Spectral Normalized U-Net for Light Field Occlusion Removal

Mostafa Farouk Senussi[1], Mahmoud Abdalla[1], Mahmoud SalahEldin Kasem[1], Mohamed Mahmoud[1], and Hyun-Soo Kang[1*]

[1]School of Information and Communication Engineering, Chungbuk National University, Cheongju, 28644, Republic of Korea

[Email: hskang@cbnu.ac.kr]

Abstract

Occlusion artifacts significantly hinder light field (LF) image reconstruction, especially in complex scenes. We propose a spectral normalized U-Net for LF occlusion removal, which begins by stacking LF views and extracting view-dependent features using a local feature encoder. To capture spatial complexity, ResASPP enable multi-scale context aggregation, while channel attention enhances occlusion-related features. Spectral normalization is applied to all convolutional layers to improve training stability and generalization. The encoder-decoder structure with skip connections preserves fine details. Experimental results show our method restores occluded regions more accurately than baselines.

Index Terms: LF Occlusion Removal, Spectral Normalization, Channel Attention, U-Net.

I. INTRODUCTION

Light field (LF) Light field (LF) imaging captures both spatial and angular information of light, enabling post-capture view synthesis and depth-aware processing. However, occlusion artifacts [1] remain a major challenge, especially in scenes with overlapping structures, degrading synthesized views and hindering tasks like object detection and image inpainting [2–4]. These artifacts disrupt scene continuity and lead to inaccuracies in localization and content restoration.

Occlusion removal in LF data is inherently difficult [5], as it requires fusing multi-view information. Traditional inpainting techniques often ignore inter-view dependencies, while early LF-specific methods fail to model the complex interactions between visible and occluded regions. Although deep learning has improved performance via convolutional and attention-based models, challenges in generalization and training stability persist.

To address this, we propose a deep learning model based on a U-Net backbone with spectral normalization. Our method stacks LF views and extracts view-specific features using a local feature encoder enhanced with ResASPP, residual blocks, and channel attention for robust occlusion reasoning. Spectral normalization ensures stable training, while skip connections preserve fine details, enabling high-quality, occlusion-free LF reconstruction. Our approach outperforms existing baselines in both quantitative and qualitative metrics, offering a simple yet effective solution for LF occlusion removal.

* Corresponding author

II. PROPOSED METHOD

We present a deep learning model for LF occlusion removal using a spectral normalized U-Net. As shown in Fig. 1, the architecture comprises a Local Feature Extractor (LFE) and an Occlusion Reconstruction (OR) module. The stacked LF input captures spatial-angular information, which the LFE encodes into multi-scale features. These are processed by the SN-enhanced U-Net for stable and accurate occlusion-free reconstruction.

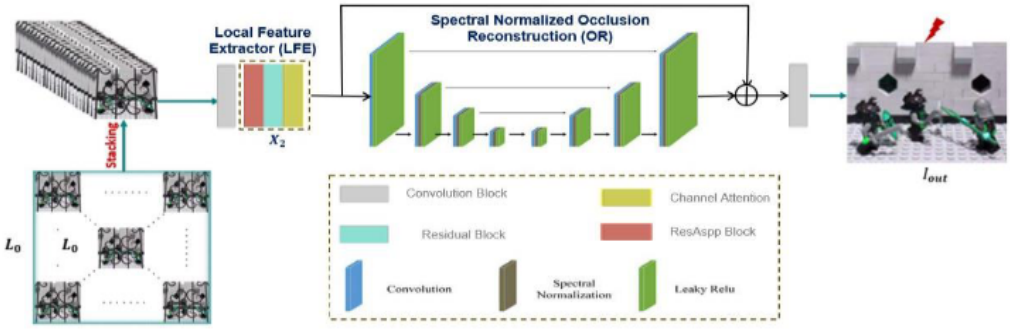

Fig. 1. Overview of the Proposed Spectral Normalized U-Net for Occlusion Removal in LF Images.

A. Local Feature Extractor (LFE)

The LFE starts by stacking all LF views into a 3D tensor L_0, which is processed by a 3×3 convolutional layer (stride 1, padding 1) , capturing spatial and angular correlations. To model multi-scale context, a ResASPP block applies parallel atrous convolutions with dilation rates $\{1,3,6,12\}$, followed by Leaky ReLU activations. Their outputs are concatenated, compressed via a 1×1 convolution, and added through a skip connection.

A residual block further refines, enhancing local structure and feature quality. To emphasize occlusion-relevant features, a channel attention module generates an attention vector using global average pooling followed by two fully connected layers and a sigmoid activation. The attended feature is then passed to the Occlusion Reconstruction module for full-scene recovery.

B. Spectral Normalized Occlusion Reconstruction (OR)

The Occlusion Reconstruction (OR) module is a U-Net architecture enhanced with spectral normalization (SN) in all convolutional layers except the first, improving training stability and generalization. The encoder contains four downsampling stages: the first uses a standard convolution with Leaky ReLU, while the remaining use SN-convolutions. These progressively reduce spatial size while enriching feature depth.

The decoder mirrors this structure with bilinear upsampling and SN-convolutions. Skip connections transfer high-resolution encoder features to corresponding decoder layers, preserving detail and enabling multi-scale context fusion, which is concatenated with encoder features. Finally, a 1×1 convolution generates the center-view occlusion-free output $\mathrm{I}_{\mathrm{out}} \in \mathbb{R}^{\mathrm{H\times W\times 3}}$.

III. EXPERIMENTS AND RESULTS

A. Experimental Setup

We trained on 1,418 DUTLF-V2 [6] images with 5×5 views (256×192), using [7]'s occlusion masks and 21 new dense masks. Training ran 500 epochs with ADAM, batch size 18, and LR 0.001 (halved every 150 epochs), completing in ~1 day on an RTX 4090. Evaluation used synthetic (4-Syn, DUTLF-V2), real (CD), and Single/Double Occ datasets from Stanford Lytro and EPFL-10 with disparities 1–4.

B. Experimental Results

1) Quantitative Results

As shown in Table 1, our method achieves the highest PSNR and SSIM across all LF datasets, confirming its effectiveness in occlusion removal. On synthetic sparse datasets, it scores 26.35 dB / 0.849 (4-Syn) and 26.79 dB / 0.857 (9-Syn), outperforming all baselines. For the real sparse CD, it achieves 25.10 dB / 0.863, demonstrating strong generalization. In dense cases, it attains top PSNR on Single Occ (30.05 dB) and Double Occ (28.47 dB), with competitive SSIM, showing robust performance under complex occlusions.

Table 1. Quantitative Analysis of Sparse and Dense LF Datasets Using PSNR/SSIM: Top Result in Red.

LF Type	Name	RFR [8]	LBAM [9]	DeOccNet [7]	Zhang [10]	Ours
Sparse (syn)	4-Syn	19.89 / 0.668	21.11 / 0.677	23.74 / 0.701	14.46 / 0.683	26.35 / 0.849
	9-Syn	20.69 / 0.672	23.04 / 0.725	23.70 / 0.715	22.00 / 0.758	26.79 / 0.857
Sparse (real)	CD	21.13 / 0.646	21.56 / 0.803	22.70 / 0.741	20.19 / 0.832	25.10 / 0.863
Dense (syn)	Single Occ	26.28 / 0.867	27.92 / 0.668	28.67 / 0.914	23.15 / 0.900	30.05 / 0.827
	Double Occ	23.25 / 0.801	24.83 / 0.827	25.85 / 0.847	18.01 / 0.823	28.47 / 0.849

2) Qualitative Results

As shown in Fig. 2, our method outperforms others in reconstructing occluded regions with high structural and textural accuracy across various LF scenes. In the first row (sparse synthetic), single-image inpainting methods like RFR [8], LBAM [9], and DeOccNet [7] exhibit blurring or artifacts, while Zhang et al. [10] fail to maintain geometry. Our model recovers fine details and spatial coherence, especially in repetitive patterns.

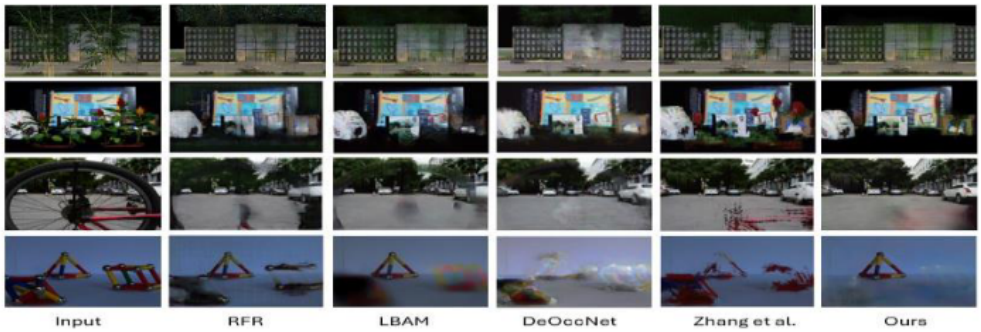

Fig. 2. Visual comparison of our method against existing approaches on sparse and dense LF datasets.

In real-world scenes (rows two and three), competing methods suffer from ghosting and photometric inconsistencies, whereas our model restores sharper, semantically consistent regions. In the dense synthetic scene (fourth row), others show color bleeding and deformation, but our method preserves object boundaries and depth structure, highlighting the effectiveness of our feature extractor and spectral normalized modules.

IV. CONCLUSION

In this paper, we introduced a spectral normalized U-Net architecture for occlusion removal in LF images, combining ResASPP for multi-scale context, residual blocks with channel attention, and spectral normalization for stable training. Our model effectively reconstructs occlusions and outperforms existing baselines.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (Ministry of Science and ICT) (RS-2023-NR076833).

REFERENCES

[1] M. F. Senussi, M. Abdalla, M. S. Kasem, M. Mahmoud, B. Yagoub and H. -S. Kang, "A Comprehensive Review on Light Field Occlusion Removal: Trends, Challenges, and Future Directions," in IEEE Access, vol. 13, pp. 42472-42493, 2025, doi: 10.1109/ACCESS.2025.3548133.

[2] Kasem, M.S.; Mahmoud, M.; Yagoub, B.; Senussi, M.F.; Abdalla, M.; Kang, H.-S. HTTD: A Hierarchical Transformer for Accurate Table Detection in Document Images. Mathematics 2025, 13, 266. https://doi.org/10.3390/math13020266.

[3] Mahmoud, M.; Yagoub, B.; Senussi, M.F.; Abdalla, M.; Kasem, M.S.; Kang, H.-S. Two-Stage Video Violence Detection Framework Using GMFlow and CBAM-Enhanced ResNet3D. Mathematics 2025, 13, 1226. https://doi.org/10.3390/math13081226.

[4] Abdalla, M.; Kasem, M.S.; Mahmoud, M.; Yagoub, B.; Senussi, M.F.; Abdallah, A.; Hun Kang, S.; Kang, H.S. ReceiptQA: A Question-Answering Dataset for Receipt Understanding. *Mathematics* **2025**, *13*, 1760. https://doi.org/10.3390/math13111760.

[5] Senussi, M.F.; Kang, H.-S. Occlusion Removal in Light-Field Images Using CSPDarknet53 and Bidirectional Feature Pyramid Network: A Multi-Scale Fusion-Based Approach. *Appl. Sci.* **2024**, *14*, 9332. https://doi.org/10.3390/app14209332.

[6] Piao, Y., Rong, Z., Xu, S., Zhang, M., & Lu, H. (2020). DUT-LFSaliency: Versatile dataset and light field-to-RGB saliency detection. arXiv preprint arXiv:2012.15124.

[7] Wang, Y., Wu, T., Yang, J., Wang, L., An, W., & Guo, Y. (2020). DeOccNet: Learning to see through foreground occlusions in light fields. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 118-127).

[8] Li, J., Wang, N., Zhang, L., Du, B., & Tao, D. (2020). Recurrent feature reasoning for image inpainting. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7760-7768).

[9] Xie, C., Liu, S., Li, C., Cheng, M. M., Zuo, W., Liu, X., ... & Ding, E. (2019). Image inpainting with learnable bidirectional attention maps. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 8858-8867).

[10] Zhang, S., Shen, Z., & Lin, Y. (2021, August). Removing Foreground Occlusions in Light Field using Micro-lens Dynamic Filter. In IJCAI (pp. 1302-1308).