4.3. Ablation Studies

4.3.1. Evaluation of the blur kernel.

We validate the effectiveness of our blur kernels from two perspectives. First, Fig. 8 demonstrates that our BlurNet accurately estimates variance maps that correspond well to the defocused regions in the input images, showing larger variance values in more blurred areas for both synthetic and real macrophotography data. Then, we quantitatively evaluate the blur kernel reconstruction accuracy on synthetic data. Clean images generated by Blender are used to construct their corresponding ground truth 3DGS models. The clean images rendered from novel views are then reblurred using the blur kernels recovered by our method and compared with the real blurred images from the corresponding views. Tab. 2 shows that our blur kernels can accurately reconstruct the defocus blur in the blurred images.

Metric	Grain	Locust	Seed	Spider	Average
PSNR \uparrow	33.58/41.35	37.98/41.92	39.86/45.69	37.54/41.35	37.24/42.58
SSIM \uparrow	0.931/0.989	0.980/0.991	0.967/0.991	0.968/0.986	0.962/0.989
LPIPS \downarrow	0.108/0.037	0.033/0.017	0.077/0.031	0.050/0.027	0.067/0.028

Table 2: Quantitative evaluation of blur kernel accuracy. Left: Metrics between the clear renders and GT blurry images. Right: Metrics between our re-blurred results and GT blurry images.

4.3.2. Effectiveness of Components.

We conduct ablation studies on the four components of our method to demonstrate their effectiveness:

- **Multi-Stage Training:** We compare the results of multi-stage training with those of a single-stage training where only high-scale inputs are used.
- **Clarity Mask:** We compare our method with and without a clarity mask \mathbf{M} . Without \mathbf{M} , we only use the RGB image and depth map as BlurNet input.
- **BlurNet:** We use a simple 2-layer CNN instead of BlurNet to implement the method. The clarity mask, RGB image, and depth map are concatenated as CNN input.
- **Depth Input:** We remove the depth map generated in training process. Without depth, BlurNet can only predict the blur kernel from RGB information.

Tab. 3 shows the ablation study results for this part. Overall, the best results are achieved when using the full model. The clarity

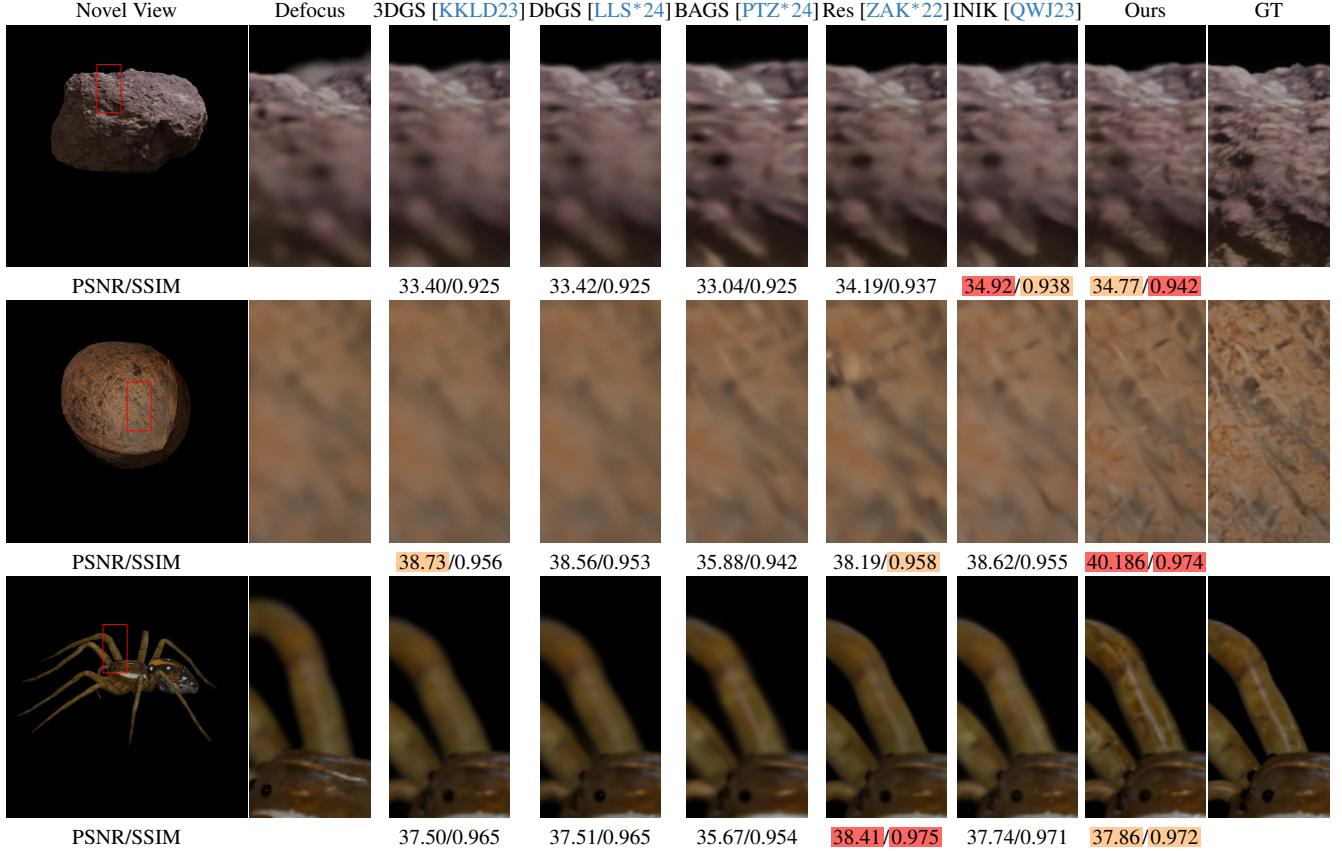


Figure 6: Visual comparisons on novel views of synthetic images dataset. DbGS, Res, and INIK are the shorthands for Deblurring-3DGs, Restormer+3DGs, and INIKNet+3DGs. We color code the **best** and the **second best**.

mask provides additional information about the camera focal plane in the blurry input and plays an important role in handling depth. Our carefully designed network further improves the performance of our method. Multi-stage training enables our method to model blur kernels of all sizes. Without it, kernels that are too large cannot be modeled. Given defocus blur as a physically inherent function of depth, depth information introduces physical constraints for BlurNet and plays a key role in deblurring. When various strategies cooperate, our method achieves improvements in all metrics. Detailed visualization results of ablation studies are provided in the Supplementary Material.

4.3.3. Size of Blur Kernels.

The size of the defocus blur kernel is a key hyper-parameter in our method. We study the impact of the blur kernel size K on the reconstruction quality in these scenarios. We try different K values using the full model on our dataset. Fig. 9a shows that larger kernels generally lead to higher performance. However, when we set $K = 8$ (i.e. $2K + 1 = 17$), the performance is almost the same as $K = 6$, while Fig. 9b shows large kernels severely slow down training and may cause OOM issues. Considering these factors, we choose $K = 6$ as a balanced choice.

5. Conclusion

This paper presented a novel self-supervised method for joint de-focus deblurring and 3D reconstruction, specifically tailored for macrophotography of sub-centimeter scale objects. Our approach uniquely models spatially-varying defocus blur kernels, leveraging 3D scene information like depth from 3D scene via our BlurNet module to accurately reflect macro-specific optical effects. This enables simultaneous optimization of a sharp 3D scene and the blur model using differentiable rendering, without requiring a mount of paired training data. The key insight is that the inherent multi-view consistency constraints in 3D reconstruction and optical prior in macrophotography provide sufficient supervision to jointly learn both scene geometry and blur characteristics, eliminating the need for explicit supervision. Extensive experiments demonstrate superior performance in both image deblurring and 3D reconstruction quality compared to existing methods. Future work will focus on enhancing depth estimation robustness and exploring applications in broader macro imaging contexts.

6. Acknowledgement

This research was supported by the National Natural Science Foundation of China(No.62441224, No.62272433), and the Fundamental Research Funds for the Central Universities.

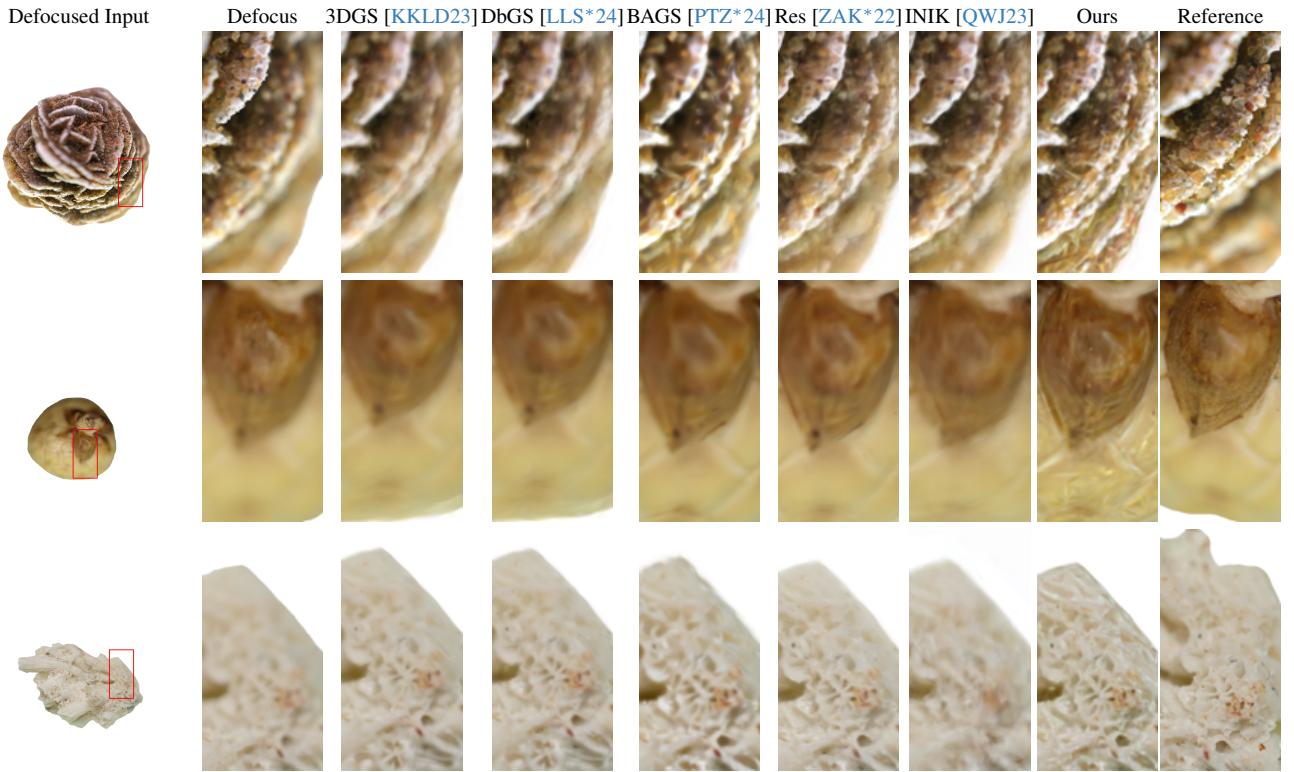


Figure 7: Visual comparisons on test views real-world macrophotography images dataset. DbGS, Res, and INIK are the shorthands for Deblurring-3DGS, Restormer+3DGS, and INIKNet+3DGS. Macro lenses can not capture all-in-focus images. Therefore, we can only compare our results with clear parts in the reference images. For the same reason, quantitative comparison cannot be carried out.

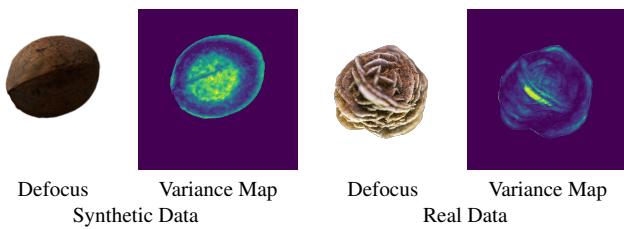


Figure 8: Visualization of estimated variance maps. Comparison between defocused input images and estimated variance maps qualitatively demonstrates our method's ability to accurately predict larger blur kernel variances in defocused regions.

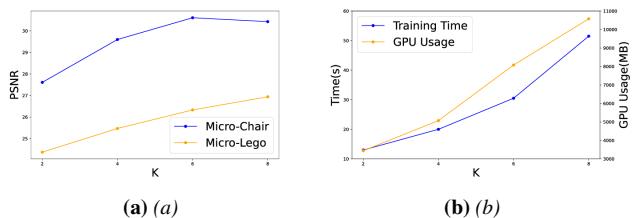


Figure 9: Ablation studies on hyper-parameter kernel size K. In (a), we measure the PSNR and SSIM of two scenes under different values of K. (b) compares the training time and GPU memory usage under different values of K.

References

[Agi16] AGISOFT: Agisoft PhotoScan Standard (Version 1.2.6) (Software), 2016. Available at: <http://www.agisoft.com/downloads/installer/>.

	Average	w/o stages	w/o mask	w/o BlurNet	w/o depth	Ours
PSNR	28.32	30.95	31.76	31.56		32.50
SSIM	0.890	0.908	0.930	0.918		0.934
LPIPS	0.107	0.107	0.081	0.101		0.077

Table 3: Quantitative ablation studies on several novel designs in our method. We color code the **best**. The results indicate that our full model achieves the best performance.

- [Bis17] BISS L.: *Microsculpture: Portraits of Insects*. ABRAMS, 2017.
- [CCCL22] CHU X., CHEN L., CHEN C., LU X.: Improving image restoration by revisiting global information aggregation. In *Computer Vision – ECCV 2022* (Cham, 2022), Avidan S., Brostow G., Cissé M., Farinella G. M., Hassner T., (Eds.), Springer Nature Switzerland, pp. 53–71. 7
- [Com18] COMMUNITY B. O.: *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. URL: <http://www.blender.org>. 7
- [CR23] CHENG P., RAMA C.: Pdrf: Progressively deblurring radiance field for fast and robust scene reconstruction from blurry images, 2023. 3
- [CSE*22] CUNY A. P., SCHLOTTMANN F. P., EWALD J. C., PELET S., SCHMOLLER K. M.: Live cell microscopy: From image to insight. *Biophysics Reviews* 3, 2 (Apr. 2022), 021302. doi:[10.1063/5.0082799](https://doi.org/10.1063/5.0082799). 1
- [GTS20] GHAMSARIAN N., TASCHWER M., SCHOEFFMANN K.: Deblurring Cataract Surgery Videos Using a Multi-Scale Deconvolutional Neural Network. In *2020 IEEE 17th International Symposium on*

- Biomedical Imaging (ISBI)* (Apr. 2020), pp. 872–876. doi:10.1109/ISBI45749.2020.9098318. 2
- [Gu23] GU Y.: Single-shot focus estimation for microscopy imaging with kernel distillation. *IEEE Transactions on Computational Imaging* 9 (2023), 542–550. doi:10.1109/TCI.2023.3281189. 3
- [Hech17] HECHT E.: *Optics*. Pearson Education, Incorporated, 2017. 4
- [JLD*20] JIANG C., LIAO J., DONG P., MA Z., CAI D., ZHENG G., LIU Y., BU H., YAO J.: Blind deblurring for microscopic pathology images using deep learning networks. *arXiv preprint arXiv:2011.11879* (2020). 2
- [JLH*17] JEFFREY KUO C.-F., LO W.-C., HUANG Y.-R., TSAI H.-Y., LEE C.-L., WU H.-C.: Automated defect inspection system for cmos image sensor with micro multi-layer non-spherical lens module. *Journal of Manufacturing Systems* 45 (2017), 248–259. doi:10.1016/j.jmsy.2017.10.004. 2
- [KB15] KINGMA D. P., BA J.: Adam: A method for stochastic optimization, 2015. 6
- [KKLD23] KERBL B., KOPANAS G., LEIMKÜHLER T., DRETTAKIS G.: 3d gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics* 42, 4 (July 2023). URL: <https://repo-sam.inria.fr/fungraph/3d-gaussian-splatting/>. 2, 3, 4, 6, 7, 8, 9
- [KSN*21] KORNILOVA A., SALNIKOV M., NOVITSKAYA O., BEGICHEVA M., SEVRUGOV E., SHCHERBAKOV K., PRONINA V., DYLOV D. V.: Deep Learning Framework For Mobile Microscopy. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)* (Apr. 2021), pp. 324–328. doi:10.1109/ISBI48211.2021.9434133. 2
- [LHCT20] LIU J., HUANG X., CHEN L., TAN S.: Deep learning-enhanced fluorescence microscopy via degeneration decoupling. *Optics Express* 28, 10 (May 2020), 14859–14873. doi:10.1364/OE.390121. 2
- [LHW05] LI L., HUANG Y., WANG Y.: *Applied Optics*. Beijing Institute of Technology Press, 2005. 4
- [LLCL19] LEE J., LEE S., CHO S., LEE S.: Deep defocus map estimation using domain adaptation. In *IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 12214–12222. 4
- [LLS*24] LEE B., LEE H., SUN X., ALI U., PARK E.: Deblurring 3d gaussian splatting. In *ECCV* (2024). 3, 7, 8, 9
- [LMZ*21] LI C., MOATTI A., ZHANG X., GHASHGHAEI H. T., GREENBAUM A.: Deep learning-based autofocus method enhances image quality in light-sheet fluorescence microscopy. *Biomedical Optics Express* 12, 8 (2021), 5214–5226. 2, 3
- [LWMC19] LIN W., WANG D., MENG Y., CHEN S.-C.: Multi-focus microscope with hilo algorithm for fast 3-d fluorescent imaging. *PloS one* 14, 9 (2019), e0222729. 2, 3
- [Mas08] MASTERS B. R.: History of the optical microscope in cell biology and medicine. *eLS* (2008). 1
- [MF21] MA L., FEI B.: Comprehensive review of surgical microscopes: technology development and medical applications. *Journal of Biomedical Optics* 26, 1 (Jan. 2021), 010901. doi:10.1117/1.JBO.26.1.010901. 1
- [MLL*22] MA L., LI X., LIAO J., ZHANG Q., WANG X., WANG J., SANDER P. V.: Deblur-nerf: Neural radiance fields from blurry images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2022), pp. 12861–12870. 3
- [MMS*22] MENSA F. S., MUZZI M., SPANI F., TROMBA G., DULLIN C., DI GIULIO A.: When the utility of micro-computed tomography collides with insect sample preparation: An entomologist user guide to solve post-processing issues and achieve optimal 3d models. *Applied Sciences* 12, 2 (2022). doi:10.3390/app12020769. 2
- [MST*20] MILDENHALL B., SRINIVASAN P. P., TANCIK M., BARRON J. T., RAMAMOORTHI R., NG R.: Nerf: Representing scenes as neural radiance fields for view synthesis. In *ECCV* (2020). 3, 5, 7
- [OP02] ORCI L., PEPPER M. S.: Microscopy: an art? *Nature Reviews Molecular Cell Biology* 3, 2 (2002), 133–137. doi:10.1038/nrm726. 2
- [PFX*24] PRAKASH K., FRANKE C., XIA F., CHATTERJEE N., SMITH C.: Microscopy at a glance: New poster article series exploring the intersection of art, science and imaging. *Journal of Microscopy* 296, 2 (2024), 111–114. doi:10.1111/jmi.13357. 2
- [PPB*19] PINKARD H., PHILLIPS Z., BABAKHANI A., FLETCHER D. A., WALLER L.: Deep learning for single-shot autofocus microscopy. *Optica* 6, 6 (2019), 794–797. 2, 3
- [PTZ*24] PENG C., TANG Y., ZHOU Y., WANG N., LIU X., LI D., CHELLAPPA R.: Bags: Blur agnostic gaussian splatting through multi-scale kernel modeling. In *ECCV* (2024). 3, 6, 7, 8, 9
- [QLZY19] QIU X., LI M., ZHANG L., YUAN X.: Guided filter-based multi-focus image fusion through focus region detection. *Signal Processing: Image Communication* (2019). 6
- [QWJ21] QUAN Y., WU Z., JI H.: Gaussian kernel mixture network for single image defocus deblurring. In *Advances in Neural Information Processing Systems* (2021), Ranzato M., Beygelzimer A., Dauphin Y., Liang P., Vaughan J. W., (Eds.), vol. 34, Curran Associates, Inc., pp. 20812–20824. URL: https://proceedings.neurips.cc/paper_files/paper/2021/file/ae1eaa32d10b6c886981755d579fb4d8-Paper.pdf. 4
- [QWJ23] QUAN Y., WU Z., JI H.: Neumann network with recursive kernels for single image defocus deblurring. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2023), pp. 5754–5763. doi:10.1109/CVPR52729.2023.00557. 7, 8, 9
- [RBpS*23] RUAN L., BEMANA M., PETER SEIDEL H., MYSZKOWSKI K., CHEN B.: Revisiting image deblurring with an efficient convnet, 2023. URL: <https://arxiv.org/abs/2302.02234>, arXiv: 2302.02234. 7
- [RFB15] RONNEBERGER O., FISCHER P., BROX T.: U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015* (Cham, 2015), Navab N., Hornegger J., Wells W. M., Frangi A. F., (Eds.), Springer International Publishing, pp. 234–241. 3
- [RMLGARG15] RODRÍGUEZ-MARTÍN M., LAGÜELA S., GONZÁLEZ-AGUILERA D., RODRÍGUEZ-GONZÁLVEZ P.: Procedure for quality inspection of welds based on macro-photogrammetric three-dimensional reconstruction. *Optics & Laser Technology* 73 (2015), 54–62. doi:10.1016/j.optlastec.2015.04.011. 2
- [SF16] SCHÖNBERGER J. L., FRAHM J.-M.: Structure-from-motion revisited. In *Conference on Computer Vision and Pattern Recognition (CVPR)* (2016). 4, 6
- [SXJ15] SHI J., XU L., JIA J.: Just noticeable defocus blur detection and estimation. In *IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 657–665. 4
- [SZPF16] SCHÖNBERGER J. L., ZHENG E., POLLEFEYS M., FRAHM J.-M.: Pixelwise view selection for unstructured multi-view stereo. In *European Conference on Computer Vision (ECCV)* (2016). 4, 6
- [TPSY22] TRUKHOVA A., PAVLOVA M., SINITSYNA O., YAMINSKY I.: Microlens-assisted microscopy for biology and medicine. *Journal of Biophotonics* 15, 9 (2022), e202200078. 2, 3
- [TSM*20] TANCIK M., SRINIVASAN P. P., MILDENHALL B., FRIDOVICH-KEIL S., RAGHAVAN N., SINGHAL U., RAMAMOORTHI R., BARRON J. T., NG R.: Fourier features let networks learn high frequency functions in low dimensional domains. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* (Red Hook, NY, USA, 2020), NIPS ’20, Curran Associates Inc. 5
- [WCC24] WANG Y., CHAKRAVARTHULA P., CHEN B.: Dof-gs: Adjustable depth-of-field 3d gaussian splatting for refocusing, defocus rendering and blur removal, 2024. URL: <https://arxiv.org/abs/2405.17351>, arXiv: 2405.17351. 3

[WH23] WANG J., HAN B.: Defocus deblur microscopy via head-to-tail cross-scale fusion, 2023. URL: <https://arxiv.org/abs/2201.02876>, arXiv:2201.02876. 3

[WXY*23] WANG Y., XU Z., YANG Y., WANG X., HE J., REN T., LIU J.: Deblurring microscopic image by integrated convolutional neural network. *Precision Engineering* 82 (July 2023), 44–51. doi: [10.1016/j.precisioneng.2023.03.005](https://doi.org/10.1016/j.precisioneng.2023.03.005). 2

[XQJ17] XU G., QUAN Y., JI H.: Estimating defocus blur via rank of local patches. In *IEEE International Conference on Computer Vision* (2017), pp. 5371–5379. 4

[XTM*17] XU J., TIAN X., MENG X., KONG Y., GAO S., CUI H., LIU F., XUE L., LIU C., WANG S.: Wavefront-sensing-based autofocus in microscopy. *Journal of Biomedical Optics* 22, 8 (2017), 086012–086012. 2, 3

[YBMA*18] YANG S. J., BERNDL M., MICHAEL ANDO D., BARCH M., NARAYANASWAMY A., CHRISTIANSEN E., HOYER S., ROAT C., HUNG J., RUEDEN C. T., SHANKAR A., FINKBEINER S., NELSON P.: Assessing microscope image focus quality with deep learning. *BMC Bioinformatics* 19, 1 (2018), 77. doi: [10.1186/s12859-018-2087-4](https://doi.org/10.1186/s12859-018-2087-4). 3

[YRH*18] YOO S., RUIZ P., HUANG X., HE K., FERRIER N. J., HERELD M., SELEWA A., DADDYSMAN M., SCHERER N., COSSAIRT O., ET AL.: 3d image reconstruction from multi-focus microscope: axial super-resolution and multiple-frame processing. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (2018), IEEE, pp. 1453–1457. 2

[ZAK*22] ZAMIR S. W., ARORA A., KHAN S., HAYAT M., KHAN F. S., YANG M.-H.: Restormer: Efficient transformer for high-resolution image restoration. In *CVPR* (2022). 7, 8, 9

[ZKC*20] ZHAO H., KE Z., CHEN N., WANG S., LI K., WANG L., GONG X., ZHENG W., SONG L., LIU Z., LIANG D., LIU C.: A new deep learning method for image deblurring in optical microscopic systems. *Journal of Biophotonics* 13, 3 (2020), e201960147. doi: [10.1002/jbio.201960147](https://doi.org/10.1002/jbio.201960147). 2

[ZSL*22] ZHANG Y., SUN Y., LI H., ZHENG S., ZHU C., YANG L.: Benchmarking the Robustness of Deep Neural Networks to Common Corruptions in Digital Pathology. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022* (Cham, 2022), Wang L., Dou Q., Fletcher P. T., Speidel S., Li S., (Eds.), Lecture Notes in Computer Science, Springer Nature Switzerland, pp. 242–252. doi: [10.1007/978-3-031-16434-7_24](https://doi.org/10.1007/978-3-031-16434-7_24). 7

[ZWS*23] ZHANG Y., WEI D., SCHALEK R., WU Y., TURNER S., LICHTMAN J., PFISTER H., FU Y.: High-Throughput Microscopy Image Deblurring with Graph Reasoning Attention Network. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)* (Apr. 2023), pp. 1–5. doi: [10.1109/ISBI53787.2023.10230473](https://doi.org/10.1109/ISBI53787.2023.10230473). 2

[ZY21] ZHANG H., YAO J.: Automatic focusing method of microscopes based on image processing. *Mathematical Problems in Engineering* 2021, 1 (2021), 8243072. doi: [10.1155/2021/8243072](https://doi.org/10.1155/2021/8243072). 3

[ZZY*24] ZHANG Y., ZHENG P., YAN W., FANG C., CHENG S. S.: A unified framework for microscopy defocus deblur with multi-pyramid transformer and contrastive learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2024), pp. 11125–11136. 2, 3, 7