UCMNet: Uncertainty-Aware Context Memory Network for Under-Display Camera Image Restoration

Supplementary Material

This supplementary material provides additional details that were omitted from the main paper due to space constraints. Sec. A describes the architectural details of the Frequency Convolution Module, and Sec. B explains the training-time update mechanism of the Context and Memory Banks. In Sec. C, we investigate the effect of varying the size of the Context and Memory Banks and identify an appropriate configuration for UCMNet. Sec. D presents the loss functions and qualitative comparisons on the SYNTH dataset, while Sec. E reports the performance of UCMNet on a real-world UDC dataset. In Sec. F, we demonstrate the capability of UCMNet for low-light image enhancement. Finally, Sec. G and Sec. H provide additional visualizations of uncertainty maps and memory token usage, respectively, and Sec. I presents limitation examples of UCMNet for future work.

A. Architectural Details of FCM

In Sec. 3.1.1, to provide reliable representations for UDC image restoration, we introduce a lightweight Frequency Convolution Module (FCM), as illustrated in Fig. 3 (b) of the main manuscript. Inspired by NAFNet [1] and DarkIR [3], the proposed FCM is designed to perform frequency-domain feature enhancement followed by spatial attention refinement. In practice, FCM is integrated into every encoder and decoder blocks to enable efficient and effective restoration.

Frequency-domain Feature Enhancement. To preserve and enhance high-frequency details, FCM transforms input feature into the Fourier domain, decomposes it into phase and amplitude, and refines the amplitude while preserving phase, which retains the structural information of the input feature. Then, through the inverse Fourier transformation, we obtain a frequency residual that compensates the original feature.

Spatial Attention Refinement. After frequency compensation, to address spatially varying distortions we apply a spatial attention module following the NAFNet design. Specifically, a lightweight Single Gate (SG) selectively amplifies salient spatial activations, while a Simplified Channel Attention (SCA) module enhances feature channels to emphasize consistently informative responses.

B. Context Bank and Memory Bank Update

In Sec. 3.2.1, we explain how the uncertainty map $F_U \in \mathbb{R}^{H' \times W' \times C}$ is used as a prior to retrieve the context feature $F_C \in \mathbb{R}^{H' \times W' \times C}$ from the Memory Bank $\mathbf{M} \in \mathbb{R}^{N \times C}$ and the Context Bank $\mathbf{C} \in \mathbb{R}^{N \times C}$, where F_C represents high-frequency information tailored to local degradation patterns. In this section, we detail the update mechanism for the Memory Bank $\mathbf{M}$ and the Context Bank $\mathbf{C}$ during training.

Memory Bank. We update the memory tokens $m_i \in \mathbb{R}^{1 \times C}$, $(i \in \{1, \ldots, N\})$ in the Memory Bank $\mathbf{M}$ according to their similarity to the uncertainty feature vectors $f_j^u \in \mathbb{R}^{1 \times C}$, $(j \in \{1, \ldots, H'W'\})$. Using the cosine similarity score s_{ij} in Eq. 3 of the main paper, we first identify the address of the most relevant memory token for each uncertainty feature vector:

$$r_j = \arg\max_i s_{ij}, \ (i \in \{1, \ldots, N\}), \tag{S.1}$$

where r_j denotes the address of the memory token most similar to the uncertainty feature vector f_j^u. The corresponding memory token m_{r_j} is then updated using a momentum-based rule following [2, 5] so that it progressively adapts to the uncertainty feature:

$$m_{r_j} \leftarrow \eta \cdot m_{r_j} + (1 - \eta) \cdot f_j^u, \tag{S.2}$$

where η is the momentum coefficient.

Context Bank. The tokens c_i in the Context Bank $\mathbf{C}$ serve as the components of the context feature F_C, as described in Sec. 3.2.1. They are updated jointly with the other network parameters during training, allowing them to gradually encode high-frequency information that supports accurate restoration of UDC images.

C. Ablation on Context-Memory Bank Size

In Table S1, we investigate the impact of the number of tokens (N) in the Memory and Context Banks. When the token number N is 256, the model provides the best trade-off between representational capacity and restoration performance.

D. SYNTH Dataset

Loss Function. Similar to BNUDC [6] and FSI [7], we employ the L1 loss on the SYNTH dataset rather than the

Figure S1. Please zoom in for the original-size comparison.

Table S1. Ablation on the size of the Context–Memory Bank (N). A compact configuration (N=256) achieves the best balance between accuracy and efficiency.

CMB size (N)	128	**256**	512
PSNR↑	33.71	**33.78**	33.60
SSIM↑	0.9624	**0.9626**	0.9619

Dataset	PPM-UNet	AlignFormer+PPM-UNet	UCMNet	AlignFormer+UCMNet
AlignFormer	19.03 / 0.7808	22.95 / 0.8581	**20.41 / 0.8168**	**23.11 / 0.8612**

Table S2. PSNR/SSIM results on real UDC dataset from AlignFormer [4]

Dataset	Restormer	GSAD	Retinexformer	DarkIR	UCMNet (Ours)
LOLv2-Real	19.94 / 0.827	20.15 / 0.846	22.79 / 0.840	**23.87 / 0.880**	23.10 / 0.862
LOLv2-Syn	21.41 / 0.830	24.47 / 0.929	**25.67** / 0.930	25.54 / 0.934	**25.67 / 0.936**

Table S3. UCMNet on low-light image enhancement task.

PSNR loss, as follows:

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{HF-UDL}} + \lambda_2 \mathcal{L}_{\text{L1}}, \tag{S.3}$$

where $\mathcal{L}_{\text{L1}}$ denotes the pixel-wise L1 loss between the prediction $\hat{I}$ and I_{gt}. The weighting values of λ_1 and λ_2 are empirically determined as 100 and 0.5.

Visualization. As the manuscript includes visual comparisons only for the TOLED and POLED datasets like Fig.S1, we provide qualitative results on the SYNTH dataset in Fig.S3. Unlike FSI [7] and BNUDC [6], our method achieves noticeably better detail restoration and effectively suppresses light-diffraction artifacts characteristic of UDC imaging conditions.

E. Real-world Dataset

Although the TOLED, POLED, and SYNTH datasets simulate real-world UDC degradations, they remain synthetic. Therefore, in Table S2, we further evaluate our method on the real-world dataset released by AlignFormer [4], where UCMNet also demonstrates robust performance.

F. UCMNet on Low-light Image Enhancement

In this section, we explore the potential of the proposed UCMNet for another image restoration task, namely low-light image enhancement, which also involves recovering images degraded by suboptimal lighting conditions. As in Table S3, we apply UCMNet to LOLv2-Real [8] and LOLv2-Syn [8] to validate its broader applicability, where degradations are spatially varying. Even without being fully optimized for this specific task, UCMNet achieves competitive performance compared to existing methods, indicating its potential for handling tasks with spatially varying degradations under challenging lighting conditions.

G. Uncertainty Map Visualization

As an extension of Fig. 8 in the main paper, additional uncertainty maps are provided in Fig. S4. We observe that shallower stages emphasize fine structural details, whereas deeper stages primarily capture coarse degradation patterns. This indicates that the network performs depth-wise restoration guided by the uncertainty prior.

H. Memory Token Usage Visualization

In Fig. S5, we visualize the usage of memory tokens by representing them as pseudo–segmentation maps, where each color corresponds to a different memory token associated with particular uncertainty feature vectors. This allows us to examine the spatial distribution of uncertainty features across the image. The results are shown depth-wise across decoder stages. In deeper stages, the maps converge to similar patterns, indicating that uncertainty becomes focused on display-induced degradations. In contrast, shallower stages highlight high-frequency structural cues such as edges and contours. These observations demonstrate that UCMNet effectively leverages diverse memory tokens to model region-specific uncertainty, enabling spatially adaptive restoration.

I. Limitations

Although the proposed network demonstrates promising restoration performance on UDC images, UCMNet still struggles on the SYNTH dataset to recover fine details in regions with severe diffraction-induced information loss, as in Fig. S2. In future work, we plan to address this limitation by incorporating stronger model priors.

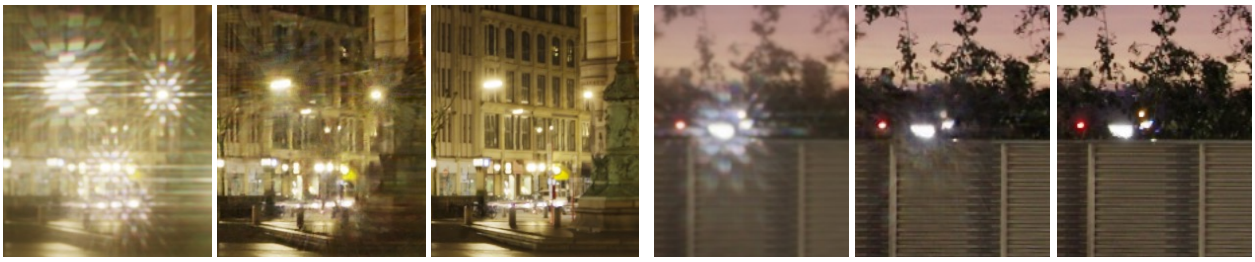

Figure S2. Limitation examples of UCMNet on SYNTH dataset.

References

[1] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*. Springer, 2022. 1

[2] Zhenxuan Fang, Fangfang Wu, Tao Huang, Le Dong, Weisheng Dong, Xin Li, and Guangming Shi. Parameterized blur kernel prior learning for local motion deblurring. In *CVPR*, 2025. 1

[3] Daniel Feijoo, Juan C Benito, Alvaro Garcia, and Marcos V Conde. Darkir: Robust low-light image restoration. In *CVPR*, 2025. 1

[4] Ruicheng Feng, Chongyi Li, Huaijin Chen, Shuai Li, Jinwei Gu, and Chen Change Loy. Generating aligned pseudo-supervision from non-aligned data for image restoration in under-display camera. In *CVPR*, 2023. 2

[5] Huaibo Huang, Aijing Yu, and Ran He. Memory oriented transfer learning for semi-supervised image deraining. In *CVPR*, 2021. 1

[6] Jaihyun Koh, Jangho Lee, and Sungroh Yoon. Bnudc: A two-branched deep neural network for restoring images from under-display cameras. In *CVPR*, 2022. 1, 2

[7] Chengxu Liu, Xuan Wang, Shuai Li, Yuzhi Wang, and Xueming Qian. Fsi: Frequency and spatial interactive learning for image restoration in under-display cameras. In *ICCV*, 2023. 1, 2

[8] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE Transactions on Image Processing*, 2021. 2

[9] Yuqian Zhou, David Ren, Neil Emerton, Sehoon Lim, and Timothy Large. Image restoration for under-display camera. In *CVPR*, 2021. 5

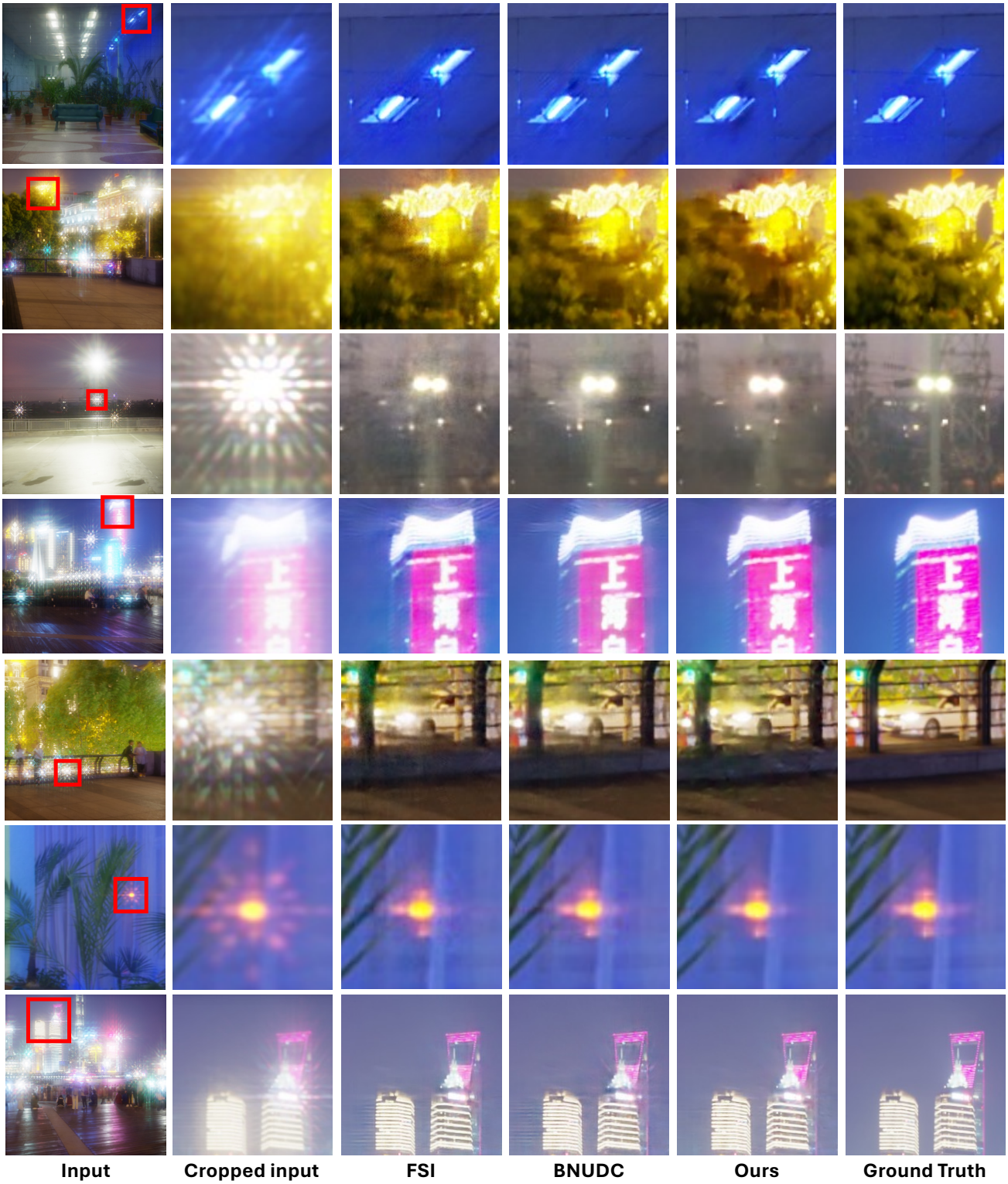

Figure S3. Visual results on the SYNTH dataset.

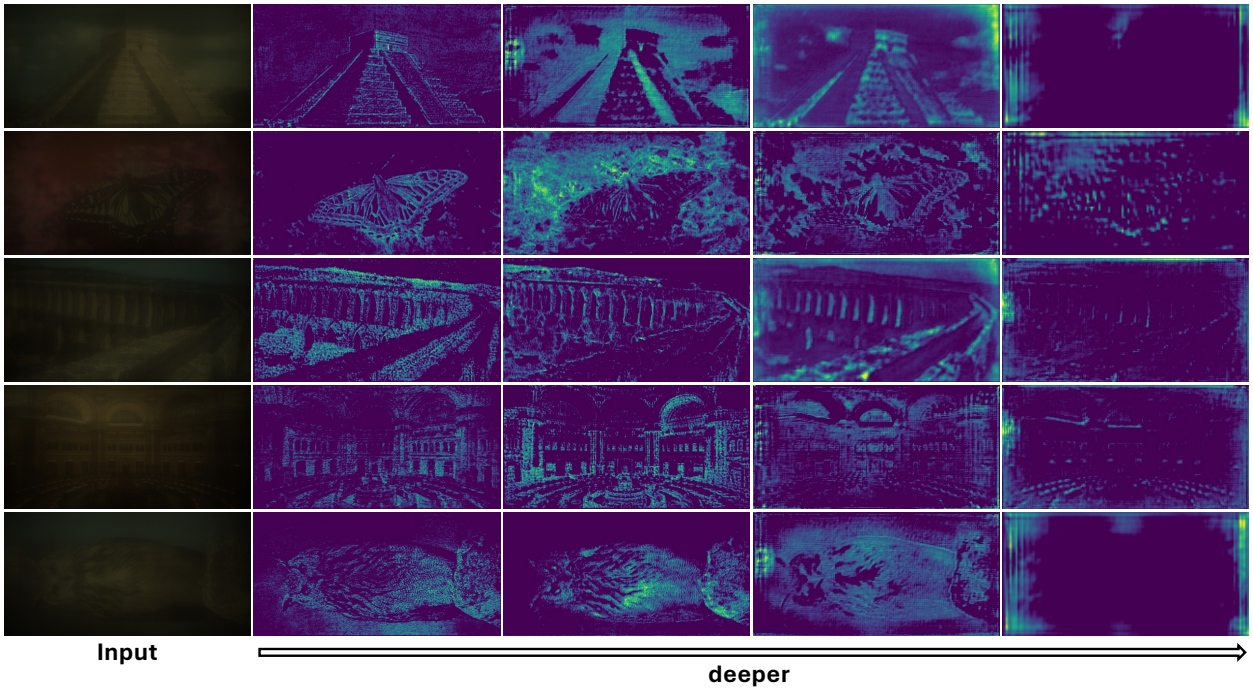

Figure S4. Estimated uncertainty map visual results on the POLED [9] dataset.

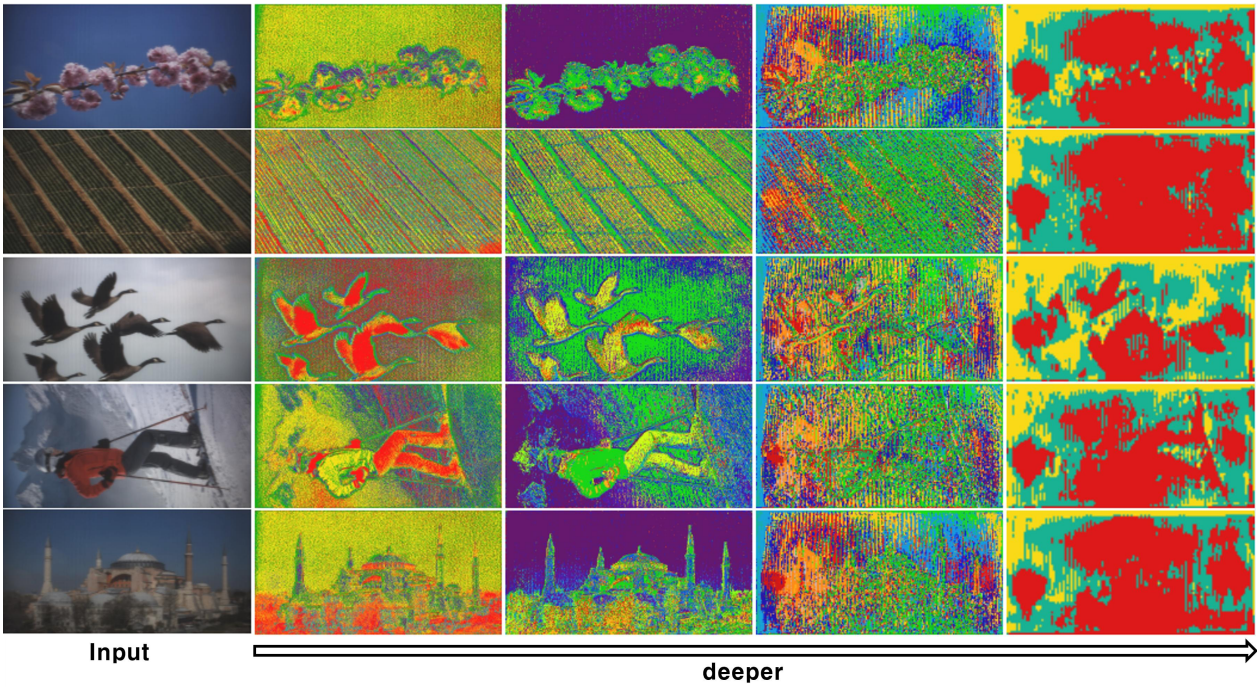

Figure S5. Visualized memory token usage as pseudo–segmentation maps on the TOLED [9] dataset, where each color represents a distinct memory token associated with specific uncertainty feature vectors, allowing inspection of their spatial distribution across the image.