Supplementary Materials: Exploring View Consistency for Scene-Adaptive Low-Light Light Field Image Enhancement

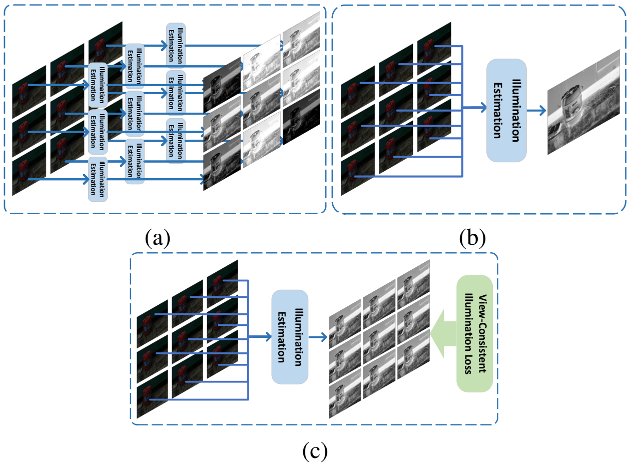
Shuo Zhang, Chen Gao, Youfang Lin*
Beijing Key Laboratory of Traffic Data Mining and Embodied Intelligence,
School of Computer Science & Technology, Beijing Jiaotong University, Beijing, China.
{zhangshuo, gaochen, yflin}@bjtu.edu.cn

In this supplementary material, we first provide further discussion of our illumination adjustment strategy and present the details of the VCRM block. We then provide additional experimental results, including more quantitative and qualitative analyses on the *Dynamic Illumination Task* and *Fixed Illumination Task*.

1. Method

1.1. Illumination Adjustment Discussion

We compare our illumination estimation strategy with others in Fig. A. One possible way for illumination estimation is to treat each view separately as shown in Fig. A(a). Since the supervision of the illumination adjustment map is lacking, the estimated illumination adjustment maps vary a lot among the views. Another common way is to estimate a uniform illumination adjustment map for each view by interacting information between views as Fig. A(b). The adjustment is the same for all views, which does not satisfy the disparity constraint. In order to obtain the illumination adjustment map that complies with the view-consistency relation, we propose to explore the relationship between views to estimate an illumination adjustment map for each view. The unsupervised illumination loss ℓ_{vc} is designed to constrain the view consistency as Fig. A(c).



(a)

(b)

(c)

Figure A. Schematic diagram of different illumination adjustment strategies. (a) The illumination map for each view is estimated separately. (b) One uniform illumination map is estimated for each view. (c) Our method considers all views and estimates the illumination map for each view. The view-consistent illumination loss is designed to keep view consistency among the illumination maps.

1.2. Overall Structure of VCRM

As shown in Fig. B, VCRM adopts a multi-scale architecture using $1, 1/2, 1/4$ size of the original input. In each scale, the View Consistent Feature Aggregation Unit (VCFAU) is implemented to explore the redundant information of the views to recover the details.

Specifically, we calculate the average of $\mathcal{L}^a$ in the angular domain and concatenate the results with the original $\mathcal{L}^a$ to alleviate the influence of noise. In each scale, we first apply 2D residual blocks on the spatial domain to extract the high-dimensional feature as F_1^k, where k represents the different scales. Then F_1^k is sent to the cascaded VCFAU blocks H_{FAU_k} and the deep features of the current scale are obtained:

$$F_i^k = H_{FAU_k}^i\left(F_{i-1}^k\right) \quad i = 1, 2, 3 \tag{1}$$

We then exploit the transposed spatial convolution layers to upscale the related features. The upsampled feature $F_3^{1/2} \uparrow, F_3^{1/4} \uparrow$ are concatenated with the F_3^1 and fed to a decoder that has a similar structure with the encoder to get the refined recovery result $\mathcal{L}^{out}$.

1.2.1. View Consistent Feature Aggregation Unit

In VCFAU, we employ deformable convolution layers on EPIs to further integrate information between different views. Since the slope of lines in EPIs corresponds to disparity information, we further import the disparity con-

[*] Corresponding author: yflin@bjtu.edu.cn

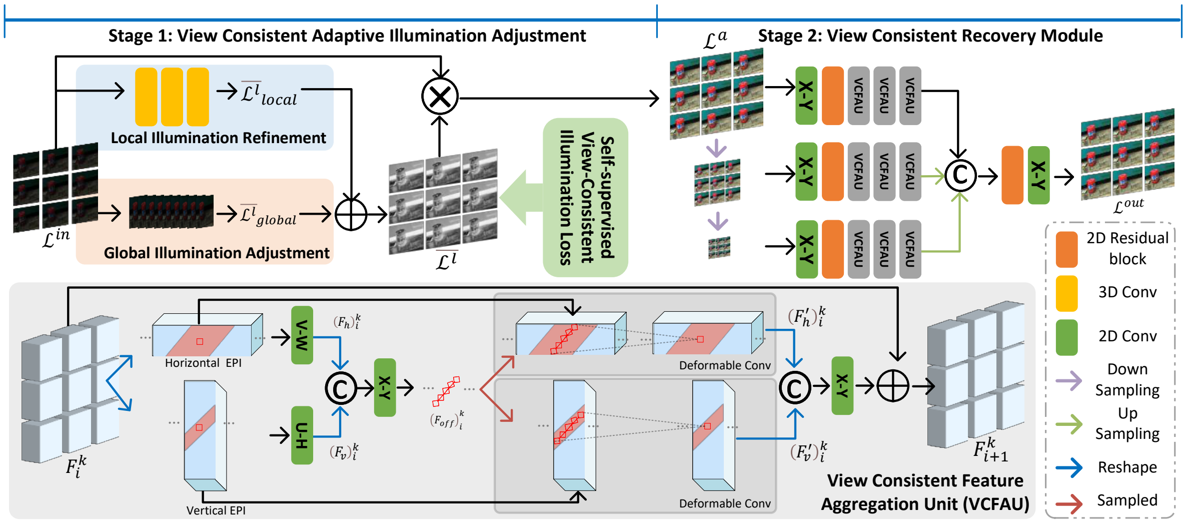

Stage 1: View Consistent Adaptive Illumination Adjustment
Stage 2: View Consistent Recovery Module
$\mathcal{L}^{l}_{local}$
Local Illumination Refinement
$\mathcal{L}^{in}$
$\mathcal{L}^{l}_{global}$
Global Illumination Adjustment
$\mathcal{L}^{l}$
Self-supervised View-Consistent Illumination Loss
$\mathcal{L}^{a}$
X-Y
VCFAU
C
$\mathcal{L}^{out}$
2D Residual block
3D Conv
2D Conv
Down Sampling
Up Sampling
Reshape
Sampled
Horizontal EPI
Vertical EPI
V-W
U-H
$(F_h)_i^k$
$(F_v)_i^k$
$(F_{off})_i^k$
F_i^k
Deformable Conv
$(F_h')_i^k$
$(F_v')_i^k$
F_{i+1}^k
View Consistent Feature Aggregation Unit (VCFAU)

Figure B. An overview of our proposed VCNet network, which consists of an illumination estimator VCAIA and a corruption restorer VCRM. In VCAIA, the illumination map is estimated and multiplied to the original input to light up the illumination. VCRM, whose basic unit is VCFAU, further recovers the details of the light-up image and outputs the final results.

straint to help the model find correct and consistent complementary information.

Firstly, the input feature $F_i^k \in \mathbb{R}^{U\times V\times H_k\times W_k\times C}$ is reshaped and fed into multiple 2D EPI convolutional layers on the V-W and U-H subspaces to extract the horizontal $(F_h)_i^k$ and vertical $(F_v)_i^k$ EPI features, respectively. To further exploit the spatial-angular correlation in EPIs, we use the deformable convolution to interact with the redundant information. Instead of respectively calculating the offset in the deformable convolution for the horizontal and vertical EPIs, we consider the disparity consistency in EPIs. As an inherent scene property, the disparity is supposed to be consistent in V-W and U-H subspaces. Therefore, we propose to calculate one uniform offset for deformable convolution layers in both the V-W and U-H subspaces. In particular, we reshape and concatenate the horizontal EPI information $(F_h)_i^k$ with the vertical EPI information $(F_v)_i^k$. Then the spatial convolution layers on spatial domain H_{off} is implemented to obtain the offsets:

$$(F_{off})_i^k = H_{off}\left(\left[(F_h)_i^k, (F_v)_i^k\right]\right), \tag{2}$$

where $(F_{off})_i^k \in \mathbb{R}^{U\times V\times H_k\times W_k\times 18}$ and *18* corresponds to 2D offsets of the 3×3 points. The $(F_{off})_i^k$ integrates information captured in both horizontal and vertical directions, which enables the model to better learn the correspondence between cross views. For each point of the EPIs, we compute the 2D offsets of the 3×3 neighborhood region, which represent the sampling positions on the related EPIs in one angular and one spatial dimension. Specifically, in horizontal EPI features $(F_h)_i^k$, the offsets represent the relative sampling positions $(\Delta v, \Delta y)$ in V-W subspace. By contrast, in vertical EPI features $(F_v)_i^k$, the offsets represent the relative sampling positions $(\Delta u, \Delta x)$ in U-H subspace. Once the $(F_{off})_i^k$ is obtained, we fed them into the 2D deformable convolution layers to interact redundant information in V-W and U-H subspaces as follows:

$$(F_h')_i^k = H_{deform_h}\left((F_h)_i^k, (F_{off})_i^k\right), \tag{3}$$

$$(F_v')_i^k = H_{deform_v}\left((F_v)_i^k, (F_{off})_i^k\right), \tag{4}$$

where H_{deform_h} means the deformable convolution layers processed on V-W subspace, while H_{deform_v} means the deformable convolution layers processed on U-H subspace. It is worth mentioning that the parameters of the deformable convolution layers differ between horizontal and vertical processing. By using uniform offsets in V-H and U-W spaces, angular information can be efficiently incorporated and maintain epipolar consistency. Then, the features $(F_h')_i^k$ and $(F_v')_i^k$ are reshaped and concatenated in the channel dimension. Finally, we employ spatial convolution layers to aggregate the information in the horizontal and vertical EPIs, whose output is added to the F_i^k to obtain the output feature F_{i+1}^k.

2. Experiments

2.1. Dynamic Illumination Task

We further show the recovered EPIs in Fig. D, where the slops of lines indicate the disparity. The EPIs recovered from single-frame-based methods are heavily aliasing, which means that they cannot recover consistent illumination of the corresponding points in different views.

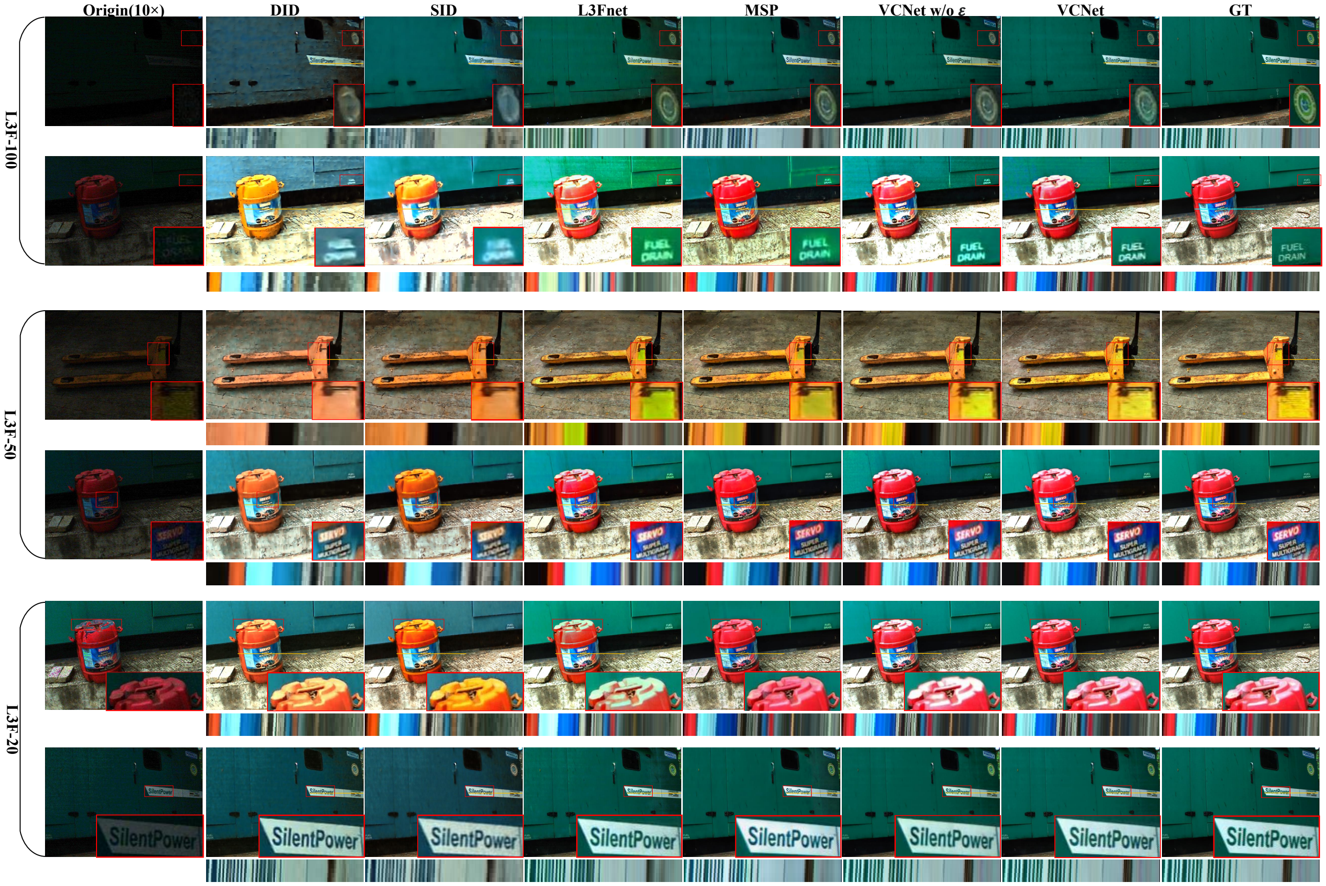


Figure C. Visual comparison on *L3F-Fixed* task, where the central view and EPIs of the recovery LFs are shown. Details are zoomed in for comparison. Other methods either collapse by noise, or distort color, or produce blurry and under-/over-exposed images. Our VCNet effectively removes the noise and reconstructs well-exposed image details. The related EPIs show that the proposed method also well maintains the LF geometry.

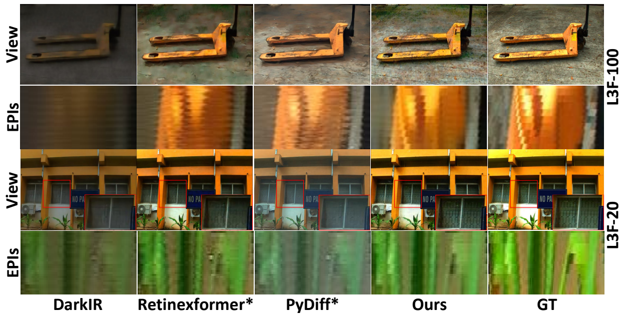


Figure D. Visual comparion of the recovery views and the EPIs.

Table A. More results on the *L3F-fixed* task.

Method	L3F-20	L3F-50	L3F-100
	PSNR↑ / SSIM ↑/ LPIPS↓	PSNR↑ / SSIM↑ / LPIPS↓	PSNR↑ / SSIM↑/ LPIPS↓
SID[1]	22.64 / 0.74 / 0.1755	21.27 / 0.67 / 0.2844	17.03 / 0.56 / 0.3425
DID[4]	23.21 / 0.74 / 0.1515	20.90 / 0.64 / 0.2865	19.22 / 0.57 / 0.3643
SGN[2]	23.43 / 0.77 / 0.1586	21.81 / 0.69 / 0.2658	20.29 / 0.60 / 0.3655
resLF[6]	23.56 / 0.77 / 0.1763	22.09 / 0.70 / 0.2611	21.40 / 0.66 / 0.3176
MTO[5]	23.18 / 0.76 / 0.1735	22.04 / 0.70 / 0.2631	20.78 / 0.62 / 0.3408
MBLLN[3]	24.57 / 0.81 / 0.2790	22.23 / 0.69 / 0.3973	20.45 / 0.60 / 0.4844
VCNet	**26.61 / 0.84 / 0.0580**	**25.13 / 0.79 / 0.0991**	**23.24 / 0.72 / 0.1716**

2.2. Fixed Illumination Task

We evaluate all the methods under the fixed illumination condition using the *L3F-Fixed* dataset. Fig. C provides a visual comparison between our method and other SOTA methods for low-light enhancement in various lighting conditions. To visualize the degree of degradation at different illumination levels, we scaled up the pixel values of the input images by a factor of 10. As shown, the input images reveal little detail due to the challenging lighting conditions. Compared with the L3F-20 and L3F-50 datasets, it is more challenging to recover images in the L3F-100 dataset. The single-frame-based methods produce rough predictions with noticeable color biases. This is because these methods do not consider redundant information between views and independently recover different views, making their final output more susceptible to image noise. In contrast, LF-based methods yield more appealing results by leveraging information from multiple views inherent in LF data during the restoration process. Compared to other techniques, our approach achieves superior quality in terms of both complex texture and color. Our results exhibit colors that closely match the ground truth, and the edges in our results appear more defined. More numeric results of other methods are supplemented in Table . All these results clearly suggest that our model is able to better interact information between views, and effectively restore images.

2.3. Ablation Study

Table A. The ablation study of ℓ_{vc} loss and VCRM block.

No.	VCRM	ℓ_{vc}	PSNR/SSIM/LPIPS
(1)	w/o Disparity Constraint	$\lambda_1 = 0$	22.76/0.73/0.1507
(2)	with Disparity Constraint	$\lambda_1 = 0$	23.59/0.75/0.1420
(3)		$\lambda_1 = 0.5$	**23.84/0.76/0.1395**
(4)		$\lambda_1 = 1$	23.74/0.76/0.1409

2.3.1. Disparity Constraint in VCRM

As the disparity constraint is the key idea in VCRM, the related ablation results are provided in Table. A. Compared with model-1, model-2 using shared offsets in horizontal and vertical EPIs outperforms nearly 1db PSNR. The shared offset makes the model more robust and better capture redundant information between views.

2.3.2. Loss Functions

We choose the weight λ_1 of our l_{vc} by conducting several experiments. The results are provided in Table A, where the other weights λ_2, λ_3 are unchanged.

References

[1] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*, pages 3291–3300, 2018. 3

[2] Shuhang Gu, Yawei Li, Luc Van Gool, and Radu Timofte. Self-guided network for fast image denoising. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2511–2520, 2019. 3

[3] Feifan Lv, Feng Lu, Jianhua Wu, and Chong Soon Lim. Mbllen: Low-light image/video enhancement using cnns. In *British Machine Vision Conference*, 2018. 3

[4] Paras Maharjan, Li Li, Zhu Li, Ning Xu, Chongyang Ma, and Yue Li. Improving extreme low-light image denoising via residual learning. In *Proceedings of the International Conference on Multimedia and Expo (ICME)*, pages 916–921, 2019. 3

[5] Shansi Zhang and Edmund Y. Lam. Learning to restore light fields under low-light imaging. *Neurocomputing*, 456:76–87, 2021. 3

[6] Shuo Zhang, Youfang Lin, and Hao Sheng. Residual networks for light field image super-resolution. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11038–11047, 2019. 3