Supplementary Material for Text-IF: Leveraging Semantic Text Guidance for Degradation-Aware and Interactive Image Fusion

Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, Jiayi Ma*
Electronic Information School, Wuhan University, Wuhan 430072, China
{yixunpeng, xu_han}@whu.edu.cn, {zhpersonalbox, linfeng0419, jyma2010}@gmail.com

1. Hyperparameters in Loss Functions of Different Tasks

For each of the fusion tasks, we design various hyperparameter settings to fit the requirements. Overall, it can be expressed as:

$$L_{task} = \{\alpha_{int}(t), \alpha_{SSIM}(t), \alpha_{grad}(t), \alpha_{color}(t)\}. \quad (1)$$

Specifically, for each degradation task, it is set as:

$$L_{low-light} = \{8, 1, 10, 12\}, L_{overexposure} = \{4, 0, 2, 12\}, \quad (2)$$

$$L_{denoise} = \{6, 1, 10, 12\}, L_{low-contrast} = \{8, 1, 10, 12\}, \quad (3)$$

where the hyperparameters are the same meanings in the Section 3.4 *loss Functions* in the paper.

2. Network Architecture

We provide detailed network architecture setting in Tab. r1. Specifically, it includes the Image Fusion Pipeline and the Text Interaction Guidance Architecture (TIGA). The adopted text semantic encoder is derived from CLIP-L [3].

Table r1. Detailed Module Setting of the Network Architecture in Text-IF

Parameters Name	Value
dim (Image Fusion Pipeline)	48
num-blocks (Image Fusion Pipeline)	[2, 2, 2, 2]
heads (Image Fusion Pipeline)	[1, 2, 4, 8]
text embedding dim (TIGA)	512

In detail, we present the multi-level detailed architecture of the semantic interaction fusion decoder. The overall module is composed of multiple architectures as shown in Fig. r1.

*Corresponding author

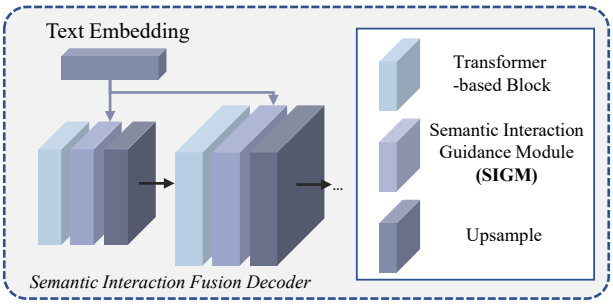

Figure r1. Detailed architecture of the semantic interaction fusion decoder, where the Transformer-based block, SIGM, and upsampling are coupled through cascading.

3. Comprehensive Semantic Interaction

Compound Text: The proposed method can be flexibly applied to various scenarios. When the requirements need to be changed, only a few words are needed to be inserted to achieve the desired effect. Simply, we take a new degradation affect as an example, as shown in Fig. r2. Besides the denoising requirements, we only add a few words to achieve the effect of the denoising and low-light enhancement at the same time. Similarly, we can make more flexible combinations through modifying the text.

Obviously, the actual results are consistent with the semantic information of the text. It is easy to remove various kinds of degradation without changing the model. All you need is a text.

4. More Comparison Experiments

In fact, the qualitative contrast of text semantic guidance image fusion is multifaceted. To further verify the effectiveness of the proposed method, we present more flexible examples in Figs. r3-r12.

Similarly, the state-of-the-art methods utilized for comparison mainly include UMF-CMGR [4], TarDAL [2], ReCoNet [1], MURF [6], U2Fusion [5], MetaFusion [7], and DDFM [8].

Specifically, the qualitative experiment of fusion directly

Figure r2. Comparative experiments on the compound text for the compound degradation addressing. Noting that only some words are added, it can be achieved that compound degradations process in the image. (**Please zoom in for the better reviewing.**)

without text guidance and the degradation perception with the interactive text are included. The specific settings are the same of the Section 4.1 *Implementation Details and Datasets* in the paper.

In both the text semantic guidance and no text semantic guidance conditions. the proposed Text-IF has significant advantages over the state-of-the-art methods. In short, Text-IF has a more comprehensive scene representation, clearer salient objectives, and more flexible fusion results, which proves the effectiveness, flexibility, and interactivity of the proposed method.

References

[1] Zhanbo Huang, Jinyuan Liu, Xin Fan, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Reconet: Recurrent correction network for fast and efficient multi-modality image fusion. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 539–555, 2022. 1

[2] Jinyuan Liu, Xin Fan, Zhanbo Huang, Guanyao Wu, Risheng Liu, Wei Zhong, and Zhongxuan Luo. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5802–5811, 2022. 1

[3] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *Proceedings of the International Conference on Machine Learning*, pages 8748–8763, 2021. 1

[4] Di Wang, Jinyuan Liu, Xin Fan, and Risheng Liu. Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, 2022. 1

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

[6] Han Xu, Jiteng Yuan, and Jiayi Ma. Murf: Mutually reinforcing multi-modal image registration and fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10): 12148–12166, 2023. 1

[7] Wenda Zhao, Shigeng Xie, Fan Zhao, You He, and Huchuan Lu. Metafusion: Infrared and visible image fusion via meta-feature embedding from object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 13955–13965, 2023. 1

[8] Zixiang Zhao, Haowen Bai, Yuanzhi Zhu, Jiangshe Zhang, Shuang Xu, Yulun Zhang, Kai Zhang, Deyu Meng, Radu Timofte, and Luc Van Gool. Ddfm: denoising diffusion model for multi-modality image fusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8082–8093, 2023. 1

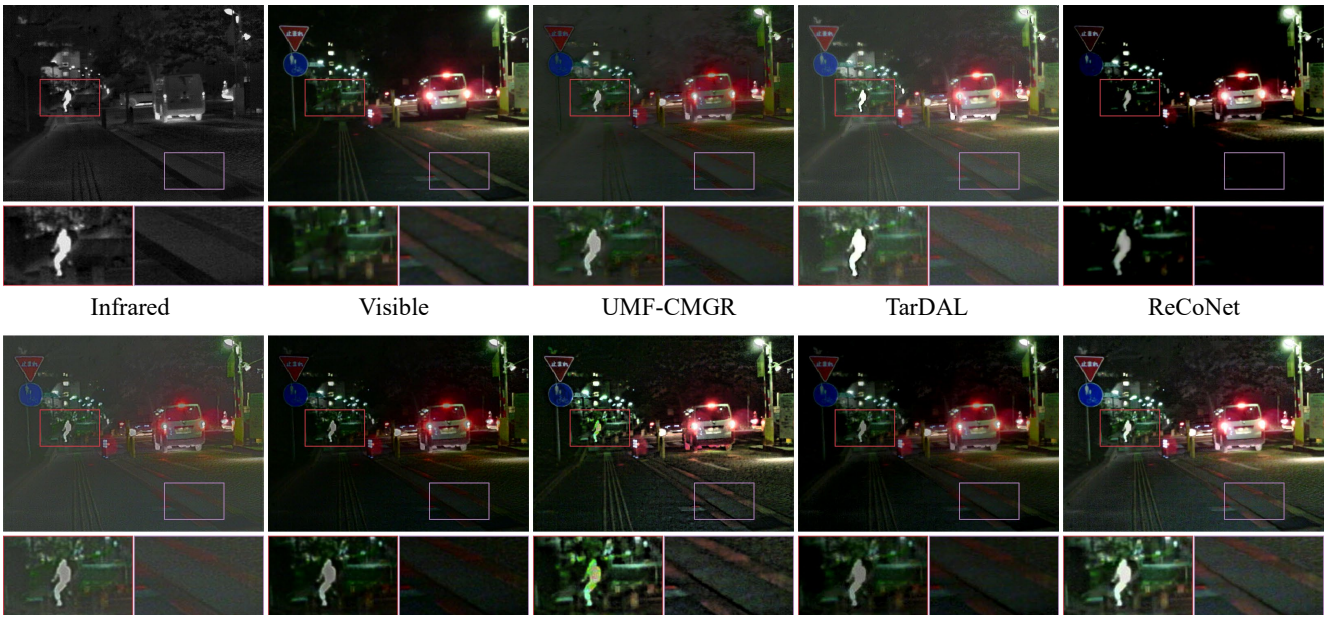

Figure r3. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

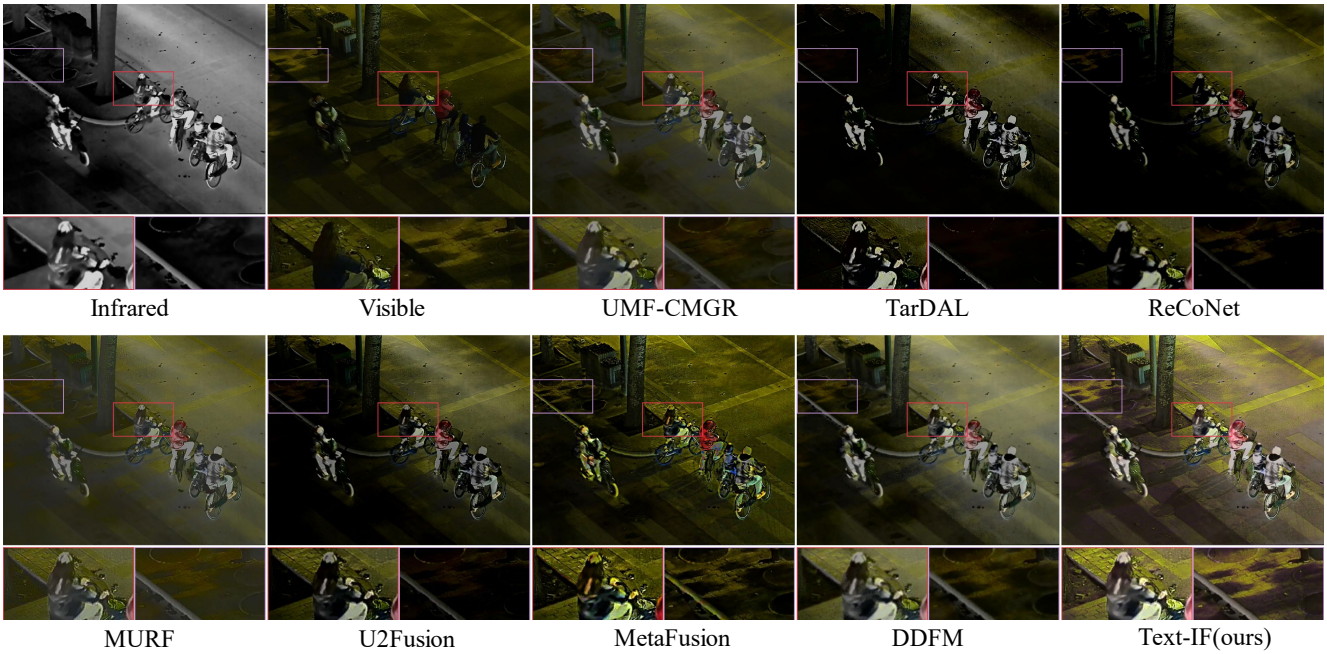

Figure r4. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

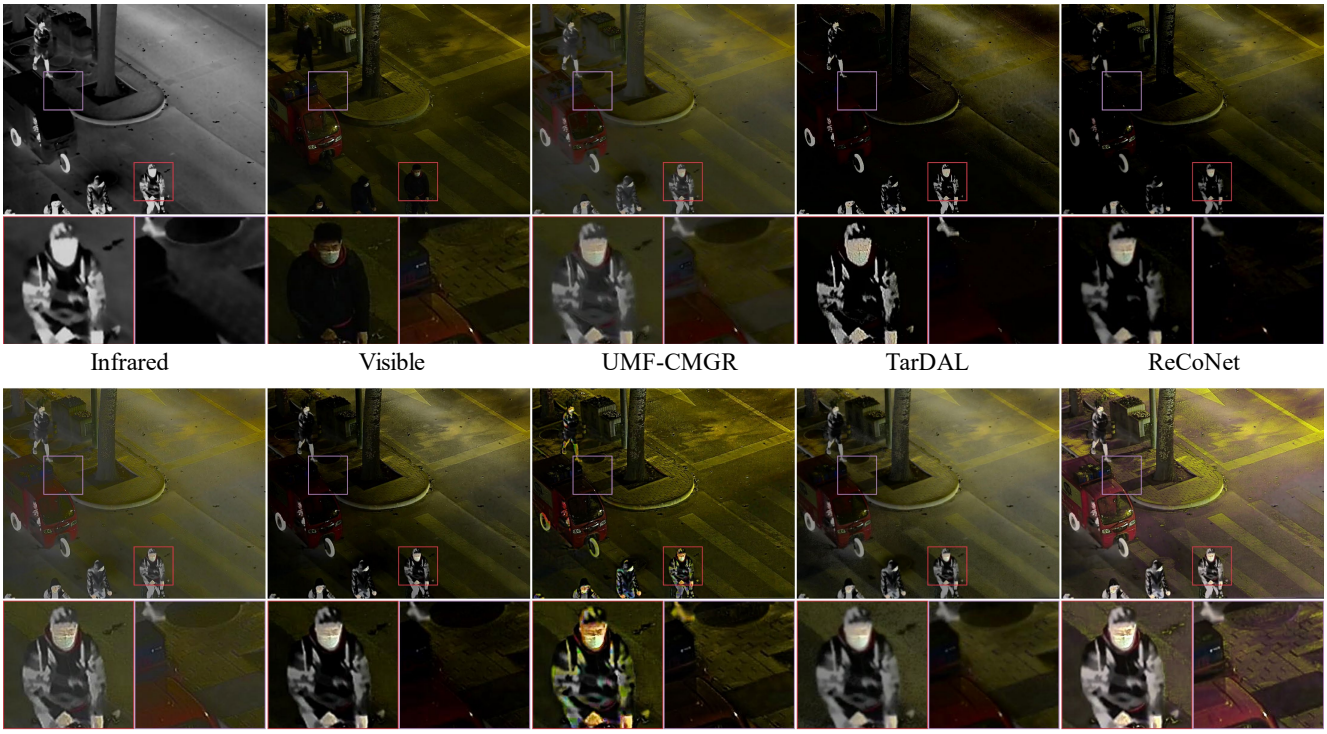

Figure r5. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

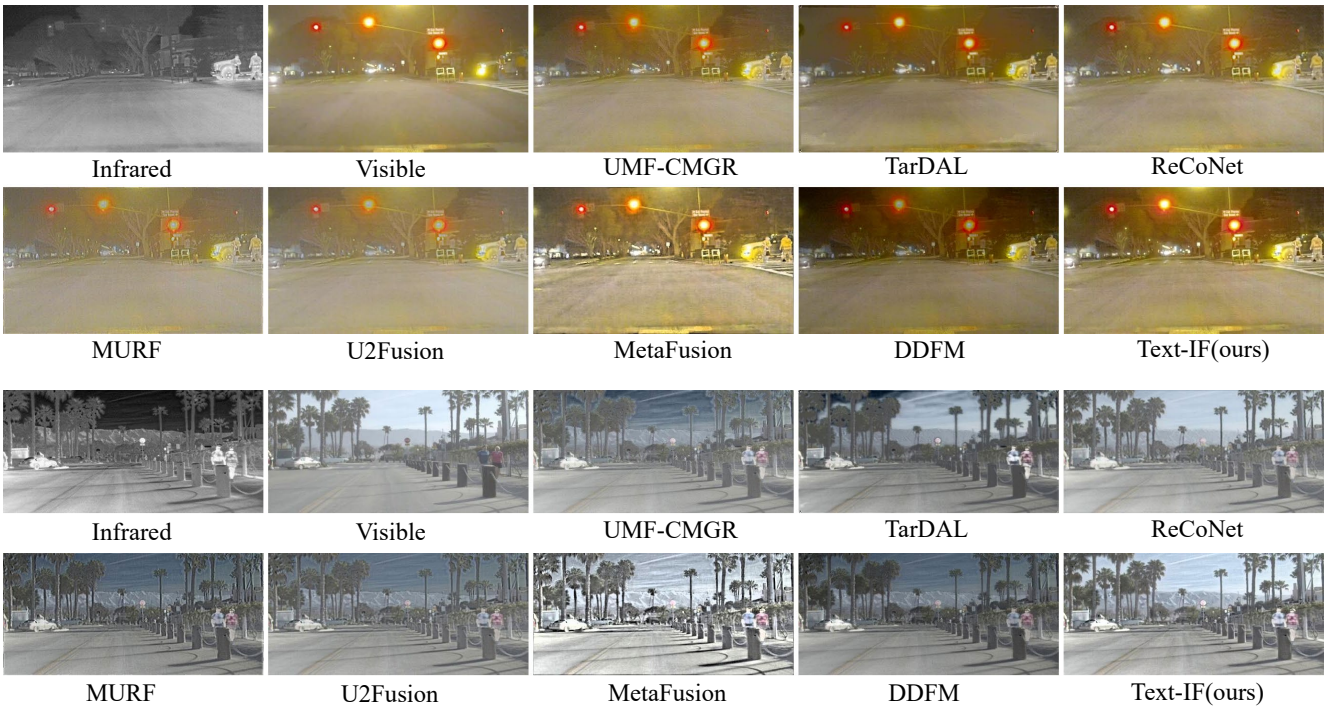

Figure r6. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

Text: *In the context of infrared-visible image fusion, visible images may suffer from reduced quality in low-light scenarios.*

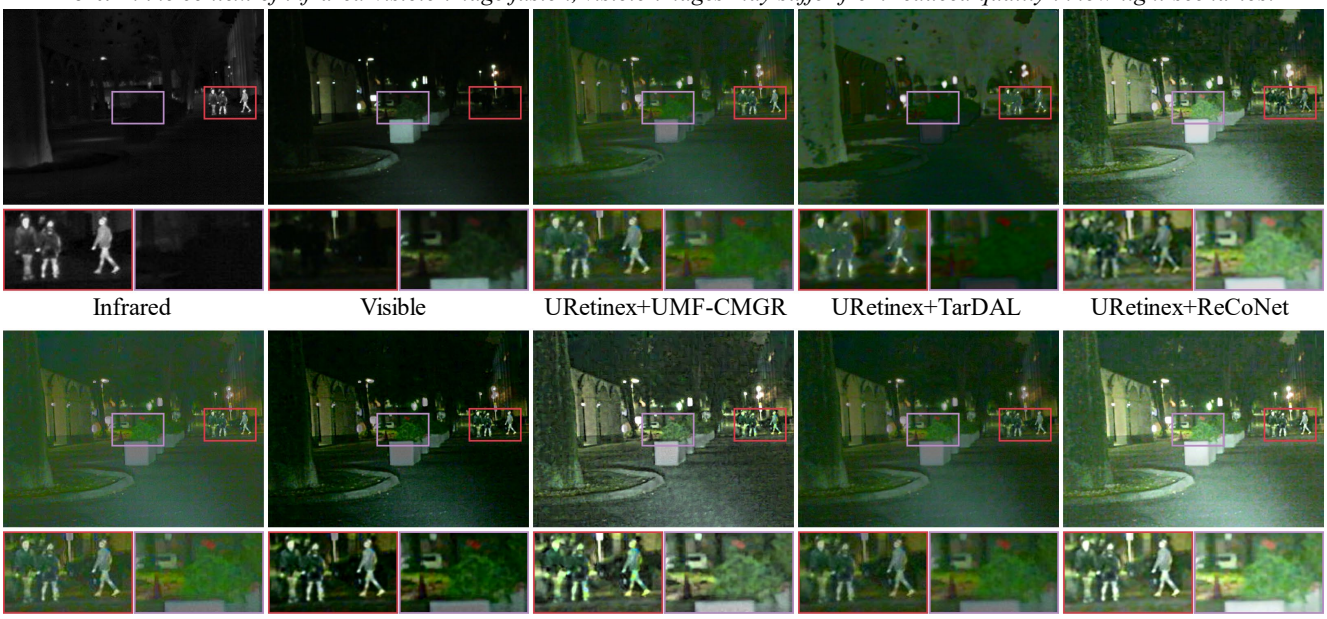

Figure r7. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

Text: *In the context of infrared-visible light fusion, visible images may suffer from reduced quality in low-light scenarios.*

Figure r8. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

Text: *We're dealing with the fusion of infrared and visible light images, and it's essential to combat the low contrast degradation in the infrared images.*

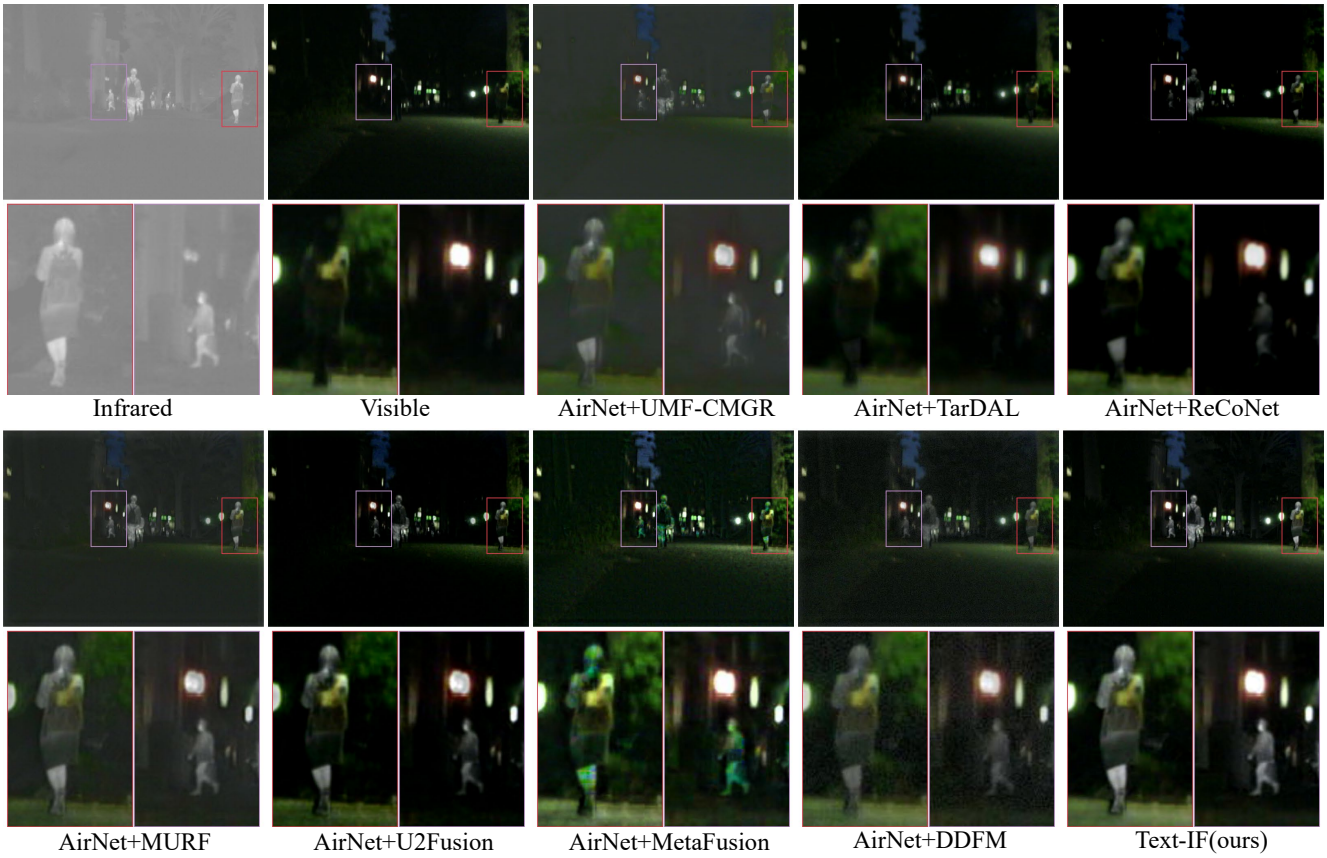

Figure r9. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

Text: *In this particular problem, we are tasked with fusing infrared and visible light images, and it's essential to address the low contrast degradation in the infrared images.*

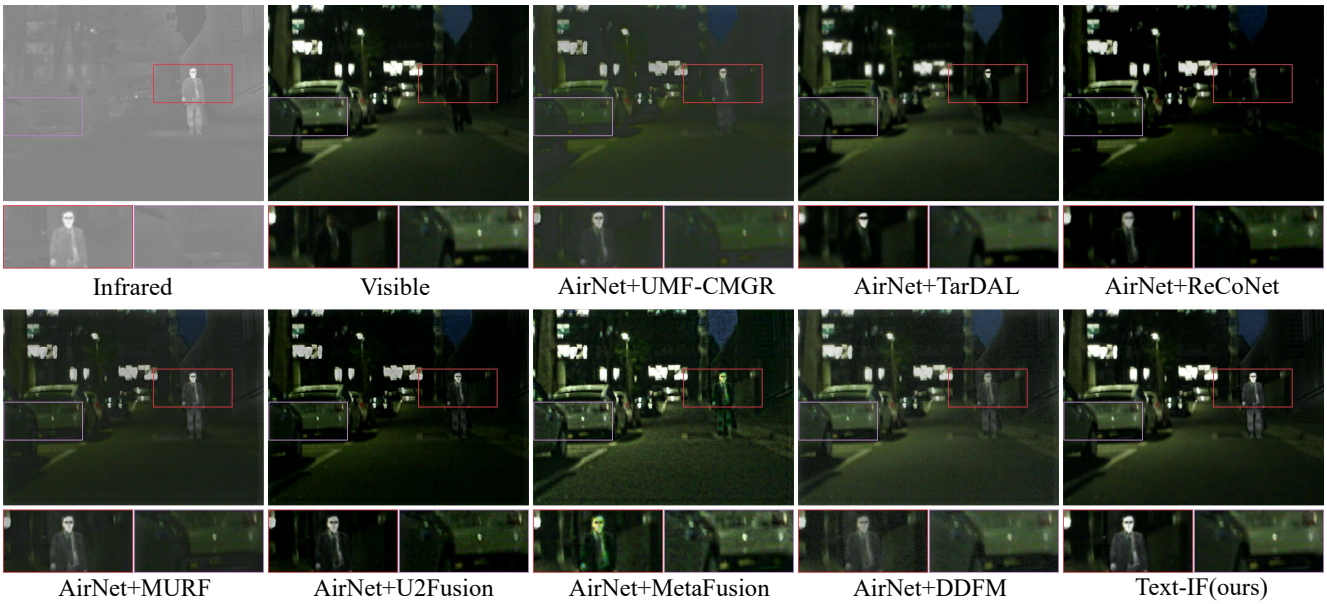

Figure r10. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

Text: *Infrared-visible light image fusion is our focus, and it's crucial to address the noise degradation in the infrared images.*

Figure r11. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)

Text: *The task involves the fusion of infrared-visible light images, and it's essential to address the noise degradation issue in the infrared images.*

Figure r12. More Comparison experiments on infrared and visible image fusion. (**Please zoom in for the better reviewing.**)