Supplementary Materials for the LE2Fusion

Yongbiao Xiao, Hui Li[⋆], Chunyang Cheng, and Xiaoning Song

International Joint Laboratory on Artificial Intelligence of Jiangsu Province, School of Artificial Intelligence and Computer Science, Jiangnan University, 214122, Wuxi, China
lihui.cv@jiangnan.edu.cn

1 The detail of LE2 network architecture

Our network contains three parts: Feature extractor (MRA and DenseBlock), LE2 module, Image reconstructor(Edge Weight Network and Reconstructor). The details are given as follows.

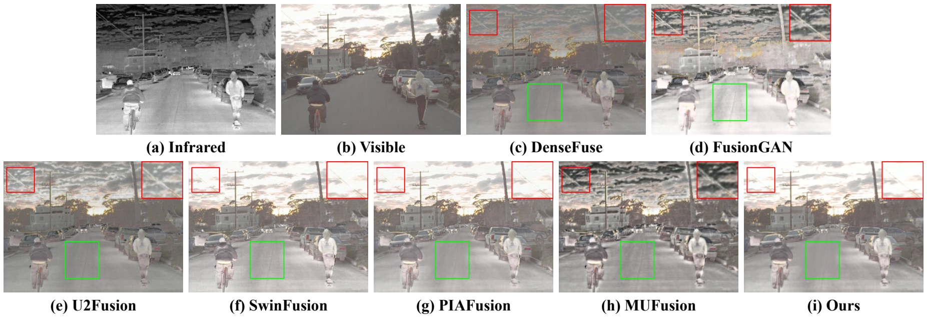
Table 1. Network architecture of our LE2Fusion. **Input** and **Output** denote the number of channels in the corresponding feature maps.

	Layers	Kernel	Input	Output	Activation
MRA Module	Layer1	1×1	1	16	LReLU
	Layer2	5×5	1	16	LReLU
DenseBlock	Layer1	3×3	16	16	LReLU
	Layer2	3×3	32	16	LReLU
	Layer3	3×3	48	16	LReLU
LE2 Module	Layer1	5×5	3	32	LReLU
	Layer2	3×3	32	64	LReLU
	Layer3	3×3	64	128	LReLU
	Layer4	3×3	128	64	LReLU
	Layer5	3×3	64	32	LReLU
	Layer6	3×3	32	2	LReLU
Edge Weight Network	Layer1	1×1	2	128	Sigmoid
	Layer2	1×1	128	64	Sigmoid
	Layer3	1×1	64	32	Sigmoid
	Layer4	1×1	32	1	Sigmoid
Image Reconstructor	Layer1	3×3	128	128	LReLU
	Layer2	3×3	128	64	LReLU
	Layer3	3×3	64	32	LReLU
	Layer4	1×1	32	1	Tanh

[⋆] Corresponding author

2 More experimental results and analysis

It is well known that the generalization ability refers to the ability of the algorithm to adapt to fresh samples. Therefore, we provide generalization experiments performed on RoadScene [1] dataset, which are used to verify the generalization ability of our method.



(a) Infrared (b) Visible (c) DenseFuse (d) FusionGAN

(e) U2Fusion (f) SwinFusion (g) PIAFusion (h) MUFusion (i) Ours

Fig. 1. Qualitative comparison of our method with six state-of-the-art methods on the RoadScene dataset. For a clear comparison, we select a texture region and zoom in it in the red box and highlight a salient region in the green box

Qualitative results The qualitative comparisons of different algorithms on the RoadScene dataset are shown in Fig. 1. In general, we can obviously observe that the thermal radiation of infrared targets is weakened from DenseFuse, U2Fusion and MUFusion, resulting in the overall darker image. FusionGAN blurs the edges of the target and severe spectral contamination appeared in the background region.

As shown in the red box, DenseFuse, FusionGAN, U2Fusion, and MUFusion are suffering from varying degrees of spectral contamination. Moreover, artifacts and distorted utility poles are showed in PIAFusion. Our method and SwinFusion preserve the sky texture detail information better. In addition, as shown in the green boxes, although SwinFusion and PIAFusion retain the details of the ground, they introduce noise that hinders human vision, resulting in too obvious thermal radiation of the ground in the fused images and affecting our senses. In our method, the edges of the local region are processed, and the pixel intensity loss function based on local region is designed to capture more meaningful information. Therefore, the visual noise caused by a lot of ground thermal radiation is reduced, and more texture details are retained.

The visualization results on various datasets demonstrate the advantages of our algorithm in local region edge feature extraction and texture preservation. We attribute the advantages to the following aspects. On the one hand, our local edge enhancement module we designed enhances local region edge information. On the other hand, we innovatively design the pixel intensity loss function to constrain the network to extract meaningful information according to the local region.

Quantitative results We randomly select 44 image pairs from the RoadScene for quantitative evaluation. The comparative results of different methods on the five metrics are shown in Table 2. Our method achieves the largest values in SD, meaning that our method can achieve a good visual effect. For MI, our fused images are slightly lower than PIAFusion, which is justifiable. Specifically, the local region features extracted by our method sacrifice some information to improve the overall visual effect. However, for EN, Q_{abf} and SCD, their performances are not as prominent as that on MSRS and LLVIP datasets. The reason for this is that the RoadScene dataset mainly contains strongly light scenes, even nighttime images, resulting in the effect of the local meaningful edge features we extract is not obvious. Although our metric is weaker after comparison with other metrics, our method can achieve optimality in terms of visual effects.

Table 2. Quantitative results on 44 image pairs from the RoadScene dataset. (**Bold**: Best, Red: Second Best, Blue: Third best)

Methods	SD	EN	MI	SCD	Q_{abf}
DenseFuse	9.6416	6.8958	2.9746	1.2099	0.3738
FusionGAN	10.0835	7.0942	2.4544	0.4270	**0.4488**
U2Fusion	9.8771	7.0498	3.0781	1.4453	0.4061
SwinFusion	10.1891	6.9561	3.6647	**1.4600**	0.4403
PIAFusion	10.1844	6.9845	**3.8336**	1.2274	0.4374
MUFusion	10.2943	**7.4382**	2.4618	1.3929	0.3656
Ours	**10.4220**	6.9800	3.7347	1.1869	0.3919

In a word, our method ranks first in all the SD metrics, which indicates that our method can achieve the best visual effects, thanking to the proposed LE2 module and pixel intensity loss function based on local region.

References

1. Xu, H., Ma, J., Jiang, J., Guo, X., Ling, H.: U2fusion: A unified unsupervised image fusion network. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(1), 502–518 (2020)