LF-BVN: Blind-View Network for Self-Supervised Light Field Denoising

Supplementary Material

1. Method

1.1. Geometric Invariance Mask Generation

This section details the generation process of our GIM.

To fully leverage the geometric invariance property of LF, we partition the set of all views into four subsets (S_A, S_B, S_C and S_D) such that each subset contains approximately one-quarter of the views. As shown in Fig. 2(a), the partition is designed to satisfy the following invariant: any view in subset S_B, S_C or S_D can be rotated by 90°, 180°, or 270°, respectively, to coincide with a view in subset S_A. As illustrated in Fig. 2(b), a specific view $\mathbf{I}_i$ in S_A maps to views $\mathbf{I}_i^{90}$, $\mathbf{I}_i^{180}$, and $\mathbf{I}_i^{270}$ in the other subsets after rotations of 90°, 180° and 270°, respectively. This allows the network to only mask and recover subset S_A during training, while still being able to denoise all views by applying the appropriate rotation during inference. However, such a view mask does not satisfy the principle of uniform blind-view coverage. To address this, we assign indices to the views within S_A and, by extension, to the corresponding views in S_B, S_C and S_D using the rotational correspondence. Instead of consistently masking the views in a single subset (e.g., S_A), we mask views on a rotating basis across the four subsets. This ensures that only one-quarter of the views in each subset are masked, guaranteeing a uniform distribution of blind views across the entire angular domain.

Algorithm 1 outlines the pseudo code for generating the GIM.

Algorithm 1: GIM Generation Algorithm

Input: LF's angular resolution n
Output: mask $M \in \{0,1\}^{n\times n}$
Init $M \leftarrow \mathbf{1}^{n\times n}$;
Init $S \leftarrow [1, 2, \ldots, \lfloor n^2/4 \rfloor]$;
$k \leftarrow \lfloor n/2 \rfloor$;
$A \leftarrow$ reshape$(S, (k, k+1))$;
for $i \leftarrow 0$ ***to*** k **do**
  for $j \leftarrow 0$ ***to*** $k-1$ **do**
    $m \leftarrow A[i,j] \bmod 4$
    $\hat{i} = k+(i-k)\cdot cos(\frac{\pi}{2}\cdot m)-(j-k)\cdot sin(\frac{\pi}{2}\cdot m)$
    $\hat{j} = k+(i-k)\cdot sin(\frac{\pi}{2}\cdot m)+(j-k)\cdot cos(\frac{\pi}{2}\cdot m)$
    $M[\hat{i},\hat{j}] \leftarrow 0$
$M[k,k] \leftarrow 0$;
return M;

1.2. Depth from Refocus

Fig. 5 demonstrates the depth map generated from the probability volume via the $soft-argmin$ [4] operation. The results indicate that our depth estimation module produces accurate depth predictions in most regions.

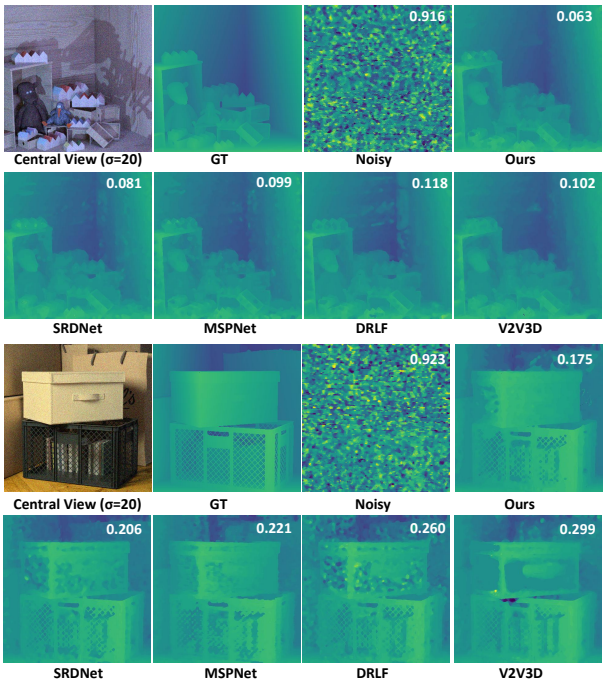


Figure 1. Comparison of depth maps (and their MAE values) estimated by OAVC from the denoised results.

Table 1. Quantitative comparison of denoising results on mixed noise (Gaussian, $\sigma = 20$ + Salt-and-Pepper, $p_s = p_p = 0.05$).

Method	HCI	HCIold	DLFD
DRLF	22.19/0.535	22.25/0.566	22.24/0.578
MSPNet	32.04/0.826	32.00/0.821	30.97/0.809
SRDNet	22.48/0.645	22.89/0.686	22.53/0.666
B2U	30.42/0.795	30.27/0.810	30.33/0.805
V2V3D	31.91/0.879	31.68/0.887	31.06/0.862
Ours	**34.75/0.927**	**34.12/0.906**	**33.54/0.902**

Table 2. Results under different disparity ranges and view number.

Disparity $U \times V$	$[-4, 4]$ 7×7	$[-4, 4]$ 5×5	$[-8, 8]$ 5×5	$[-16, 16]$ 3×3
SRDNet	38.17/0.93	35.94/0.93	34.43/0.90	33.57/0.89
V2V3D	36.21/0.91	34.10/0.89	33.74/0.86	31.85/0.80
Ours	**37.86/0.960**	**35.44/0.940**	**35.08/0.930**	**33.85/0.900**

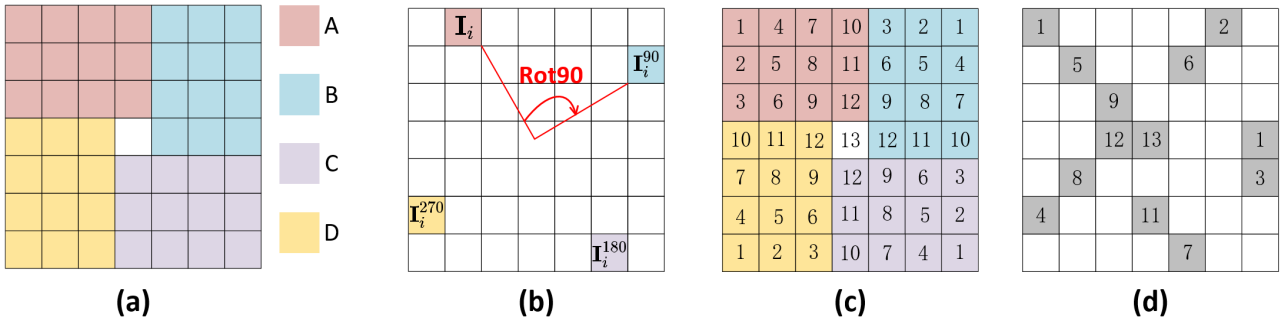


Figure 2. Generation pipeline of our GIM. The process consists of four main steps: (a) First, all views of the light field are partitioned into four subsets (S_A, S_B, S_C and S_D). (b) A rotational correspondence is established between views in different subsets. (c) Based on this correspondence, a unified indexing scheme is applied to all views across the subsets. (d) Finally, the GIM is generated by masking views in a round-robin manner: the view with index 1 is masked in S_A, index 2 in S_B, and so on. This ensures uniform coverage across the angular domain.

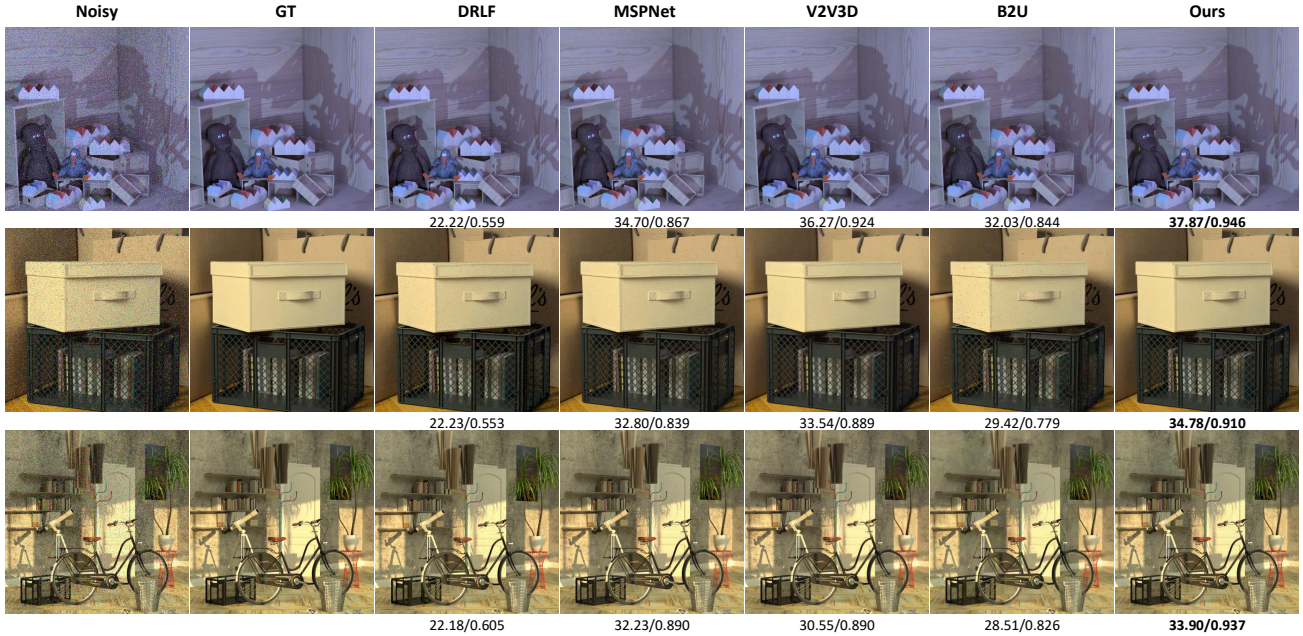


Figure 3. Qualitative comparison of denoising results on mixed noise (Gaussian, $\sigma = 20$ + Salt-and-Pepper, $p_s = p_p = 0.05$).

2. Experiments

2.1. Robustness to Disparity and Angular Resolution

To further evaluate the robustness of our method, we conduct experiments under varying disparity ranges and angular resolutions (i.e., different numbers of views). As shown in Table 2, we test the models under diverse configurations, ranging from dense views with small disparities ($7 \times 7, [-4, 4]$) to sparse views with large disparities ($3 \times 3, [-16, 16]$). Naturally, increasing the disparity range and reducing the angular resolution make the denoising task significantly more challenging due to the reduced multi-view context and larger occlusion areas. However, our proposed method consistently outperforms the self-supervised baseline V2V3D by a large margin across all settings. Moreover, it achieves highly competitive or superior performance compared to the supervised state-of-the-art method SRDNet, particularly exhibiting better structural preservation (consistently higher SSIM).

2.2. Depth Estimation From Denoised LFs

Fig. 1 presents a visual comparison of depth maps estimated by OAVC [3] from the denoised results of our method and other counterparts. Since OAVC computes depth based on multi-view photometric consistency, which is highly sensi-

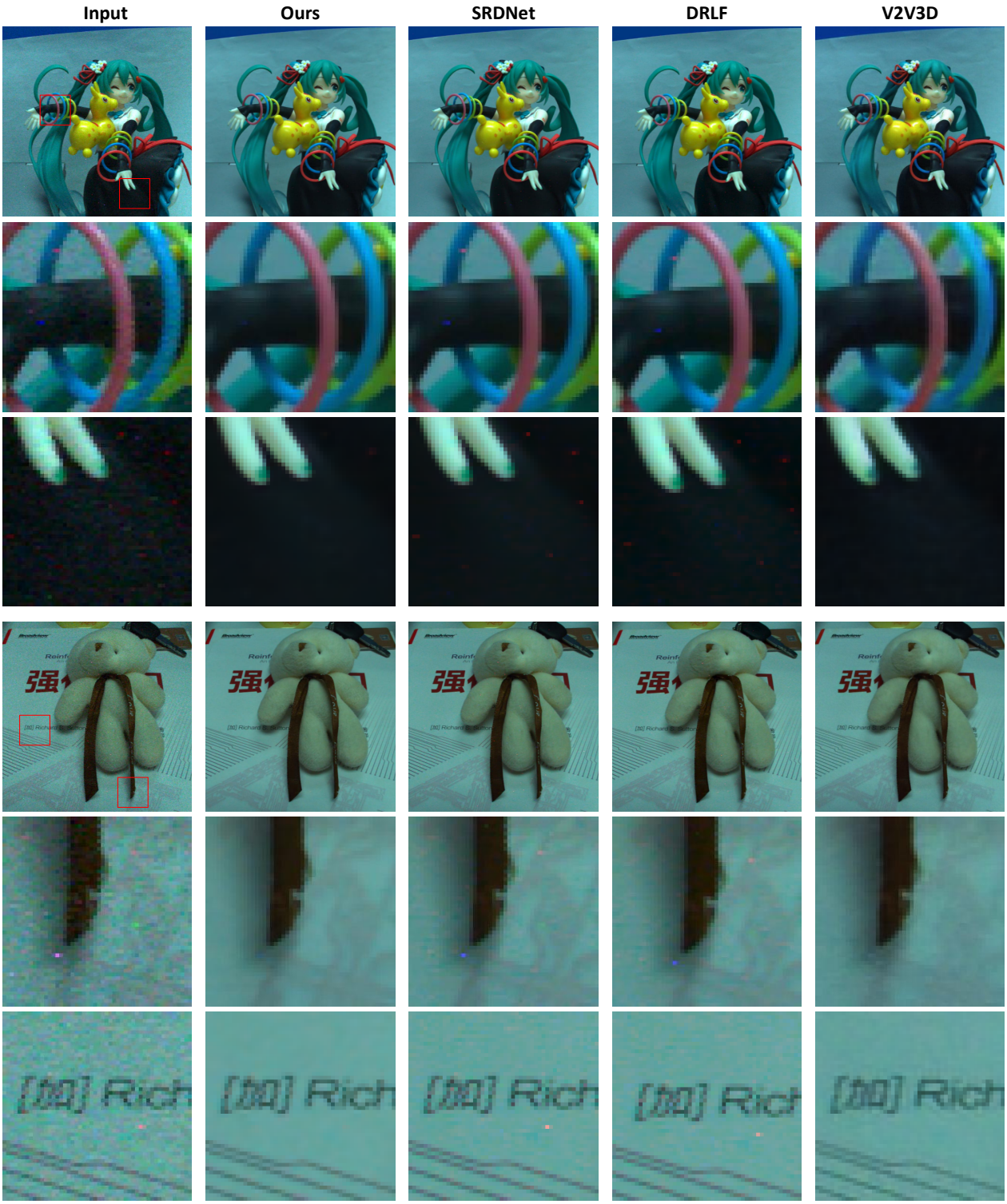
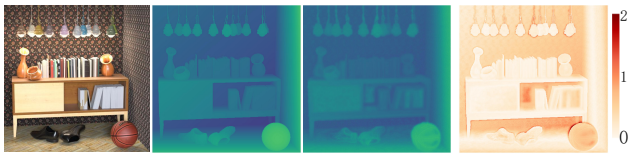


Figure 4. Visual comparison of denoising results on real-world LF images captured by a Lytro Illum camera.

tive to noise, the quality of the depth map is a direct indicator of the denoising performance. The results demonstrate that our method best preserves photometric consistency, particularly in weak-texture regions, and accurately reconstructs fine structures such as the netting.



(a) View (b) GT depth (c) Estimated depth (d) Error map

Figure 5. The visualization of the estimated depth map from the depth probability volume. (a) RGB view. (b) Ground truth depth map. (c) Depth map obtained by our depth estimation module. (d) Absolute error map.

Figure 6. Visualization of denoising results on non-lambertian surfaces.

2.3. Results on Real-world LFs

We show two LF images captured by a Lytro Illum camera and their denoised results in Fig. 4. All models were trained on the HCI dataset corrupted with Gaussian noise at a fixed level of $\sigma = 20$. This choice is motivated by the fact that real-world noise is typically mild. Models trained on noise with random intensities ($\sigma \in [5, 50]$) tend to specialize in removing high-level noise, consequently leading to compromised performance on the more common low-level noise encountered in practical scenarios.

As evidenced by the results, supervised methods SRDNet [1] and DRLF [2] demonstrate good texture reconstruction but are ineffective against impulse noise. Meanwhile, V2V3D [7] suffers from significant artifacts near edges. Our method addresses both limitations: it effectively eliminates impulse noise while also achieving high-fidelity texture restoration.

2.4. Results for Non-Lambertian Surfaces

Our method only requires the content of the blind view to be available in some of the unblind views, even if view consistency is not satisfied across all views. Therefore, our method can handle reflections and specular surfaces as in Fig. 6.

2.5. Results for Non-zero-mean Noise

Monocular BSN like B2U [6] are fundamentally limited by their reliance on the zero-mean noise assumption, causing them to fail on non-zero-mean noise such as salt-and-pepper noise. In contrast, our LF-BVN leverages the photometric consistency across LF views—a geometric constraint independent of noise statistics—enabling robust performance even under such challenging noise conditions. Therefore, we evaluate all models, which were trained on Gaussian noise with $\sigma = 20$, on data corrupted with a mixture of salt-and-pepper noise ($p_s = p_p = 0.05$) and Gaussian noise ($\sigma = 20$). The results are presented in Table 1 and Fig.3. The supervised methods DRLF and SRDNet, being designed specifically for Gaussian noise, fail completely when presented with salt-and-pepper noise. In contrast, MSPNet [5], which was originally developed for low-light enhancement, exhibits a degree of inherent robustness to it. However, its lack of explicit mechanisms to enforce cross-view consistency results in poor recovery of weakly-textured regions. While V2V3D can handle this noise type by explicitly constructing an MPI representation of the LF, this approach often leads to significant artifacts. By leveraging the proposed GIM and the reconstruction consistency loss, our method effectively handles salt-and-pepper noise while recovering high-fidelity textures.

References

[1] Song Chang, Youfang Lin, Wenqi Wang, Da An, and Shuo Zhang. Learning light field denoising with symmetrical refocusing strategy. *IEEE Transactions on Computational Imaging*, 10:1786–1798, 2024. 4

[2] Mantang Guo, Junhui Hou, Jing Jin, Jie Chen, and Lap-Pui Chau. Deep spatial-angular regularization for light field imaging, denoising, and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6094–6110, 2021. 4

[3] Kang Han, Wei Xiang, Eric Wang, and Tao Huang. A novel occlusion-aware vote cost for light field depth estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8022–8035, 2021. 2

[4] Yu-Ju Tsai, Yu-Lun Liu, Ming Ouhyoung, and Yung-Yu Chuang. Attention-based view selection networks for light-field disparity estimation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12095–12103, 2020. 1

[5] Xianglang Wang, Youfang Lin, and Shuo Zhang. Multi-stream progressive restoration for low-light light field enhancement and denoising. *IEEE Transactions on Computational Imaging*, 9:70–82, 2023. 4

[6] Zejin Wang, Jiazheng Liu, Guoqing Li, and Hua Han. Blind2unblind: Self-supervised image denoising with visible blind spots. In *Proceedings of the IEEE/CVF Conference*

on Computer Vision and Pattern Recognition (CVPR), pages 2027–2036, 2022. 4

[7] Jiayin Zhao, Zhenqi Fu, Tao Yu, and Hui Qiao. V2v3d: View-to-view denoised 3d reconstruction for light field microscopy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 26451–26461, 2025. 4