

Article

Ambiguity-Reduced Depth from Defocus via Single-Shot Polarizer-Free Dual-Focus Imaging

Wenjie Lai ¹ , Fanyu Zeng ², Xiao Hu ¹, Shaowei He ¹, Ziji Liu ^{1,*}, Huiling Tai ^{1,*} and Yadong Jiang ¹

¹ School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610051, China; wjlai@uestc.edu.cn (W.L.)

² School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China

* Correspondence: zjliu@uestc.edu.cn (Z.L.); taitai1980@uestc.edu.cn (H.T.)

Abstract

Single-image depth from defocus is limited by blur-radius ambiguity: two object distances on different sides of the focal plane can produce similar point spread functions (PSFs). We study single-shot dual-focus imaging (DFI) with a polarizer-free liquid crystal (LC) lens, where the ordinary-ray component remains unmodulated while the extraordinary-ray component is refocused. Unlike conventional multi-capture DFD, the proposed system records a dual-focus superposition in one exposure and selects the optical setting before network training using the ambiguity-interval length A and PSF correlation C_r . We show that DFI does not reduce ambiguity unconditionally: its benefit depends on the LC-lens power. A deblurring-based depth estimation network with a physics-calibrated Wiener bank translates the selected dual-focus cue into quantitative depth. On model-matched synthetic DFI data with an 8 m focus setting, the proposed configuration reduces RMS error from 0.227 ± 0.001 m to 0.190 ± 0.001 m (mean $\pm$ s.d. over four independent runs) relative to single-focus imaging. A 116-pair indoor prototype dataset provides feasibility evidence under the tested configuration, but is not used to claim broad physical generalization.

Keywords: depth from defocus; liquid crystal lens; single-shot dual-focus imaging; point spread function; Wiener deconvolution

1. Introduction

Depth from defocus (DFD) estimates scene depth from defocus blur and is attractive because the cue is encoded directly by the optics, without a second viewpoint or active illumination. It has been studied for decades with both conventional optics [1–5] and, more recently, learned estimators [6–11]. Multi-focus or dual-image DFD can reduce depth ambiguity by comparing images captured at different focus settings [1,3,7], but multiple captures introduce registration, magnification, and dynamic-scene limitations. Single-image DFD avoids these acquisition problems [12,13], yet a conventional single-focus image can map two depths to the same blur radius, making the PSF insufficiently discriminative near symmetric positions around the focal plane.

Several strategies modify the optics to mitigate this single-focus ambiguity. Coded apertures reshape the defocus PSF to make blur more depth-discriminative [14], and end-to-end “deep optics” jointly optimize a phase mask and a depth network [15]. Dual-pixel sensors provide a sub-aperture disparity that has been used for depth and defocus deblurring [16,17], while purely learned monocular estimators infer depth from semantic



Received: 10 July 2026

Revised: 27 July 2026

Accepted: 27 July 2026

Published: 31 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

context without an explicit blur model [18,19]. Dual-focus imaging is a complementary route in which two focus settings are combined optically, as in plasmonic dual-focus lenses [20] and dual-focus stereo and microscopy systems [21,22]. Liquid crystal (LC) lenses are a natural way to realize an electrically tunable focus for DFD [23–25], since the focal length is switched by voltage without moving parts. Recent LC-integrated and LC-enabled terahertz metalenses further demonstrate electrically switchable single/bifocal and varifocal achromatic operation [26,27]; these devices broaden the context of tunable LC optics but are not direct DFD baselines. Our work builds on tunable LC optics but removes the polarizer, so a single exposure superposes two focus settings instead of capturing them sequentially.

We address the single-focus ambiguity with single-shot DFI using a polarizer-free LC lens. Unlike conventional polarizer-based LC-lens imaging, the proposed acquisition keeps both ordinary-ray (*o*-ray) and extraordinary-ray (*e*-ray) components in one exposure. The polarizer-free LC lens is placed in front of the main lens and produces a radially varying effective refractive-index profile for the extraordinary-ray component. Natural unpolarized light is therefore decomposed into two copropagating components: the *o*-ray, which sees a uniform ordinary refractive index and passes through essentially unmodulated, and the *e*-ray, which is refocused by the voltage-dependent LC phase profile. Consequently, the *o*-ray is imaged with the original focal length, while the *e*-ray is refocused to a different focal length determined by the applied LC voltage. The sensor records the incoherent superposition of the two components, so one exposure encodes two focus settings without a polarizer or multiple captures. Our earlier report introduced the polarizer-free acquisition concept and used part of the present apparatus for multi-state capture [28]. The LC lens used here was newly fabricated. The present study adds the single-shot DFI formulation, the A/C_r design criteria and LC-power sweep, the physics-calibrated DBDENet pipeline, and the associated ambiguity, accuracy, and robustness analyses; a component-by-component comparison is provided in the Supplementary Materials.

The main point is not that dual focus always removes ambiguity. Instead, the second focus setting must be selected carefully. We introduce two criteria for this selection: the ambiguity-interval length A , which measures the depth region where ambiguity remains, and the PSF correlation coefficient C_r , which measures how separable the PSFs are across depths. A useful DFI configuration should reduce both A and C_r .

This article is organized around three contributions. First, we formulate a polarizer-free LC-lens system as a single-shot DFI camera for DFD, rather than as a multi-capture refocusing device. Second, we show that the useful operating range of DFI can be screened and ranked using the joint behavior of A and C_r , which turns the second focus setting into a quantitative optical-design indicator. Third, a compact deblurring-based network confirms that the encoded dual-focus cue improves depth estimation and that this improvement is observed across several learned estimator architectures. Together, the optical design and the physics-calibrated network form an integrated framework for single-shot DFI depth estimation.

It is useful to distinguish this goal from two adjacent lines of work. With respect to estimating depth from focus, the camera or lens is swept through multiple focal positions, and the depth is inferred from the focus response. This produces rich optical information, but it also requires multiple captures and is not well matched to dynamic scenes. In dual-image DFD, two separately captured images with different focus settings are compared [1,7], which directly addresses the single-focus ambiguity but still requires precise registration and is sensitive to scene motion. Our earlier polarizer-free LC-lens work [28] also falls into this multi-capture category, acquiring images at different lens states. The proposed DFI setting is more constrained: only one intensity image is available, and the two focus

components are superposed rather than separately observed. The benefit is single-shot acquisition without inter-frame registration; the cost is that the mixed PSF must be made sufficiently discriminative by optical design.

This distinction also explains why we avoid treating DFI as a purely algorithmic data augmentation. If the second focus setting is poorly selected, the additional e -ray component may provide little useful depth-dependent signature, making the resulting PSFs insufficiently distinguishable across depths. Conversely, if the second focus setting creates a complementary PSF, the same scene texture carries two defocus responses in one measurement. The role of A/C_r is to determine which of these cases applies before training a depth estimator.

2. Principle and Design Criteria

Figure 1 summarizes the optical principle used in the following formulation. In a single-focus image, depth is represented by one blur radius, so two object distances can become ambiguous when they produce the same radius around the focus plane. In the proposed DFI image, the o -ray preserves this original blur radius R_o , while the LC-modulated e -ray produces a second radius R_e at the same object distance. The depth cue is therefore the paired response (R_o, R_e) rather than either radius alone. If the LC power is selected properly, depths that are ambiguous in the o -ray branch are separated by the e -ray branch; if it is selected poorly, the paired response can remain highly correlated. This motivates the blur-radius and PSF-based criteria derived below.

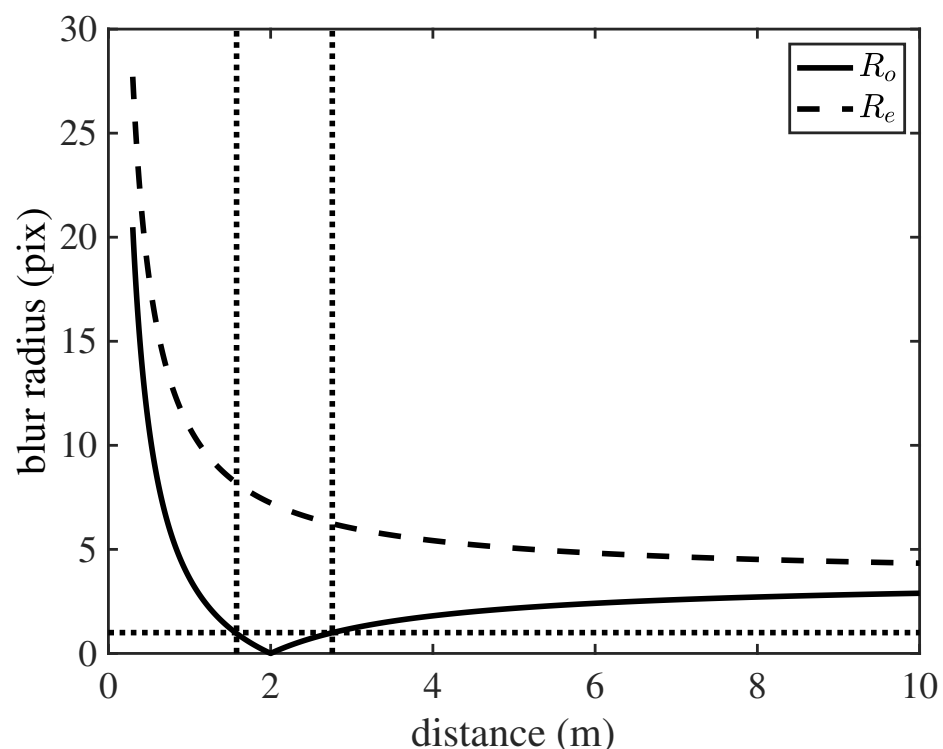


Figure 1. Principle of DFI-based ambiguity reduction. The o -ray keeps the original blur radius R_o , while the LC lens changes the e -ray blur radius R_e . The solid curve denotes R_o , the dashed curve denotes R_e , the vertical dotted lines mark two depths with the same o -ray blur radius, and the horizontal dotted line marks that shared blur-radius level. When selected properly, the pair (R_o, R_e) is more discriminative than a single blur radius.

For a main lens with focal length f and an LC lens power P_e , the focal length of the e -ray component is

$$f_e = \frac{f}{1 + fP_e}, \quad (1)$$

while the o -ray focal length remains $f_o = f$. The defocus PSF is approximated by a Gaussian kernel

$$K = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right), \quad \sigma = R/\sqrt{2}, \quad (2)$$

where R is the blur radius [29]. Under natural unpolarized illumination, the o -ray and e -ray components are approximately balanced. The DFI PSF is therefore modeled as

$$K_{\text{DFI}} = 0.5K_e + 0.5K_o. \quad (3)$$

The Gaussian kernel (following the standard Fourier-optics convention [30]) is a nominal near-axis approximation for the controlled configuration, where a radially symmetric blur-width model provides a tractable link between depth and the Wiener PSF bank. It is not a measured or exact model of the prototype PSF. Measured PSFs/MTFs, field dependence, aberrations, and polychromatic deviations were not available for the present device and remain limitations of the physical validation. The two blur radii are

$$\begin{aligned} R_e &= R_a D(1/f + P_e - 1/u - 1/D), \\ R_o &= R_a D(1/f - 1/u - 1/D), \end{aligned} \quad (4)$$

where u is the object distance, D is the distance from the imaging plane to the optical system center, and $R_a = f/(2N)$ is the aperture radius with N the f-number.

The parameters in Equation (4) have distinct practical roles. The focal length f and aperture radius R_a are fixed by the main lens, while D is obtained through camera calibration. After calibration, changing P_e electronically changes only the e -ray focus and leaves the o -ray component as a built-in reference focus. This is the main acquisition advantage over mechanical dual-focus capture: the two focus components are recorded in the same exposure, so scene motion and inter-frame registration are avoided. The trade-off is that the image is a superposed measurement, so the selected LC power must make the combined PSF sufficiently informative.

The 50/50 weighting in Equation (3) follows from the use of unpolarized illumination in the ideal polarizer-free setting. In a real system, birefringence, transmission loss, and surface reflection can shift this ratio. Therefore, the 50/50 model should be interpreted as the nominal design point, and nonideal o/e ratios are treated as a robustness factor in the supplementary experiments.

The blur-radius curves provide an intuitive interpretation of A . In a conventional single-focus image, two depths at opposite sides of the focal plane can yield the same blur radius. When the LC lens changes only the e -ray focal length, the ambiguous pair under the o -ray branch will generally have different e -ray radii. If this difference is large enough, the combined PSF becomes distinguishable. If the second focus lies in an unfavorable position, however, the e -ray branch can introduce another near-symmetry, and the ambiguous interval may persist or even expand. This is why we explicitly include a failure case in the experiments rather than reporting only the best LC power.

The PSF-correlation criterion captures a different issue. Even when two depths are not exactly ambiguous, their PSFs can remain very similar. Learned estimators can exploit image context, but if the physical PSFs are highly correlated over a wide range, the depth cue provided by defocus remains weak. A lower C_r indicates that small changes in depth lead to more visible changes in the combined blur kernel. In practice, A is a coarse ambiguity

measure, and C_r is a continuous discriminability measure. The two should therefore be interpreted together: A rejects obviously poor dual-focus settings, while C_r ranks settings that have similar ambiguity intervals.

The design process is then straightforward. For a fixed main lens and sensor, we sweep feasible LC powers, compute R_o and R_e over the operating depth range, synthesize the corresponding PSFs, and evaluate A and C_r on the same working depth interval. Candidate powers with large ambiguity intervals are removed first. Among the remaining powers, the setting with lower PSF correlation is preferred, provided that the blur magnitude is still compatible with the target image resolution and signal-to-noise ratio. In our LC calibration, the voltage-to-optical-power response is approximately linear over the normalized range $[-1, 1]$, which defines the admissible range used for implementation. This gives a physics-first rule for selecting the optical setting before training the depth estimator.

Figure 2 shows this sweep for the three focus distances used later. The ambiguity interval A behaves as a hard constraint: it is zero over a contiguous range of negative powers and then increases abruptly, so the first design step is simply to stay inside the $A = 0$ region. Within that region, C_r varies more smoothly and selects among the admissible powers. In our LC calibration, the voltage-to-optical-power response is approximately linear over the normalized range $[-1, 1]$; therefore, the admissible optimum is attained at the lower bound, $P = -1$. This setting sits in the $A = 0$ region with a near-minimal C_r for every focus distance, illustrating how the two criteria jointly fix the optical setting without trial-and-error training.

We evaluate an LC power by two quantities over the working depth range. First, A measures the cross-ray blur-radius ambiguity interval. Let d_{ambig} denote the boundary depth at which the cross-ray blur-radius overlap begins under the calibrated thin-lens model. The corresponding depth set is

$$\mathcal{B} = \begin{cases} [\max(d_{\text{ambig}}, d_{\min}), d_{\max}], & 0 < d_{\text{ambig}} < d_{\max}, \\ \emptyset, & \text{otherwise,} \end{cases} \quad (5)$$

and the ambiguity-interval length is the one-dimensional measure

$$A = \mu(\mathcal{B}). \quad (6)$$

where μ denotes a one-dimensional Lebesgue measure, i.e., the length along the depth axis. Thus, A reports the depth span within the evaluated working range over which the cross-ray blur-radius ranges overlap under the nominal 50/50 DFI model. The analytical expression for d_{ambig} is provided in the Supplementary Materials. Second, C_r measures PSF separability continuously. For two object distances u_i and u_j , we compute the normalized PSF correlation as

$$C_r(u_i, u_j) = \frac{\langle K_{\text{DFI}}(u_i), K_{\text{DFI}}(u_j) \rangle}{\|K_{\text{DFI}}(u_i)\|_2 \|K_{\text{DFI}}(u_j)\|_2}, \quad (7)$$

and report the average correlation over the evaluated depth pairs. A lower C_r means the PSFs change more strongly with depth and are easier for either a classical or learned estimator to separate.

Both criteria are needed. A zero or short ambiguity interval means that the paired blur radii remove the most obvious symmetric ambiguity, but the resulting PSFs may still be highly correlated over broad depth ranges. Conversely, a low PSF correlation is less useful if the optical setting creates a large ambiguous interval. Therefore, the preferred LC power should minimize A and keep C_r low. Figure 3 illustrates why this condition is necessary: the SFI ambiguity interval and the DFI interval for the suboptimal $P = 0.5$ setting are explicitly marked by -45°

and 45° hatching, respectively, showing that some LC powers reduce ambiguity whereas others can make the DFI image less informative than the single-focus image.

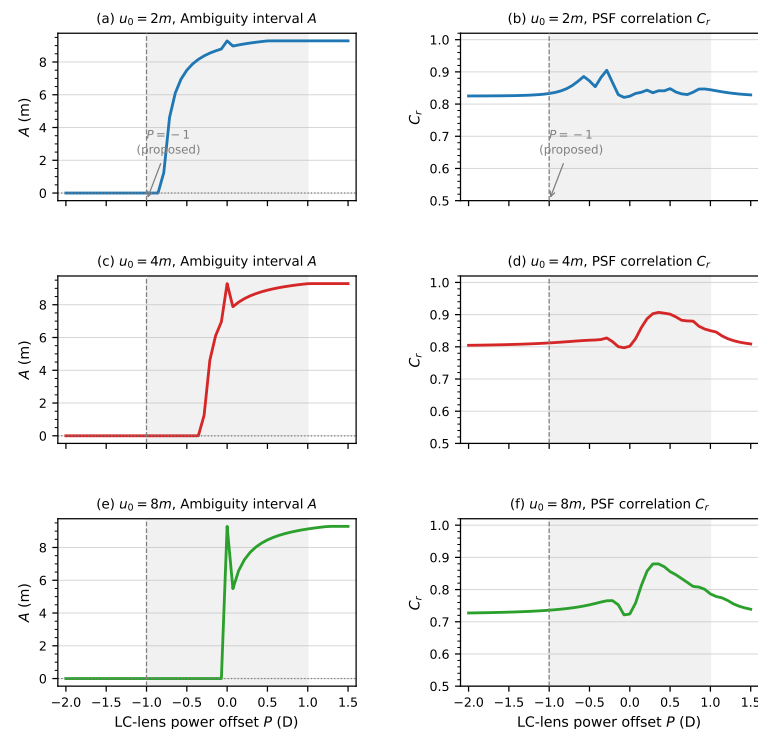


Figure 2. Ambiguity-interval length A (left) and PSF correlation C_r (right) as functions of the LC-lens power offset P , for focus distances $u_0 = 2, 4$, and 8 m. The differently colored curves correspond to the three focus distances. Both criteria are evaluated over the NYUv2 working depth range $[0.71, 10]$ m. The gray shaded band marks the calibrated linear LC operating range, $P \in [-1, 1]$, within which powers are considered admissible for implementation, and the vertical dashed line marks the selected setting $P = -1$. A exhibits a sharp transition: it stays at zero for sufficiently negative P and rises steeply once the e -ray focus crosses an unfavorable region. The proposed setting $P = -1$ keeps $A = 0$ while attaining a low C_r at all three focus distances, which is why it is selected before training the depth-estimation network.

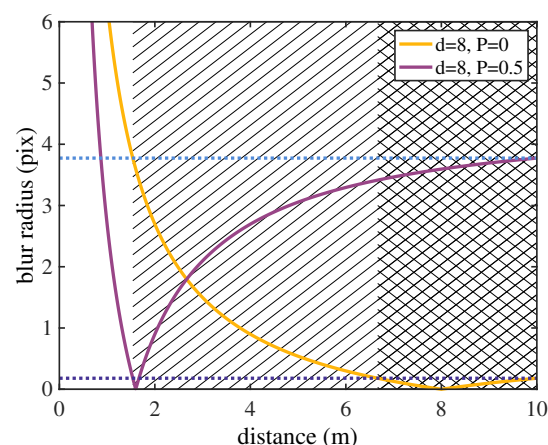


Figure 3. Ambiguity intervals for SFI and a suboptimal DFI setting at 8 m focus. The orange curve ($P = 0$) is the inactive-LC SFI baseline, and the purple curve ($P = 0.5$) is the DFI setting. The dotted lines indicate the corresponding blur-radius relationships used to locate ambiguous depth intervals. The black diagonal hatching distinguishes the marked ambiguity regions: the -45° hatched region marks the SFI ambiguity interval, while the 45° hatched region marks the DFI ambiguity interval introduced by the suboptimal second focus. This example shows that DFI should be treated as an optical design problem rather than a universal ambiguity-removal mechanism.

3. Depth Estimator

To translate the DFI design into quantitative depth estimates, we use the joint optical–computational pipeline shown in Figure 4. The optical criteria are evaluated before network training: A screens out LC powers that retain cross-ray blur-radius ambiguity, and C_r ranks PSF separability among the feasible settings. The selected power and its calibrated depth-dependent PSFs then define the Wiener candidate bank used by DBDENet. For $N = 32$ candidate depths d_i , PSFs are computed from Equations (2)–(4). The physics-calibrated Wiener deconvolution bank produces a stack of deblurred images,

$$I_i = F^{-1} \left(\frac{F(K_i)^* F(x_b)}{|F(K_i)|^2 + \lambda} \right), \quad (8)$$

where x_b is the defocused image, F and F^{-1} denote the Fourier and inverse Fourier transforms, respectively, $F(K_i)^*$ is the complex conjugate of the PSF spectrum, and $\lambda = 0.005$ is the Wiener regularization coefficient. We consider N candidate depths, indexed by $i \in \{1, \dots, N\}$. For each candidate depth d_i , the corresponding PSF K_i is used to obtain a deblurred plane I_i . The weights w_i shown in Figure 4 are normalized over all N candidates. The selected values are supported by an ablation and a sensitivity analysis: $N = 32$ reduces RMS from 0.216 to 0.189 m (12.5%) relative to raw input ($N = 0$), and $\lambda = 0.005$ is the minimum-error point in the tested regularization sweep. The full N and λ sweeps are reported in the Supplementary Materials.

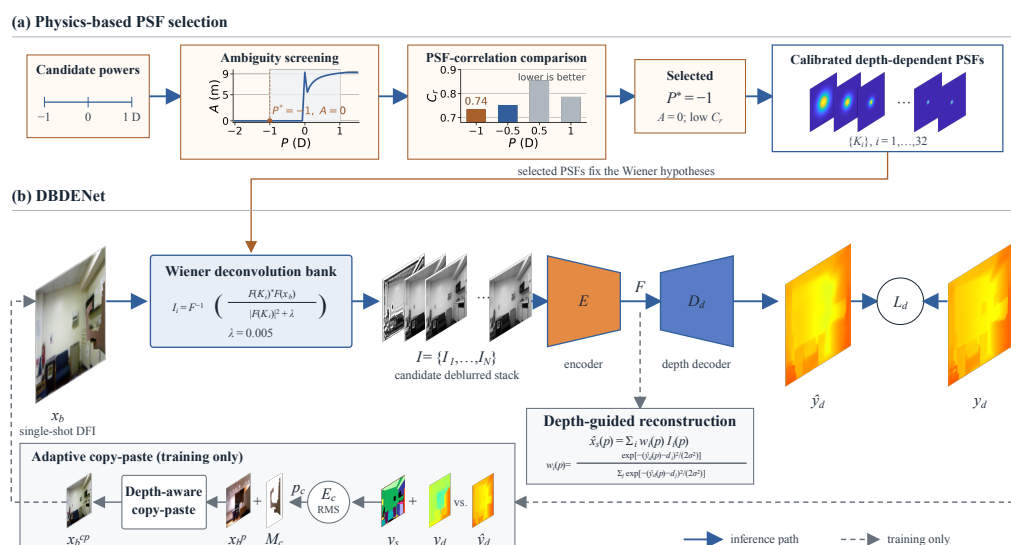


Figure 4. Joint optical-design and depth-estimation framework. (a) Candidate LC-lens powers are screened by the ambiguity interval A and ranked by the PSF correlation C_r . The inset plots show the criteria at the 8 m focus setting: $P = -1$ gives $A = 0$ and the lowest C_r among the compared admissible powers. Orange elements mark the selected optical-design path, blue elements mark the calibrated PSF bank and inference flow, and gray dashed arrows indicate training-only paths. This setting defines the calibrated depth-dependent PSF bank. (b) DBDENet applies the selected PSFs in a Wiener deconvolution bank, producing candidate deblurred images for encoder–decoder depth estimation. During training, the depth-guided reconstruction branch forms an all-in-focus estimate $\hat{x}_s$. The adaptive copy–paste branch builds on copy–paste augmentation [31] and computes category-wise RMS error E_c from $\hat{y}_d$, y_d , and semantic labels y_s ; the resulting sampling probability p_c selects semantic masks for depth-aware copy–paste augmentation. The asterisk in $F(K_i)^*$ denotes complex conjugation, as in Equation (8).

For a 480×640 input, the Wiener bank forms a $32 \times 480 \times 640$ candidate stack. A four-stage MSCAN-T encoder from SegNeXt [32] maps it to feature widths of 32, 64, 160,

and 256 channels, with stage depths of 3, 3, 5, and 2. A 256-channel LightHam decoder from SegNeXt [32], built around the Hamburger matrix-decomposition module [33], fuses the multiscale features and upsamples them to a one-channel full-resolution depth map. A final ReLU constrains predictions to non-negative depth; no upper clamp is applied. During training only, a depth-guided reconstruction branch converts the predicted depth into normalized Gaussian weights over the candidate planes and uses their weighted sum to reconstruct the all-in-focus target.

Training uses AdamW with initial learning rate 3×10^{-4} , polynomial decay (power 0.9) to 10^{-6} over 200 epochs, weight decay 0.1, gradient clipping at 1, a training batch size of 4, and horizontal-flip plus bounded additive-noise augmentation. The total objective is $L_d + 0.01L_b$, where L_d combines an L_1 depth term with a gradient term weighted by 0.001, and the all-in-focus image reconstruction loss L_b combines a Charbonnier penalty [34] and an SSIM term [35] with a relative SSIM weight of 4. Controlled runs use fixed seeds 1–4, deterministic algorithms, and the last-epoch checkpoint. Full equations and reproducibility settings are provided in the Supplementary Materials.

DBDNet provides a compact, physics-guided computational component of the proposed DFI framework. It explicitly converts optical priors into a candidate deblurring stack, so the downstream encoder can compare sharpness cues across candidate depths. In the current implementation, DBDNet has 19.10 M parameters and 37.81 G MACs and runs at 97.69 fps for 480×640 input on an RTX 4090; this end-to-end timing includes the Wiener deconvolution bank. It is therefore substantially lighter than heavier depth backbones such as iDisc [36] and DCDepth [37] while remaining competitive in depth accuracy. Controlled single-run ablations show that the 32-depth Wiener stack reduces RMS from 0.216 to 0.189 m relative to raw input ($N = 0$), and adaptive learning reduces RMS from 0.213 to 0.189 m. The reported Wiener-bank, candidate-depth, adaptive-learning, regularization-sensitivity, and accuracy–efficiency results are provided in the Supplementary Materials.

The candidate-depth design is also aligned with the optical model. Each deconvolution filter corresponds to one hypothesized object distance and the associated DFI PSF. If the hypothesized distance is close to the true depth, the deconvolved image tends to recover sharper local structure; if the hypothesis is wrong, residual blur or ringing remains. The encoder therefore receives a stack whose variation is physically tied to focus consistency. This differs from feeding the raw defocused image alone, where the network must infer the optical ambiguity implicitly.

We keep the network description compact in the main text because most implementation details are not specific to the DFI optical principle. The detailed loss definitions, reported component ablations and sensitivity results, and full method comparison including parameters, MACs, and speed are provided in the Supplementary Materials.

4. Experiments

Synthetic defocus images are generated from NYUv2 RGB-D pairs [38,39] following the standard NYUv2 monocular-depth protocol [40]; the split contains 795 training and 654 testing pairs. Unless stated otherwise, the main lens uses an f-number of 2.8, pixel size of $5.6 \mu\text{m}$, focal length of 15 mm, and candidate depth range of 0–10 m. After converting stored depth labels with the dataset scale factor $10/2550$, the NYUv2 ground-truth depths span approximately $[0.71, 10.0]$ m with a median of 2.5 m. The optical criteria A and C_r are therefore computed over $[0.71, 10]$ m, matching the actual dataset depth range while excluding the most saturated near-field PSFs. We also evaluate a small indoor real DFI dataset captured with a prototype using a camera from Guangzhou Micro-shot Technology Co., Ltd. (Guangzhou, China), a self-fabricated LC lens, and an Intel RealSense LiDAR Camera L515 (Intel Corporation, Santa Clara, CA, USA) for depth acquisition [28]. It

contains 116 RGB-D pairs, with 78 for training and 38 for testing, so conclusions from real data are limited to this prototype and indoor configuration. All learned methods are trained separately for the SFI and DFI settings under matched data splits. The main text reports a compact metric set, RMS error, and threshold accuracy δ_1 , with extended comparison tables provided in the Supplementary Materials. RMS is emphasized because the reviewed task is metric depth from defocus, and the main optical claim is that the DFI PSF makes depth errors smaller near ambiguity-prone regions. For the LC-power study, each setting is generated with the same scene split and network configuration so that the observed changes can be attributed to the optical design rather than to a training protocol change.

The real prototype data are used more conservatively. They provide evidence that the single-shot DFI signal is physically measurable and useful, but the dataset is too small to support a broad claim about scene generalization. For this reason, the real-data table is reported as a compact feasibility result, and the limitations section explicitly lists the robustness tests needed before stronger deployment claims.

The synthetic and real evaluations play different roles. The synthetic data isolate the optical variable because the all-in-focus images, depth maps, and camera parameters are controlled. This makes it possible to compare SFI and DFI under the same scene content and training protocol and to attribute the performance trend to the PSF design. At the same time, the synthetic setting is a calibrated, model-matched validation: the Gaussian PSF family used to render the data is consistent with the PSF model used by the Wiener deconvolution bank. Therefore, the synthetic gains should be interpreted as evidence that the selected DFI encoding is beneficial under ideal calibrated optics, rather than as a complete test of physical generalization. The real prototype data introduce uncontrolled effects such as sensor noise, imperfect LC-lens response, calibration residuals, and deviations from the Gaussian PSF assumption. We therefore use the real results as an initial mismatch check to test whether the DFI signal survives beyond the model-matched synthetic setting. The computational experiments were implemented in Python 3.10 using PyTorch (<https://pytorch.org>, accessed on 26 July 2026). Runtime was measured on an NVIDIA GeForce RTX 4090 GPU (NVIDIA Corporation, Santa Clara, CA, USA).

The focus-distance study is another part of the optical validation. At 2 m, 4 m, and 8 m, the same LC power does not produce identical blur-radius geometry because the relation among focus distance, object depth, and sensor position changes. The observation that $P = -1$ remains favorable in the tested configurations suggests that the selected second focus is broadly compatible with the simulated indoor depth range. However, this should not be generalized to arbitrary cameras or target distances. A different aperture, pixel size, lens focal length, or working range would require recomputing A and C_r .

The failure at 8 m with $P = 0.5$ is particularly important for interpreting the results. It shows that adding a second focus component is not equivalent to adding more information in a monotonic sense. The added component can be redundant or even confusing if it increases PSF similarity in the relevant depth range. This supports this paper's central recommendation: DFI should be designed with an explicit optical criterion, and network training should be performed after the optical setting has passed this criterion.

Table 1 shows the compact comparison between DFI and SFI. For DBDENet, DFI reduces RMS error from 0.227 ± 0.001 m to 0.190 ± 0.001 m, corresponding to a 16.4% reduction. The same trend appears for D3-Net, 2HDED, iDisc, DCDepth, and FastDepth, with RMS reductions of 12.5–25.0% among the estimators in Table 1. This cross-network behavior is important because it indicates that the gain comes from the optical encoding rather than a network-specific tuning effect. Extended metrics are moved to Supplementary Materials to keep this article focused.

Figure 5 shows a representative synthetic example. Although the DFI input looks more blurred than the SFI input because of the added e -ray component, the depth map estimated from it follows the ground-truth structure more faithfully, particularly at object boundaries and in the relative ordering of foreground and background. This is the visual counterpart of the design criteria: the superposed defocus signature is more depth-discriminative even though it is less visually sharp.

Table 1. Representative DFI and SFI comparison on the synthetic dataset at 8 m focus. The main text uses the compact metric set RMS and δ_1 ; extended metrics are in the Supplementary Materials. Single-run results are shown for the comparison methods; [†] DBDNet values are mean $\pm$ s.d. over four independent runs.

Method	Input	RMS (m)	δ_1 (%)
D3-Net [8]	SFI	0.324	95.2
D3-Net [8]	DFI	0.243	97.8
2HDED [9]	SFI	0.256	97.9
2HDED [9]	DFI	0.199	98.9
iDisc [36]	SFI	0.223	98.5
iDisc [36]	DFI	0.184	98.8
DCDepth [37]	SFI	0.329	95.8
DCDepth [37]	DFI	0.288	97.3
FastDepth [41]	SFI	0.270	97.5
FastDepth [41]	DFI	0.221	98.3
DBDNet	SFI	0.227 ± 0.001 [†]	98.34 ± 0.02 [†]
DBDNet	DFI	0.190 ± 0.001 [†]	98.87 ± 0.03 [†]

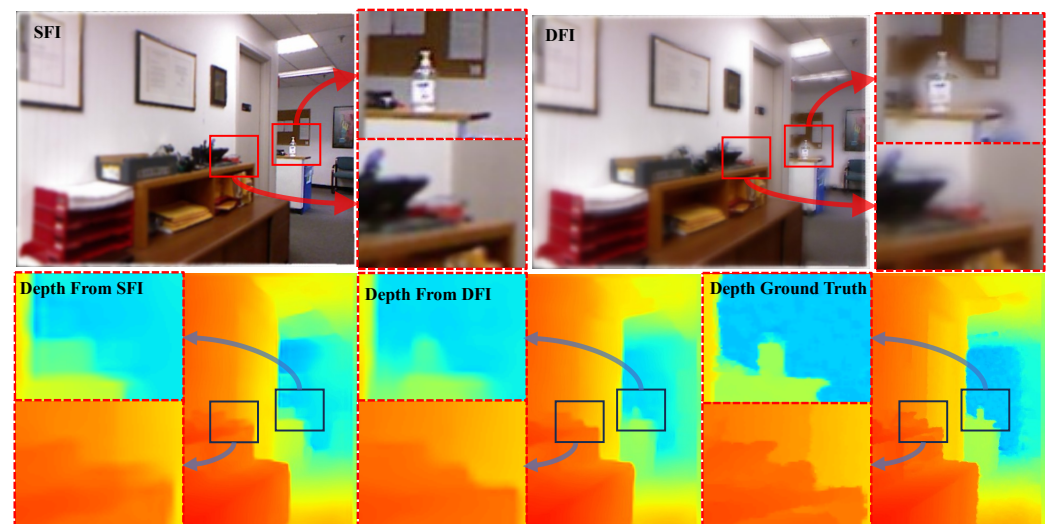


Figure 5. Qualitative comparison on synthetic data. (**Top**) the SFI and DFI inputs with zoomed insets; the DFI image carries the superposed o -/ e -ray defocus signature. (**Bottom**) depth maps estimated by DBDNet from the SFI and DFI inputs, with the ground truth. Colored depth maps encode distance according to the displayed depth colors, and arrows/squares mark regions used for local visual comparison. The DFI result recovers depth boundaries and relative ordering (highlighted regions) closer to the ground truth than the SFI result, consistent with the quantitative gain in Table 1.

Table 2 verifies that DFI is sensitive to the second focus setting. The LC power $P = -1$ gives the lowest RMS error for all evaluated focus distances and also yields zero ambiguity interval in this configuration. By contrast, at 8 m focus with $P = 0.5$, DFI increases RMS error from 0.227 m to 0.282 m under the same single-run sweep protocol. This failure

case is important: it supports the design-criteria view and shows that dual focus must be optically configured.

To test the criteria quantitatively rather than through individual cases, we sweep 12 LC-power configurations (three focus distances $\times$ four powers) and correlate A and C_r with the resulting depth accuracy (Figure 6). Across all configurations, C_r and RMS show a significant positive Spearman rank correlation ($\rho = 0.608$, $p = 0.036$), indicating that lower PSF correlation is generally associated with lower depth error. One exception appears in the unnormalized RMS metric at 8 m focus, but the original observation was confirmed by three additional controlled runs, and the same pair follows the criterion on the relative metrics. The full statistics and this counterexample are provided in the Supplementary Materials. Therefore, A and C_r , computed purely from the optical model, should be interpreted as optical screening and ranking indicators rather than as a guarantee of monotonic improvement in every metric. The small sample ($n = 12$) and the residual focus-distance confound also limit the strength of the correlation claim. On the real prototype dataset, DFI also improves the tested estimators in the current indoor setup. For DBDNet, the SFI input is processed with the ordinary-ray bank K_o , whereas the DFI input is processed with the corrected nominal bank $K_{DFI} = 0.5K_o + 0.5K_e$. In four controlled paired-seed runs evaluated with the last-epoch checkpoints, RMS decreases from 0.7424 ± 0.0096 m with SFI to 0.7040 ± 0.0160 m with DFI, and δ_1 increases from $61.28 \pm 1.79\%$ to $63.47 \pm 0.32\%$. The RMS improvement is positive in all four seeds, with a mean paired reduction of 0.0384 m. Since the real dataset is small, we use it only as a feasibility check and report the per-seed values, extended metrics, depth-bin RMS differences, and descriptive failure-analysis cases in the Supplementary Materials.

Table 3 summarizes the real-data trend under the same compact metric set used in the synthetic comparison. The improvement is smaller and less uniform than in the synthetic data, which is expected because the real prototype introduces calibration error, sensor noise, nonideal PSFs, and imperfect depth labels. We therefore avoid claiming broad real-world generalization from this small dataset. Instead, the result shows that the single-shot polarizer-free DFI signal remains usable in a physical indoor setup and motivates a larger capture campaign.

Table 2. LC lens power, ambiguity criteria, and RMS error for DBDNet. P is the LC lens power in diopters; $P = 0$ is the inactive-lens (SFI) baseline. The tested nonzero powers lie within the calibrated linear operating range $[-1, 1]$. A is the ambiguity-interval length in meters. Lower A and C_r values are preferred. Both criteria are computed over $[0.71, 10]$ m. Single-run RMS values are shown for a matched power-sweep comparison; multi-run statistics for the selected 8 m SFI/DFI pair are reported in Table 1. A dash indicates that A and C_r are not applicable to the inactive-lens baseline.

Focus	P	RMS (m)	A	C_r
2 m	0	0.248	-	-
2 m	-1	0.223	0	0.833
2 m	0.5	0.239	9.29	0.848
4 m	0	0.299	-	-
4 m	-1	0.205	0	0.812
4 m	0.5	0.288	8.89	0.901
8 m	0	0.227	-	-
8 m	-1	0.189	0	0.736
8 m	0.5	0.282	8.46	0.856

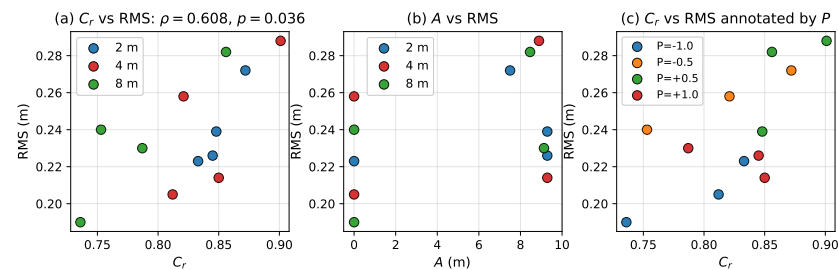


Figure 6. Optical design criteria versus depth estimation error over 12 LC-power configurations (three focus distances $\times$ four powers $P \in \{-1, -0.5, 0.5, 1\}$). **(a)** C_r versus RMS, colored by focus distance, with the global Spearman rank correlation. **(b)** A versus RMS. **(c)** C_r versus RMS annotated by P . Different colors distinguish the 2, 4, and 8 m focus settings, and text labels in (c) give the LC-lens power. Lower C_r is generally associated with lower error, supporting its use as a pre-training optical design indicator.

Table 3. Compact real-data comparison on the indoor prototype dataset. The main text uses the same RMS and δ_1 metric set as Table 1. The dataset contains 116 RGB-D pairs, so these results support feasibility rather than broad generalization. [†] DBDNet is mean $\pm$ s.d. over four controlled paired seeds.

Method	SFI RMS	DFI RMS	SFI δ_1	DFI δ_1
D3-Net [8]	0.938	0.878	52.3	54.0
2HDED [9]	0.996	0.778	50.6	60.4
iDisc [36]	0.695	0.620	63.7	66.9
DBDNet [†]	0.7424 ± 0.0096	0.7040 ± 0.0160	61.28 ± 1.79	63.47 ± 0.32

5. Robustness Considerations

The optical formulation assumes calibrated values of f , R_a , and image distance D , a Gaussian PSF approximation, and a near 50/50 o/e contribution under unpolarized illumination. These assumptions are reasonable for the controlled prototype but should be stress-tested before making stronger claims. The Supplementary Materials therefore report perturbation studies on both synthetic and real data. On the synthetic benchmark, we perturb the aperture radius (R_a) from -20% to $+20\%$, the focal length (f) and object-side focus setting (u_0 , through the resulting image distance D) by $\pm 10\%$, the nonideal o/e ratio from 30/70 to 70/30, and the Wiener regularization coefficient λ . On the real prototype dataset, we hold the captured observations and trained model unchanged while applying matched reconstruction-bank perturbations to PSF width, f , u_0 , and the o/e ratio from 30/70 to 70/30, with a diagnostic continuation to the pure- K_o limit. Overall, the results show stability under the tested moderate single-parameter perturbations; they do not establish general robustness to severe or combined physical mismatch.

For the synthetic calibration study, DBDNet remains insensitive to aperture-radius mismatch up to $\pm 10\%$: RMS changes by $+2.3\%$ for $R_a - 10\%$ and $+2.5\%$ for $R_a + 10\%$. The largest synthetic single-parameter calibration degradation is caused by $f + 10\%$, which increases RMS by 17.0% . The Wiener regularization sweep gives a clear optimum at $\lambda = 0.005$; using $\lambda = 0.001$ or 0.01 raises RMS by 50.2% and 19.6% , respectively.

We additionally held the 38 real test images and the trained DFI model fixed while perturbing the nominal $0.5K_o + 0.5K_e$ Wiener bank. Across width scales 0.80 – 1.20 , RMS ranges from 0.6918 to 0.7123 m, compared with 0.6954 m at the nominal scale. Perturbing f by $\pm 10\%$ gives RMS values of 0.6936 and 0.7134 m, while perturbing u_0 by $\pm 10\%$ gives 0.6984 and 0.6942 m. For the ordinary-ray fraction $\alpha = 0.30$ – 1.00 , RMS reaches its lowest value near $\alpha = 0.60$ and rises to 0.7224 m at the pure- K_o endpoint, so the ordinary-only reconstruction bank is not the best fixed-bank choice for captured DFI observations. Because

the captured images and network weights are unchanged throughout this sweep, these results summarize inference-time sensitivity to the assumed reconstruction-bank mixture.

These observations suggest a practical deployment procedure. First, calibrate the camera and LC-lens response and estimate the expected operating depth range. Second, compute A and C_r for feasible LC powers under the calibrated model. Third, run robustness checks by perturbing calibration parameters and the o/e ratio before finalizing the optical configuration and collecting data for depth-estimator training. Finally, train or fine-tune the depth estimator using the selected optical configuration. This workflow keeps the optical design and learning system coupled, while avoiding a blind search over lens powers.

6. Discussion

The results explain why dual focus is useful only after optical screening. The ambiguity interval A acts as a feasibility test: a nonzero value identifies depths that retain cross-ray blur-radius ambiguity. Among admissible settings, C_r ranks how distinctly the complete normalized PSF changes across depth. Their roles are complementary rather than interchangeable, and the observed 8 m counterexample confirms that neither quantity should be interpreted as a universal predictor of every error metric. In the tested calibrated range, $P = -1$ is selected before training because it satisfies $A = 0$ and gives a low C_r across all three focus settings.

Relative to multi-capture DFD, the proposed route removes inter-frame registration and scene-motion errors by recording the two focus components in one exposure, at the cost of observing only their superposition. Dual-pixel sensing instead exploits two sub-aperture views and thus requires a compatible sensor. Coded apertures and deep optics create depth-dependent PSFs through a fixed or jointly optimized mask, whereas the LC lens provides an electrically selectable focus offset while preserving a conventional reference-focus component. LC-integrated terahertz metalenses [26,27] likewise show the broader potential of tunable LC optics but operate in different spectral and device regimes and are not competing depth-estimation methods.

The synthetic and real results support different conclusions. The synthetic study isolates the optical design variable and shows consistent DFI gains across multiple estimators under model-matched Gaussian rendering and Wiener preprocessing. The 116-pair real prototype dataset, four paired-seed DBDNet runs, depth-bin RMS analysis, descriptive failure-analysis cases, and PSF-bank perturbations show that the cue remains usable under the tested indoor device configuration. The small dataset precludes a claim of broad real-world generalization.

The main physical limitations are the nominal Gaussian/50:50 o/e PSF model and the absence of measured PSFs/MTFs for the prototype. The real-data width, f , u_0 , and o/e sweeps test reconstruction-bank sensitivity on fixed observations; broader deployment still requires measured PSF/MTF characterization and validation on larger, more diverse real datasets.

7. Conclusions

We presented a single-shot DFI scheme based on a polarizer-free LC lens for depth from defocus. The key result is that dual focus should be designed through the ambiguity interval A and PSF correlation C_r ; it reduces ambiguity only when the LC-lens power is selected appropriately. A compact deblurring-based depth network exploits this optical encoding and demonstrates improved depth estimation under model-matched synthetic optics, while the small 116-pair indoor dataset provides prototype feasibility evidence. The perturbation studies quantify bounded but asymmetric sensitivity to the tested single-

parameter mismatches; broader deployment requires measured PSF/MTF characterization and validation on larger, more diverse real datasets.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/photonics13080732/s1>: Supplementary Materials, Sections S1–S20: DFI prototype and optical path; DFI PSF expansion; DBDENet and losses; Experimental settings and reproducibility; Adaptive learning; Synthetic defocus dataset generation; Evaluation metrics; Calibration parameter sensitivity; Matched PSF-bank perturbations on real observations; Run-to-run variability; Adaptive learning ablation; Extended SFI/DFI comparison tables; Spearman correlation between design criteria and accuracy; Method comparison and computational cost; Wiener candidate-depth ablation; o/e polarization mixing-ratio sensitivity; Wiener regularization sensitivity; Previous work versus the present study; Limitations; Interval estimates for the real-data RMS comparison. Figures S1–S7: DFI camera setup; Optical path of the DFI setup; Adaptive learning visualization; Qualitative comparison on synthetic DFI and SFI datasets across focal distances; Qualitative comparison on the real indoor prototype dataset; Representative real-prototype cases for failure analysis; Depth distribution of the NYUv2 training and test set. Tables S1–S26: Compact DBDENet architecture; Principal settings used for the reported DBDENet experiments; Calibration sensitivity under aperture-radius perturbation; Calibration sensitivity under focal-length/focus-setting perturbations; Fixed-checkpoint sensitivity to PSF-bank width mismatch on real prototype images; Fixed-checkpoint real-data sensitivity to focal length and focus setting; Fixed-checkpoint real-data sensitivity to ordinary/extraordinary mixing ratio; DBDENet DFI four-run variability; DBDENet SFI four-run variability; Adaptive learning ablation; Extended synthetic DFI/SFI metrics; Extended real DFI/SFI metrics; Controlled paired-seed real prototype results; DBDENet across 2 m/4 m/8 m focus distances; Real-data RMS difference by depth quartile; Extended LC-power comparison; Spearman rank correlation between C_r and depth metrics; Three-run recheck of the 8 m $P = -0.5$ vs. $P = +1$ pair; Method comparison on synthetic DFI dataset; Method comparison on real DFI dataset; Computational cost; Wiener candidate-depth ablation; Polarization mixing-ratio sensitivity; Wiener regularization sensitivity; Relationship between earlier report and present study; Paired interval estimates for real prototype RMS comparison.

Author Contributions: Conceptualization, W.L.; methodology, W.L.; software, W.L.; investigation, W.L. and F.Z.; data curation, X.H. and S.H.; writing—original draft preparation, W.L.; writing—review and editing, F.Z., Z.L., H.T. and Y.J.; supervision, Z.L., H.T. and Y.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Major Science and Technology Project of Henan Province, grant number 251100220100; the National Science Fund for Distinguished Young Scholars, grant number 62225106; the Sichuan Innovation Research Group Project, grant number 2025NSFTD0008; and the Postdoctoral Fellowship Program of the China Postdoctoral Science Foundation, grant number GZC20250551.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings of this study are not publicly available at this time but are available from the corresponding authors upon reasonable request. The synthetic dataset was derived from the publicly available NYU Depth V2 dataset as described in this manuscript.

Acknowledgments: The authors thank Wei-Ming Zhu for support and guidance.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

DFD	Depth from defocus
DFI	Dual-focus imaging
SFI	Single-focus imaging
LC	Liquid crystal
PSF	Point spread function
RMS	Root mean square
AL	Adaptive learning

References

1. Watanabe, M.; Nayar, S.K. Rational filters for passive depth from defocus. *Int. J. Comput. Vis.* **1998**, *27*, 203–225. [\[CrossRef\]](#)
2. Horn, B.K.P. *Focusing*. AI Memo AIM-160; MIT Artificial Intelligence Laboratory: Cambridge, MA, USA, 1968.
3. Xiong, Y.; Shafer, S.A. Depth from focusing and defocusing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; IEEE: Piscataway, NJ, USA, 1993; pp. 68–73.
4. Schechner, Y.Y.; Kiryati, N. The optimal axial interval in estimating depth from defocus. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*; IEEE: Piscataway, NJ, USA, 1999; Volume 2, pp. 843–848.
5. Schechner, Y.Y.; Kiryati, N. Depth from defocus vs. stereo: How different really are they? *Int. J. Comput. Vis.* **2000**, *39*, 141–162. [\[CrossRef\]](#)
6. Levin, A.; Fergus, R.; Durand, F.; Freeman, W.T. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph. (TOG)* **2007**, *26*, 70. [\[CrossRef\]](#)
7. Tang, H.; Cohen, S.; Price, B.; Schiller, S.; Kutulakos, K.N. Depth from defocus in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: Piscataway, NJ, USA, 2017; pp. 4773–4781. [\[CrossRef\]](#)
8. Carvalho, M.; Le Saux, B.; Trounev-Peloux, P.; Almansa, A.; Champagnat, F. Deep depth from defocus: How can defocus blur improve 3D estimation using dense neural networks? In *Proceedings of the Computer Vision—ECCV 2018 Workshops, Munich, Germany, 8–14 September 2018*; Springer: Cham, Switzerland, 2019; pp. 307–323. [\[CrossRef\]](#)
9. Nazir, S.; Vaquero, L.; Mucientes, M.; Brea, V.M.; Coltuc, D. Depth estimation and image restoration by deep learning from defocused images. *IEEE Trans. Comput. Imaging* **2023**, *9*, 607–619. [\[CrossRef\]](#)
10. Gur, S.; Wolf, L. Single image depth estimation trained via depth from defocus cues. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, 15–20 June 2019; pp. 7683–7692.
11. Si, H.; Zhao, B.; Wang, D.; Gao, Y.; Chen, M.; Wang, Z.; Li, X. Fully Self-Supervised Depth Estimation from Defocus Clue. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, BC, Canada, 18–22 June 2023; pp. 9140–9149.
12. Anwar, S.; Hayder, Z.; Porikli, F. Deblur and deep depth from single defocus image. *Mach. Vis. Appl.* **2021**, *32*, 34.. [\[CrossRef\]](#)
13. Anwar, S.; Hayder, Z.; Porikli, F. Depth estimation and blur removal from a single out-of-focus image. In *Proceedings of the British Machine Vision Conference (BMVC)*, London, UK, 4–7 September 2017; pp. 113.1–113.12. [\[CrossRef\]](#)
14. Zhou, C.; Lin, S.; Nayar, S.K. Coded aperture pairs for depth from defocus and defocus deblurring. *Int. J. Comput. Vis.* **2011**, *93*, 53–72.
15. Ikoma, H.; Nguyen, C.M.; Metzler, C.A.; Peng, Y.; Wetzstein, G. Depth from defocus with learned optics for imaging and occlusion-aware depth estimation. In *Proceedings of the 2021 IEEE International Conference on Computational Photography (ICCP)*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–12.
16. Garg, R.; Wadhwa, N.; Ansari, S.; Barron, J.T. Learning Single Camera Depth Estimation Using Dual-Pixels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Republic of Korea, 27 October–2 November 2019.
17. Pan, L.; Chowdhury, S.; Hartley, R.; Liu, M.; Zhang, H.; Li, H. Dual Pixel Exploration: Simultaneous Depth Estimation and Image Restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Virtual, 19–25 June 2021; pp. 4340–4349.
18. Bhoi, A. Monocular depth estimation: A survey. *arXiv* **2019**, arXiv:1901.09402.
19. Godard, C.; Mac Aodha, O.; Firman, M.; Brostow, G.J. Digging into self-supervised monocular depth estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3828–3838.
20. Williams, C.; Montelongo, Y.; Wilkinson, T.D. Plasmonic Metalens for Narrowband Dual-Focus Imaging. *Adv. Opt. Mater.* **2017**, *5*, 1700811. [\[CrossRef\]](#)
21. Li, F.; Sun, J.; Wang, J.; Yu, J. Dual-focus stereo imaging. *J. Electron. Imaging* **2010**, *19*, 043009. [\[CrossRef\]](#)
22. Watanabe, T.M.; Sato, T.; Gonda, K.; Higuchi, H. Three-dimensional nanometry of vesicle transport in living cells using dual-focus imaging optics. *Biochem. Biophys. Res. Commun.* **2007**, *359*, 1–7. [\[CrossRef\]](#) [\[PubMed\]](#)

23. Bao, R.; Cui, C.; Yu, S.; Mai, H.; Gong, X.; Ye, M. Polarizer-free imaging of liquid crystal lens. *Opt. Express* **2014**, *22*, 19824–19830. [[CrossRef](#)] [[PubMed](#)]
24. Ye, M.; Chen, X.; Li, Q.; Zeng, J.; Yu, S. Depth from defocus measurement method based on liquid crystal lens. *Opt. Express* **2018**, *26*, 28413–28420. [[CrossRef](#)] [[PubMed](#)]
25. Zheng, L.; Chen, X.; Zhang, Y.; Ye, M. Influence of the maximum blur radius on depth sensor based on liquid crystal lens. *Photonic Sens.* **2021**, *11*, 371–375. [[CrossRef](#)]
26. Wang, Q.; Ge, S.; Yu, H.; Hu, W. Liquid Crystal Integrated Dynamic Terahertz Metalenses. *Laser Photonics Rev.* **2024**, *18*, 2400869. [[CrossRef](#)]
27. Wang, Q.; Xu, Y.; Guo, Z.; Chen, L.; Huang, C.; Hu, W. Liquid-Crystal-Enabled Varifocal Achromatic Metalens. *Laser Photonics Rev.* **2025**, *19*, e01901. [[CrossRef](#)]
28. Lai, W.; Liu, Z.; Sun, T.; Hu, X. Depth Estimation Using Polarizer-Free Liquid Crystal Lens. *Acta Opt. Sin.* **2023**, *43*, 1415002.
29. Subbarao, M.; Surya, G. Depth from defocus: A spatial domain approach. *Int. J. Comput. Vis.* **1994**, *13*, 271–294. [[CrossRef](#)]
30. Born, M.; Wolf, E. *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th expanded ed.; Cambridge University Press: Cambridge, UK, 1999.
31. Ghiasi, G.; Cui, Y.; Srinivas, A.; Qian, R.; Lin, T.-Y.; Cubuk, E.D.; Le, Q.V.; Zoph, B. Simple Copy-Paste is a strong data augmentation method for instance segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 2918–2928. [[CrossRef](#)]
32. Guo, M.H.; Lu, C.Z.; Hou, Q.; Liu, Z.; Cheng, M.M.; Hu, S.M. SegNeXt: Rethinking convolutional attention design for semantic segmentation. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 1140–1156. [[CrossRef](#)]
33. Geng, Z.; Guo, M.H.; Chen, H.; Li, X.; Wei, K.; Lin, Z. Is attention better than matrix decomposition? In Proceedings of the International Conference on Learning Representations, Virtual, 3–7 May 2021.
34. Charbonnier, P.; Blanc-Feraud, L.; Aubert, G.; Barlaud, M. Two deterministic half-quadratic regularization algorithms for computed imaging. In Proceedings of the 1st International Conference on Image Processing, Austin, TX, USA, 13–16 November 1994; Volume 2, pp. 168–172. [[CrossRef](#)]
35. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [[CrossRef](#)] [[PubMed](#)]
36. Piccinelli, L.; Sakaridis, C.; Yu, F. IDisc: Internal Discretization for Monocular Depth Estimation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 21477–21487.
37. Wang, K.; Yan, Z.; Fan, J.; Zhu, W.; Li, X.; Li, J.; Yang, J. DCDepth: Progressive monocular depth estimation in discrete cosine domain. *Proc. Adv. Neural Inf. Process. Syst.* **2024**, *37*, 64629–64648. [[CrossRef](#)]
38. Hasinoff, S.W.; Kutulakos, K.N. A Layer-Based Restoration Framework for Variable-Aperture Photography. In Proceedings of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil, 14–21 October 2007.
39. Silberman, N.; Hoiem, D.; Kohli, P.; Fergus, R. Indoor segmentation and support inference from RGBD images. In *Proceedings of the Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 746–760.
40. Eigen, D.; Puhersch, C.; Fergus, R. Depth map prediction from a single image using a multi-scale deep network. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; Volume 27, pp. 2366–2374.
41. Wofk, D.; Ma, F.; Yang, T.-J.; Karaman, S.; Sze, V. FastDepth: Fast Monocular Depth Estimation on Embedded Systems. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.