Supplementary Material to "Dispel Darkness for Better Fusion: A Controllable Visual Enhancer based on Cross-modal Conditional Adversarial Learning"

Hao Zhang, Linfeng Tang, Xinyu Xiang, Xuhui Zuo, Jiayi Ma*
Electronic Information School, Wuhan University, Wuhan, China
{zhpersonalbox, linfeng0419, jyma2010}@gmail.com, {xiangxinyu, xuhuizuo2001}@whu.edu.cn

1. Interactive Executable Interface

We offer a user-friendly interface for our proposed DDBF, as shown in Fig. 1. This interface includes all the functional buttons of DDBF, enabling users (possibly without programming knowledge) to interactively use DDBF. Next, we provide detailed instructions for using this interface.

- **File Upload:** used to specify original input that is multi-modal data directly captured by sensors, including low-light visible images and infrared images.
- **Monitoring Mode:** used to specify the content displayed in the left "Input" window. There are three modes available for selection: infrared mode, visible mode, and mixed display mode.
- **Correction Ratio:** a customized button used to control the enhancement ratio r of DDBF, supporting flexible adjustment by sliding the mouse pointer. It enables users to obtain visualizations of enhancement and fusion that meet personal visual preferences in a WYSIWYG (What You See Is What You Get) manner.
- **Night/Day:** used to specify the lighting conditions when the input data is taken, so as to automatically recommend a correction ratio to achieve promising enhancement and fusion.
- **Enhancement Mode:** used to specify the content displayed in the right "Enhance" window. There are three modes available for selection: enhancement mode, fusion mode, and mixed display mode.

Such an interactive interface is very suitable for deployment on the software side of monitoring and reconnaissance, such as traffic monitoring and individual helmets, as shown in Fig. 2. It will greatly improve visibility at night and break through the visual limitations brought by low-reflected light. For instance, in complex and changing low-light conditions, the security personnel/police can easily adjust the enhancement ratio on their mobile terminal for more effective monitoring. This interactive executable interface will be publicly available for free use in https://github.com/HaoZhang1018/DDBF.

*Corresponding author

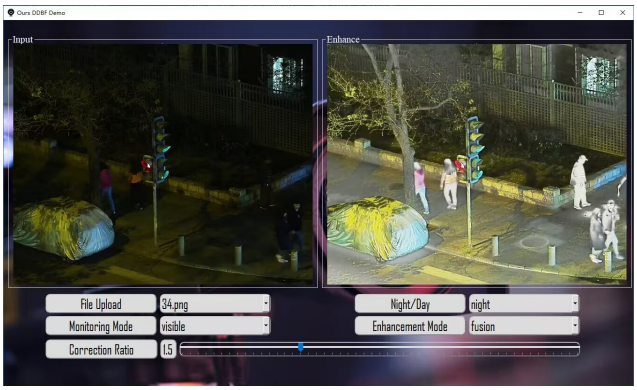

Figure 1. An example of the developed executable interactive interface.

Figure 2. Application scenarios of the controllable enhancement.

2. Adjustment of Enhancement Ratio

The multi-modal image fusion is a low-level visual task that aims to provide images that are more favorable for the user's visual perception. In our method, users can easily adjust the enhancement ratio by dragging the enhancement ratio button in the provided interactive executable interface, allowing them to interactively customize the fused results to their visual preferences. The increase of the enhancement ratio gradually eliminates the information mismatch between the infrared and visible modalities, so that the fused image can effectively fuse the effective information of the two modalities. However, when the enhancement ratio is too large, the visible image will dominate the fusion process due to the excessive intensity, causing a weakening of thermal radiation information and thus reducing fusion performance. Therefore, please obtain visualizations of enhancement and fusion that meet your visual preferences in a WYSIWYG

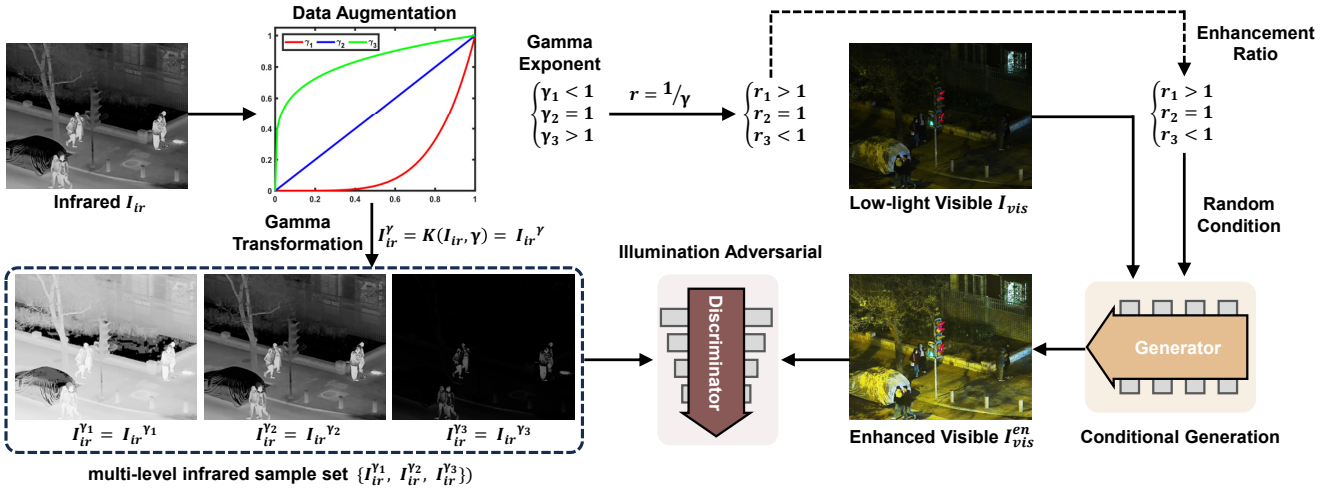

Figure 3. Pipeline of the guided restoration module.

(What You See Is What You Get) manner.

3. Pipeline of The Guided Restoration Module

Because the operations and variables involved in the guided restoration module are relatively numerous and complex, we additionally show the pipeline of the guided restoration module here to help understanding. In the guided restoration module, γs are parameters (*i.e.*, exponent) of gamma transformation, which are used to control the non-linear adjustment of image intensity. We use multiple gammas for original infrared images, constructing a multi-level infrared sample set that reflects the desired multiple illumination/contrast distributions. This set acts as a positive sample in adversarial learning, guiding the generator to have the ability to adjust multi-level illumination. Considering that the value of gamma is negatively correlated with brightness adjustment ($\gamma > 1$ means dimming), we define an enhancement ratio as the reciprocal of gamma, thus serving as the conditional input of the generator. Notably, only one enhancement ratio and a corresponding transformed infrared sample are randomly fed into the GAN for each iteration. We provide the pipeline of training the guided restoration module in Fig. 3 to illustrate the above process intuitively. Besides, we add a list of variables and describe their paraphrases in Table 1.

Table 1. List of variables.

Notations	Paraphrase
γ	Exponent in gamma transformation
r	Enhancement ratio $r = 1/\gamma$
K	Gamma transformation
G	Generator
D	Discriminator
L	Gaussian low-pass filter
U	YUV Color removal function
a, b, c	Probability labels

4. More Visual Comparisons

As the length of the main text is limited, we provide more visual comparisons here to demonstrate the advantages of our DDBF.

4.1. Visual Results of Low-light Enhancement

ExDark Dataset. Two groups of visual results of low-light enhancement on the ExDark dataset [3] are shown in Figs. 4 and 5. It can be seen that our DDBF can generate enhanced results with better visual clarity, presenting more vivid texture structures. For example, our DDBF can best preserve the trees on mountains in Fig. 4, and patterns on bottles in Fig. 5. Furthermore, the flexibility of our method for illumination adjustment is very attractive, which realistically simulates the gradual appearance change of the scene as the lighting increases. For instance, as the enhancement ratio increases, the results of our method in Fig. 4 resemble the transition from early morning to sunrise.

AGLIE Dataset. Figs. 6 and 7 demonstrate the enhanced results of different low-light enhancement methods on the AGLIE dataset. As the ground truth is available in the AGLIE dataset [4], we can refer to it to evaluate the low-light enhancement performance. It can be observed that by adjusting the enhancement ratio, our method effectively increases the brightness while maintaining the most consistent color with the ground truth. These results demonstrate that our method can achieve more promising low-light enhancement than other comparative methods.

4.2. Visual Results of Low-light Multi-modal Fusion

LLVIP Dataset. Figs. 8 and 9 show the visualization of multi-modal fusion on the LLVIP dataset [2], where the visible modality suffers from the limitation of low light. Clearly, the fused results of the comparative methods still

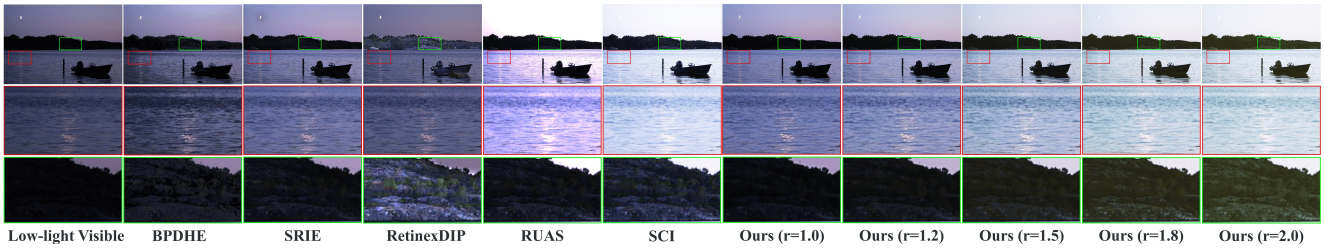

Figure 4. Visualization of low-light enhancement on the LLVIP dataset.

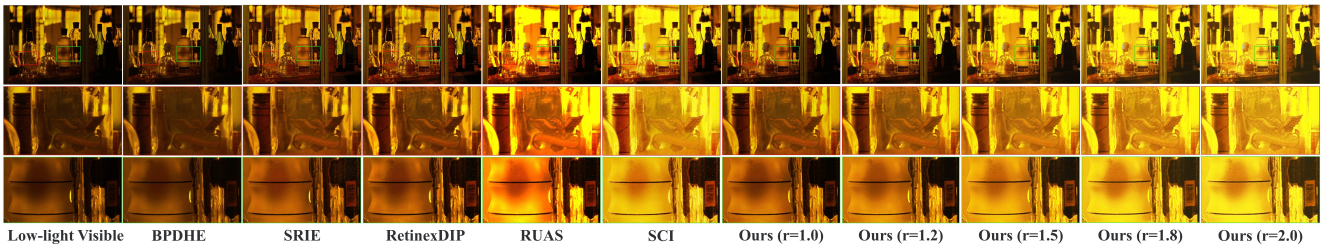

Figure 5. Visualization of low-light enhancement on the LLVIP dataset.

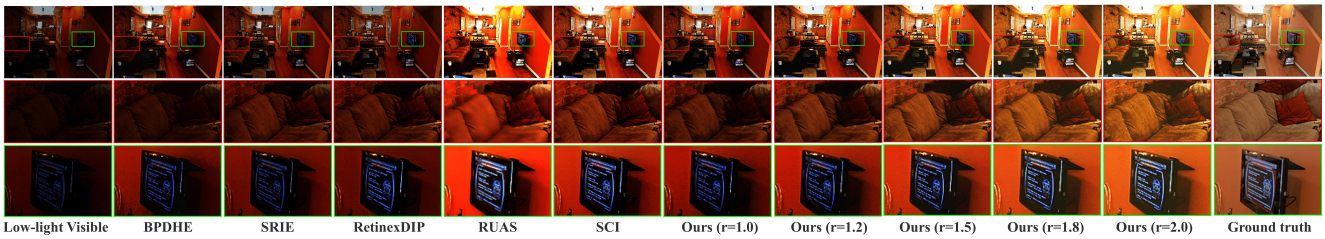

Figure 6. Visualization of low-light enhancement on the AGLIE dataset.

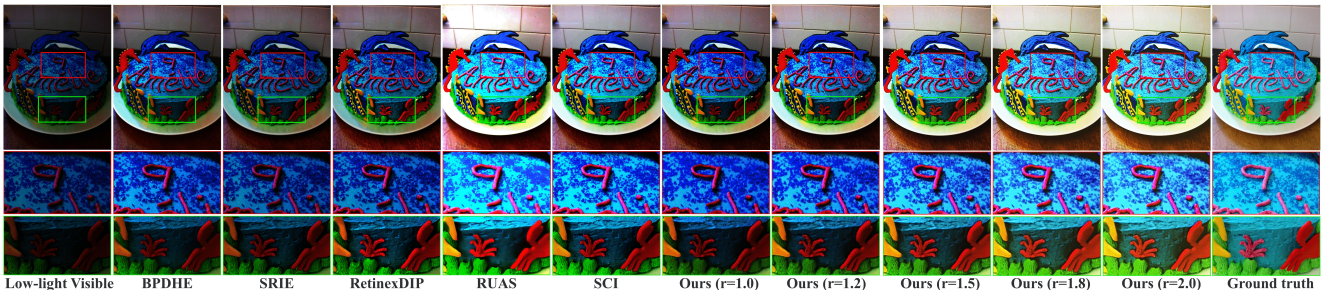

Figure 7. Visualization of low-light enhancement on the AGLIE dataset.

exhibit low visibility due to their failure to consider the loss of beneficial information in the low-light visible modality. In contrast, our DDBF is capable of preserving rich textures (*e.g.*, the traffic light in Fig. 9) and salient thermal objects (*e.g.*, the pedestrians in Fig. 8) by reducing the information mismatch between the infrared and visible modalities. Besides, our information fusion module is designed based on information saliency. Therefore, we can control the saliency of visible images by adjusting the enhancement ratio, thereby producing fused results with different appearances. Such a convenient adjustment strategy allows users to flexibly customize fused images that match their perceptual preferences.

MFNet Dataset. The visual results on the MFNet dataset [1] are presented in Figs. 10 and 11. Similarly, our method still achieves outstanding fusion performance. For example, in Fig. 10, the fused result of our method contains the clearest railing, while maintaining the thermal pedestrians and the background lightened trees. In Fig. 11, our method effectively illuminates the scene, presenting clear tree canopies, in which small infrared thermal objects are also preserved well.

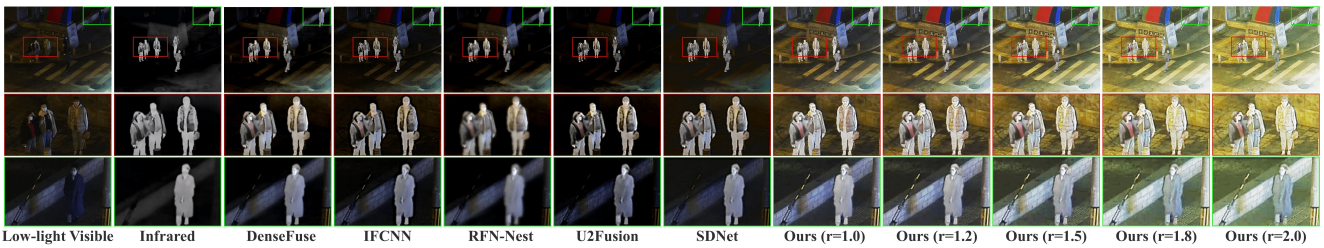

Figure 8. Visualization of multi-modal fusion methods on the LLVIP dataset.

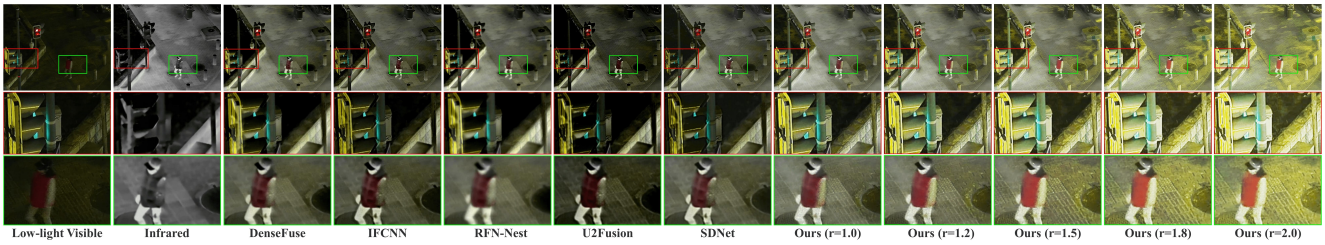

Figure 9. Visualization of multi-modal fusion methods on the LLVIP dataset.

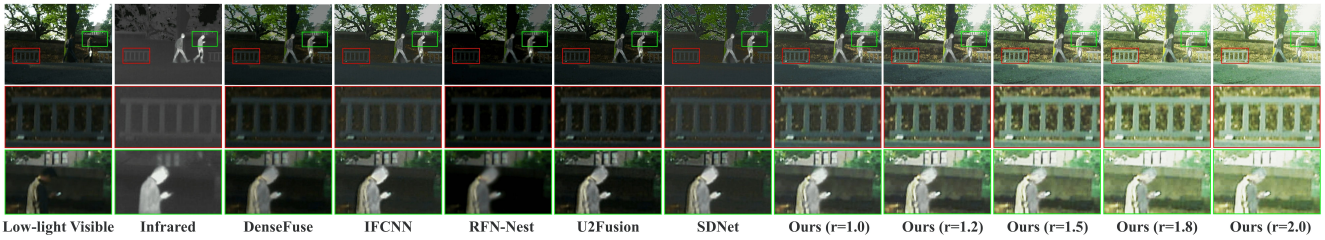

Figure 10. Visualization of multi-modal fusion methods on the MFNet dataset.

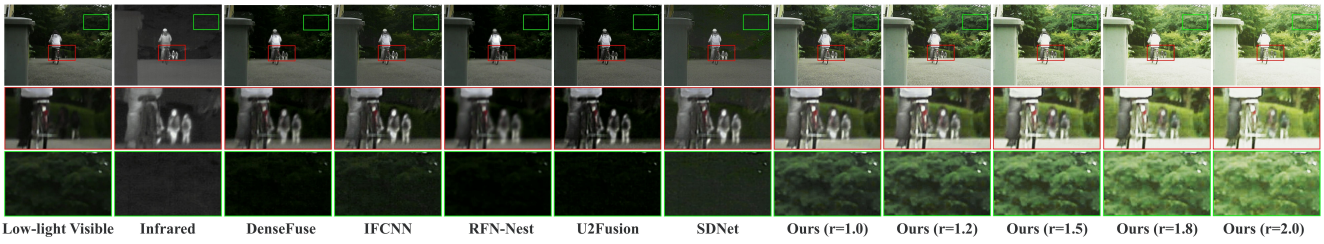

Figure 11. Visualization of multi-modal fusion methods on the MFNet dataset.

4.3. Visual Results of Daytime Multi-modal Fusion

Figs. 12 and 13 demonstrate the results of our method being applied to the daytime multi-modal data from RoadScene dataset [8]. It can be seen that our method exhibits good generalization performance, which can also deal with the daytime scenarios well. More concretely, our DDBF corrects the low contrast in daytime visible images and restores vivid colors, improving the visual quality. Furthermore, by adjusting the enhancement ratio, our obtained fused results maintain both the rich structure from the corrected visible images and the salience from the infrared images.

5. Different Numbers of Levels for Constructing The Infrared Sample Set

During the training process, we center on the original infrared image ($r = 1$) to construct a three-level infrared sample set ($r_1 < 1 = r_2 < r_3$). In this way, our proposed guided restoration module can be trained to have the functions of low-light enhancement and overexposure correction at the same time, and enable progressive illumination adjustment. We conduct experiments to verify the influence of the level number of the infrared sample set on our method, and the visual results are presented in Fig. 14. First, we use two different two-level infrared sample sets

Figure 12. Visual results of generalization to the daylight RoadScene dataset.

Figure 13. Visual results of generalization to the daylight RoadScene dataset.

($r1 < 1 = r2$, $r1 = 1 < r2$) to participate in the adversarial training. It can be seen that when two-level infrared sample sets are used, the trained model only has a single-sided controllable adjustment function (low-light enhancement or overexposure correction), and is not very sensitive to the changes of the enhancement ratio. In addition, we also construct a five-level infrared sample set ($r1 < r2 < 1 = r3 < r4 < r5$) to participate in the adversarial training of our method. Clearly, both the five-level infrared sample set and our used three-level one can guide the proposed model to achieve both low-light enhancement and overexposure correction in a controllable manner. Therefore, we choose a relatively simple 3-level configuration to achieve the desired function of image enhancement.

6. Comparison with Specialized Low-light Fusion Methods

Research on low-light visible and infrared image fusion is currently limited in the community. PIAFusion [6] and DIVFusion [7] are rare methods that consider lighting in image fusion tasks, so we compare our DDBF with these two methods on the LLVIP [2] and MFNet [1] datasets. PIAFusion [6] supplements lost information in low-light visible images by increasing the retention ratio of infrared information. However, it still cannot recover scene information from dark areas in visible images. DIVFusion [7] roughly enhances low-light visible images for better fusion representation, yet has inflexible and noticeable color distortion drawbacks. We compare our DDBF with PIAFusion and DIVFusion on the LLVIP [2] and MFNet [1] datasets to show the performance difference. Fig. 15 (a) validates the above points. Fig. 15 (a) validates the above points. In contrast, our method achieves flexible, high-color-fidelity fusion in low-light conditions. Table 2 objectively proves the advantages of our method.

Table 2. Quantitative comparison with specialized low-light fusion methods.

Dataset	LLVIP				MFNet			
Metric	MI ↑	VIF ↑	AG ↑	SD ↑	MI ↑	VIF ↑	AG ↑	SD ↑
PIAFusion	**3.113**	0.453	6.010	0.174	3.362	0.457	3.550	0.148
DIVFusion	2.120	0.459	5.587	**0.208**	2.658	0.470	4.410	**0.209**
Ours (r=1.0)	2.976	0.473	8.020	0.189	3.365	0.485	4.133	0.162
Ours (r=1.2)	2.914	0.486	8.634	0.197	3.378	**0.490**	4.398	0.176
Ours (r=1.5)	2.904	0.499	9.150	0.204	3.392	0.485	4.591	0.189
Ours (r=1.8)	2.910	**0.504**	**9.252**	0.200	3.405	0.480	**4.649**	0.193
Ours (r=2.0)	2.927	0.503	8.951	0.191	**3.435**	0.476	4.568	0.190

7. Comparison of Enhancement Plus Fusion

Further, we use the state-of-the-art low-light enhancement method SCI [5] as a precursor to other comparative fusion

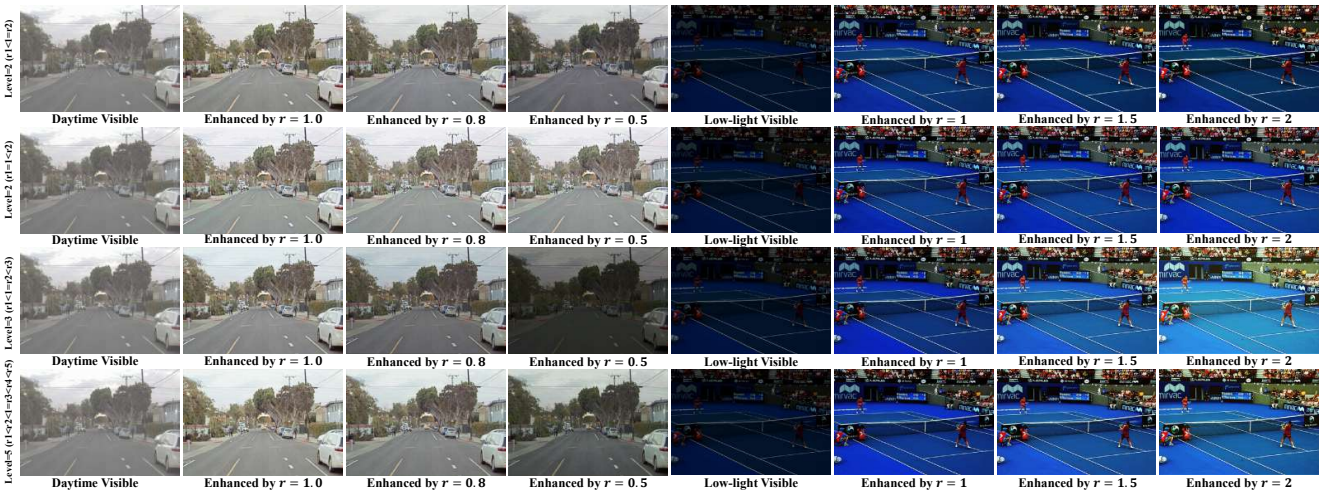

Figure 14. Analysis of the levels in the infrared sample set.

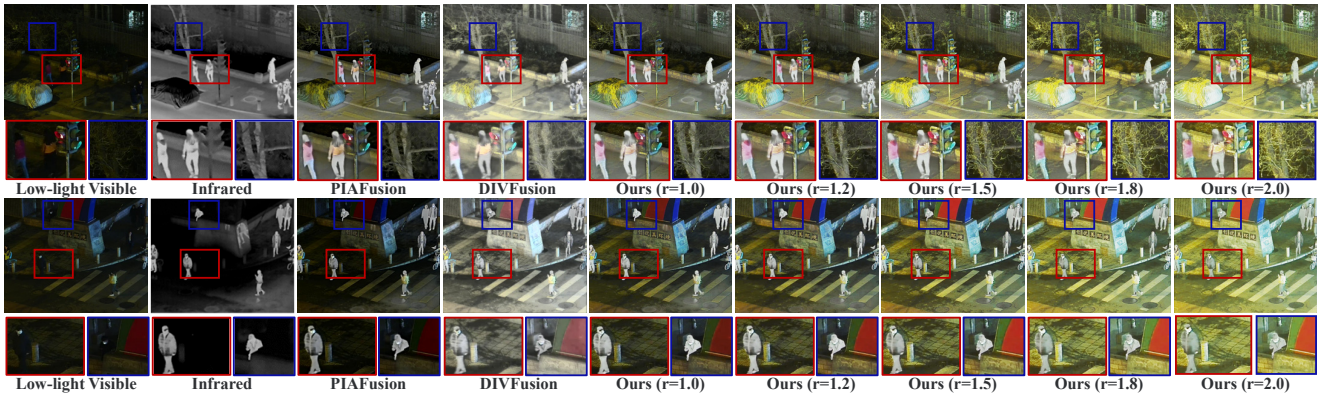

Figure 15. Visual comparison with specialized low-light fusion methods.

methods, reducing the information mismatch in low-light environments. Thus, a comparison of enhancement plus fusion is performed on the LLVIP dataset [2], and the visual results are shown in Figs. 16 and 17. It clearly demonstrates that existing fusion methods fail to achieve satisfactory lighting and color fidelity, even when the input visible image is enhanced. Moreover, we conduct quantitative analysis using objective metrics, as presented in Table 3. Our information fusion module outperforms the other fusion methods across a majority of the metrics, substantiating its effectiveness and superiority.

Table 3. Quantitative comparison of enhancement plus fusion.

Dataset	LLVIP			
Metric	MI ↑	VIF ↑	AG ↑	SD ↑
DenseFuse	2.354	0.446	6.027	0.183
IFCNN	2.359	0.475	**10.051**	0.197
RFN-Nest	2.100	0.442	4.591	0.188
U2Fusion	2.025	0.425	7.841	0.194
SDNet	2.000	0.344	7.106	0.149
PIAFusion	3.113	0.453	6.010	0.174
DIVFusion	2.120	0.459	5.587	0.208
Ours (r=1.0)	**2.976**	0.473	8.020	0.189
Ours (r=1.2)	2.914	0.486	8.634	0.197
Ours (r=1.5)	<u>2.904</u>	0.499	9.150	**0.204**
Ours (r=1.8)	2.910	**0.504**	**9.252**	<u>0.200</u>
Ours (r=2.0)	2.927	<u>0.503</u>	8.951	0.191

8. Additional Ablation Studies

To more fully validate the effectiveness of the specific designs in our DDBF, we implement additional ablation studies. Firstly, this paper employs Gaussian low-pass filtering for blurring (following the famous single-scale Retinex), so as to satisfy the local smoothness properties. We evaluate the impacts of it by using other blurring operations, including mean and bilateral filtering. Secondly, we remove the scene fidelity loss of the GRM to see its role. Thirdly, an automatic estimation strategy of enhancement ratio is developed for comparison. Specifically, five users are invited to adjust "r" according to their aesthetics on 300 images with

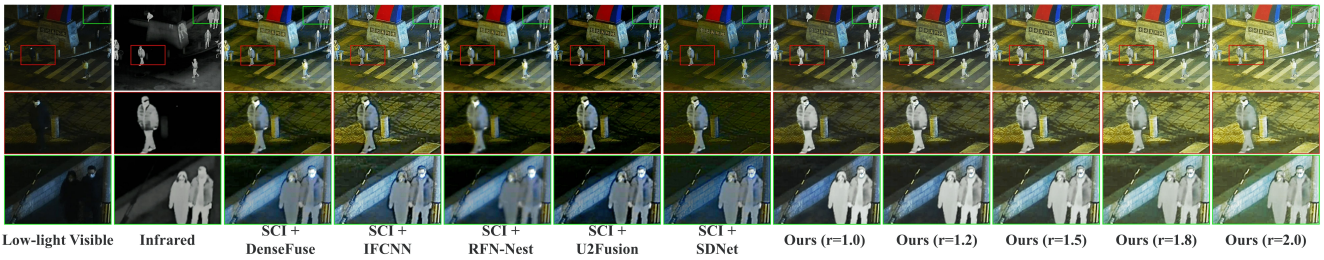

Figure 16. Visual comparison of enhancement plus fusion.

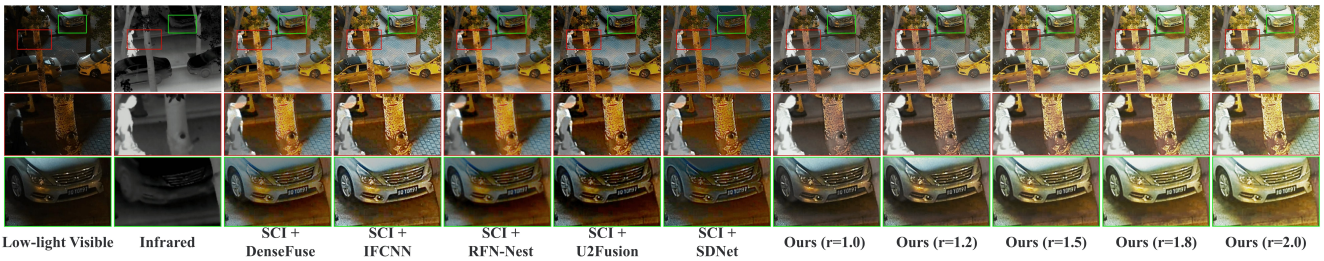

Figure 17. Visual comparison of enhancement plus fusion.

progressive brightness, and the results are shown in Fig. 18. Intuitively, there is a clear inverse correlation between the user-selected "r" and the mean intensity. Thus, a rich dictionary of enhancement ratios for images covering different brightness levels was established. Based on the current input image's average intensity, we can identify its position in the dictionary and interpolate the ratios from the nearby Top 5 samples to obtain the tailored "r".

The visual comparison on the AGLIE dataset [4] is presented in Fig. 19. Obviously, the more thorough the scene texture removal, the better the enhancement performance. This shows that the removal of scene texture can prompt GAN to focus more on distinguishing and adjusting brightness. Besides, removing the scene fidelity loss causes complete color deviation, proving its protective effect on scene colors. Finally, the automatically generated "r" achieves comparable performance to manual selection. The dictionary will improve with expanded users and images in the future to reduce bias. Moreover, basing judgments on a more comprehensive basis rather than just mean intensity would also help to provide a more accurate estimate of the enhancement ratio "r". The quantitative results in Table 4 further support the above conclusions.

Table 4. Quantitative results of low-light enhancement.

		Manual			Automatic
Dataset	Metric	Mean	Bilateral	Ours	Ours
AGLIE	SSIM ↑	0.658	0.547	**0.706**	0.691
	PSNR ↑	15.747	13.635	**16.434**	16.088

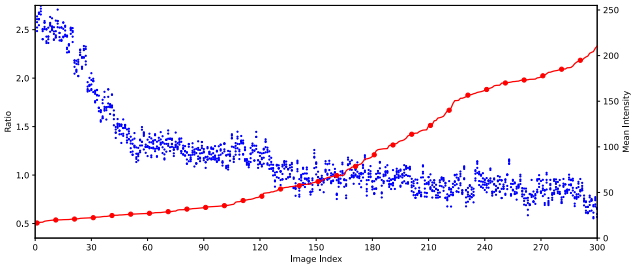

Figure 18. User research on enhancement ratio "r".

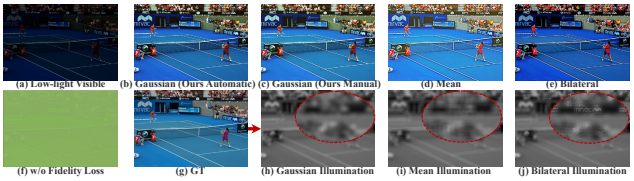

Figure 19. Visual results of additional ablation studies.

References

[1] Qishen Ha, Kohei Watanabe, Takumi Karasawa, Yoshitaka Ushiku, and Tatsuya Harada. Mfnet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5108–5115, 2017. 3, 5

[2] Xinyu Jia, Chuang Zhu, Minzhen Li, Wenqi Tang, and Wenli Zhou. Llvip: A visible-infrared paired dataset for low-light vision. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3496–3504, 2021. 2, 5, 6

[3] Yuen Peng Loh and Chee Seng Chan. Getting to know low-

light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 2

[4] Feifan Lv, Yu Li, and Feng Lu. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *International Journal of Computer Vision*, 129(7): 2175–2193, 2021. 2, 7

[5] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5637–5646, 2022. 5

[6] Linfeng Tang, Jiteng Yuan, Hao Zhang, Xingyu Jiang, and Jiayi Ma. Piafusion: A progressive infrared and visible image fusion network based on illumination aware. *Information Fusion*, 83:79–92, 2022. 5

[7] Linfeng Tang, Xinyu Xiang, Hao Zhang, Meiqi Gong, and Jiayi Ma. Divfusion: Darkness-free infrared and visible image fusion. *Information Fusion*, 91:477–493, 2023. 5

[8] Han Xu, Jiayi Ma, Junjun Jiang, Xiaojie Guo, and Haibin Ling. U2fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Mmachine Intelligence*, 44(1):502–518, 2022. 4