EdgeFusion: Infrared and Visible Image Fusion Algorithm in Low Light

Zikun Song, Pinle Qin[✉], Jianchao Zeng, Shuangjiao Zhai, Rui Chai, and Junyi Yan

North University of China, Taiyuan, China
qpl@nuc.edu.cn

Abstract. Under low-light conditions, visible light imaging technology exhibits poor imaging performance, whereas infrared thermal imaging technology can effectively detect and identify targets. To solve the target imaging problem in low-light environments, multimodal image fusion technology can combine the advantages of both aforementioned methods. Existing fusion methods focus excessively on information from infrared images, obscuring the original texture details of the targets and resulting in low-quality images. Therefore, in this study, we propose a multi-scale edge-fusion network for infrared and visible images called EdgeFusion, which can produce an edge-fusion image. Specifically, the network utilises infrared multiscale gradient information to enhance the edges of the thermal targets, thereby improving the ability to identify them. By designing a balanced loss, EdgeFusion suppresses the global information from infrared images that obscured the fine texture details of the original images. In addition, a residual gradient method is introduced to enhance the textural details of the generated images. After extensive experimentation on the public datasets LLVIP and TNO, the results indicate that EdgeFusion outperforms existing state-of-the-art methods in preserving fine-grained infrared edges and enhanced image texture details.

Keywords: Image fusion · Convolution network · Multiscale gradient retention module

1 Introduction

Image fusion techniques [1,2] integrate multimodal images into a single image, combining the advantages of different modalities. Under low-light conditions, visible images struggle to differentiate between background and targets, whereas infrared images lack texture information. Combining the two can generate images

that highlight targets and preserve texture. These advantages enable a wide range of applications in military security [3], target detection [4], semantic segmentation [5], and other fields.

Over decades of development, various methods for infrared and visible image fusion have emerged. Among them, traditional methods [6–8] involve three steps: feature extraction, fusion, and inverse transformation. However, owing to coarse fusion rules such as the maximum and mean value strategies, these methods face challenges in adapting to feature specificity, leading to information distortion. In recent years, fusion methods based on deep learning have shown better results; they include autoencoders [4,9–12], CNN (convolutional neural networks) [13–15], and GAN (generative adversarial networks) [4,16–19]. Although current deep learning-based image fusion algorithms can generate images with complementary information, they face significant challenges in preserving edges and enhancing fine texture details. As shown in Fig. 1, Fig. 1(a) and Fig. 1(b) respectively represent the infrared image and visible image captured in low-light conditions, Fig. 1(c) represents the fusion results using existing methods, which focus on the overall infrared information of foreground thermal targets, leading to complete highlighting of the targets and obscuring certain color and texture details [20]. Figure 1(d) depicts the effect of edge-fusion image, which effectively distinguishes the targets from the background and enhances the person's color and texture details, and facilitates subsequent tasks.
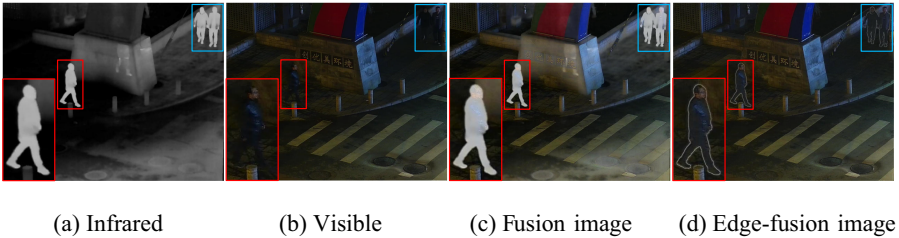

(a) Infrared   (b) Visible   (c) Fusion image   (d) Edge-fusion image

Fig. 1. Fusion results of infrared and visible images under low-light conditions.

To get the edge-fusion images, we propose a multiscale edge-fusion network EdgeFusion, for infrared and visible images. This method extracts the infrared thermal radiation gradient information and visible texture information for fusion. Specifically, a gradient operating module (GOM) has been introduced to enable feature reuse and enhance fine-grained details through residual gradient flow. This enables the extraction of rich texture information from visible images. Moreover, for the infrared image, a multiscale gradient retention module (MGRM) is designed to merge features from different scales of convolution and gradient operations, which allows the extraction of the gradient information from the infrared images. Additionally, to prevent the dominant infrared information from overshadowing the target texture, a balance loss is proposed to achieve convergence between the intensity and texture losses, allowing for the coexistence of intensity

and texture information. This significantly preserved the rich information in the original image. In summary, this paper provides the following key contributions:

- EdgeFusion is designed for infrared and visible images. This network effectively combines fine texture information from visible images with edge gradient information from infrared images, producing fusion images that are more visually appealing in low-light environments.
- An MGRM is proposed to extract fine-grained infrared edge information and accurately distinguish targets from the background. The design of the balance loss helps preserve the internal texture details of the visible image while reducing the interference from infrared information and increasing the amount of preserved information.
- Extensive experiments were conducted on the LLVIP and TNO datasets, and the outcomes were qualitatively and quantitatively compared with that of the state-of-the-art methods. The experimental results verify the effectiveness of EdgeFusion, and the experimental results for target detection demonstrate the benefits of this method for downstream tasks.

2 Related Work

2.1 AE-Based Image Fusion Methods

Methods based on autoencoder networks have strong interpretability, and their feature fusion processes rely mainly on manually designed fusion rules. Li et al. proposed DenseFuse [10], which is composed of three layers: encoding, fusion, and decoding. Owing to the limited feature extraction capability of autoencoders, Li et al. proposed NestFuse [4], which introduces a nested connection network and a spatial attention model. They also proposed spatial and channel attention as fusion strategies. RFN-Nest [11] employs a novel loss function for detail preservation and a loss function for feature enhancement, integrating more information into the network. DRF [12], proposed by Xu et al., employs various fusion strategies to decompose the original image into representations associated with scene and sensor modalities.

2.2 CNN-Based Image Fusion Methods

Methods based on CNNs use specially designed network architectures and loss functions for feature extraction, fusion, and image reconstruction. DeepFuse [13] exhibits robustness to varying input conditions and can handle extreme exposure without artefacts, thereby addressing the limitations of multiexposure image fusion. Zhang et al. proposed the IFCNN [14], which is a generic image fusion framework that utilises a linear element fusion rule to fuse the convolutional features of multiple input images and designs a universal loss function for various fusion tasks. Xu et al. have designed a unified multitask fusion framework [21] that can perform different types of fusion tasks. STDFusionNet [15] leverages the salient characteristics of thermal targets in infrared images to guide fusion using saliency masks.

2.3   GAN-Based Image Fusion Methods

GANs are unsupervised learning frameworks highly suitable for image fusion. FusionGAN, proposed by Ma et al. [4], was the first model to introduce GANs into the field of image fusion. However, owing to the issue of mode collapse, Ma et al. further proposed DDcGAN [17], which utilises a dual-discriminator GAN to improve the robustness and ensure that the fused image retains the characteristics of both source images simultaneously. Additionally, Ma et al. introduced a multiclassification-constrained GAN [18]. Li et al. designed a GAN that incorporates a multiscale attention mechanism, enabling the network to focus more on the key information of both infrared and visible images [19].

However, current existing fusion methods of various kinds emphasise the usefulness of global infrared information while neglecting situations in which the texture information of the fused image is obscured. Therefore, it is necessary to design a method that preserves the internal details and highlights the edge information of thermal targets. This study innovatively extracts edge information from an infrared image and fuses it with texture information from a visible image, resulting in fused images that contain the infrared edges of thermal targets while preserving the details and textures of the image.

3   Design of EdgeFusion

3.1   Problem Formulation

In low-light environments, distinguishing between the targets and backgrounds in visible images is extremely challenging. To address this issue, an MGRM is specifically designed for infrared images to extract multiscale edges. Consider a pair of visible and infrared images denoted as I_{vi} and I_{ir}, respectively. The visible image is converted into the YCrCb colour space and only the Y channel is extracted and denoted as I_{vi}^Y. The GOM [22] feature extraction module, denoted as E_G, was used to extract texture features from the visible image. The MGRM is denoted as E_V. Thus, it can be expressed as:

$$F_{vi} = E_G\left(I_{vi}^Y\right), F_{ir} = E_V\left(I_{ir}\right) \tag{1}$$

where F_{vi} represents the extracted visible image feature information from the Y channel of the visible image and F_{ir} represents the extracted infrared image feature information from the infrared image.

The MGRM was used to extract edge features from the infrared image. With I_{ir} as the input infrared image, F_m^n as the single-layer output feature, and F_{ir} as the module output feature, the MGRM can be represented as:

$$F_m^n = \nabla\left(Conv_{n*n}\left(I_{ir}\right)\right), (n = 1, 3, 5) \tag{2}$$

$$F_{ir} = C\left(F_m^1, F_m^3, F_m^5\right) \tag{3}$$

The symbol ∇ represents the Sobel gradient operator, $Conv_{n*n}$ denotes the convolution operation, where n × n denotes the size of the convolutional kernel.

Function $C(\cdot)$ represents the concatenation along the channel dimension. The MGRM obtains rich edge texture information through multiscale convolution and gradient computations.

The input feature of the GOM was obtained by passing the visible Y-channel image through a convolutional layer and can be represented as:

$$F_{in\_vi} = Conv\left(I_{vi}^Y\right) \tag{4}$$

Taking F_{in_vi} as the input feature of GOM and F_{vi} as the output feature, the formula of GOM can be expressed as:

$$F_{vi} = Conv^n\left(F_{in\_vi}\right) \oplus Conv\left(\nabla F_{in\_vi}\right) \tag{5}$$

where $Conv(\cdot)$ represents a convolutional layer, $Conv^n(\cdot)$ represents n cascaded convolutional layers, and $\oplus$ denotes element-wise addition in features. Finally, image reconstruction was used to reconstruct the fused image, denoted as I_f, where F_f represents the fusion features. $R(\cdot)$ represents image reconstruction. This can be represented as:

$$I_f = R\left(F_f\right) \tag{6}$$

3.2 Network Architecture

The proposed network architecture for the EdgeFusion is shown in Fig. 2. In this architecture, two GOM modules are used to extract features from the visible image, and an MGRM is used to extract edge features from the infrared image. The visible image is first converted from the RGB colour space to the YCrCb colour space. The Y channel is then extracted and passed through a 3×3 convolutional layer using the activation function of LReLU. Subsequently, two GOM modules are employed to extract the texture and detailed features from the visible image.

The MGRM was used to extract edge features from an infrared image, and its detailed structure is shown in Fig. 3(a). This module first performs convolution operations using kernels of different scales to extract features from an infrared image. Specifically, 1×1, 3×3, and 5×5 convolutional kernels were applied to the infrared image, and the LReLU activation function was used. After multiscale convolution, Sobel gradient operations were performed on the infrared features to extract edge features at different scales. The obtained multiscale features were then concatenated along the channel dimensions. Subsequently, two 3×3 convolutional layers were applied to extract deep-level features. Finally, a 1×1 convolutional layer was used to eliminate the difference in channel dimensions.

The GOM was used to extract texture details from the visible image, as shown in the Fig. 3(b). It comprises two 3×3 LReLU convolutional layers and a 1×1 shared convolutional layer, utilising dense connections to fully leverage the features extracted by each convolutional layer. The residual flow undergoes gradient computation, followed by a 1×1 convolution and addition to the main-stream flow, achieving the fusion of deep and fine-grained features.
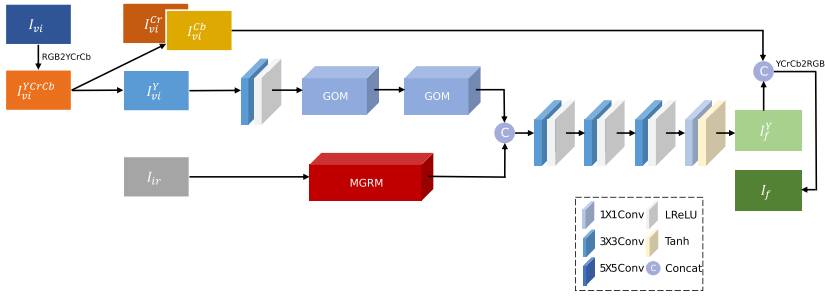
Fig. 2. EdgeFusion network structure diagram.

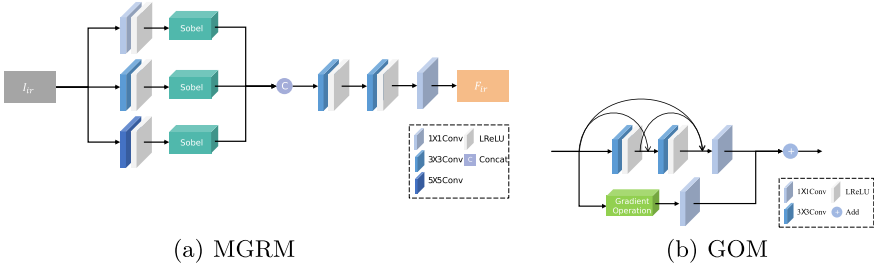(a) MGRM   (b) GOM

Fig. 3. Specific structure of MGRM and GOM modules.

Channel-wise concatenation was performed to combine the visible and infrared features, and the resulting feature maps were fed into the image reconstructor for image reconstruction. The decoder consists of three 3×3 convolutional layers and one 1×1 convolutional layer. The 3×3 convolutional layer used LReLU as the activation function, whereas the 1×1 convolutional layer used tanh as the activation function. No downsampling was introduced into the fusion network, and the feature maps during the fusion process were maintained consistent with the size of the source images, thus avoiding information loss.

3.3 Loss Function

To incorporate more essential and useful information into the fused image, detail loss was introduced in this study. The detail loss comprises three components: intensity loss $\mathcal{L}_{int}$ [22], texture loss $\mathcal{L}_{texture}$ [22], and balance loss $\mathcal{L}_{balance}$. The detailed loss is defined as follows:

$$\mathcal{L}_{detail} = \mathcal{L}_{int} + \alpha \mathcal{L}_{texture} + \beta \mathcal{L}_{balance} \tag{7}$$

where $\mathcal{L}_{int}$ is used to constrain the overall intensity performance of the fused image, $\mathcal{L}_{texture}$ is used to encourage the fused image to retain more intricate texture details, and $\mathcal{L}_{balance}$ is used to enforce a balanced proportion between the intensity and texture losses. The parameters α and β are used to adjust the

weights of the texture loss and the balance loss, respectively. The intensity loss is defined as follows:

$$\mathcal{L}_{int} = \frac{1}{HW} \left\| I_f - max\left(I_{vi}, I_{ir}\right) \right\|_1 \tag{8}$$

In the formula, H and W represent the height and width of the image, $\left\|\cdot\right\|_1$ denotes the l_1 norm, and $max(\cdot)$ represents the selection of the maximum value among the elements. Here, it represents the fusion of salient pixels from the visible and infrared images. The texture loss is defined as follows:

$$\mathcal{L}_{texture} = \frac{1}{HW} \left\| |\nabla I_f| - max\left(|\nabla I_{vi}|, |\nabla I_{ir}|\right) \right\|_1 \tag{9}$$

The symbol ∇ represents the Sobel gradient operation, and $|\cdot|$ denotes the absolute value operation. This formulation indicates that the texture of the fused image tends to be the maximum union of textures from the visible and infrared images. The balance loss is defined as follows:

$$\mathcal{L}_{balance} = \left(1 - \left(\mathcal{L}_{int}/\gamma\mathcal{L}_{texture}\right)\right)^2 \tag{10}$$

The balance loss is defined as follows: The parameter γ is used to constrain the proportion between the intensity loss and texture loss. The intensity loss significantly affects the fusion result, and the proportion between it and the texture loss affects the visual performance of the fusion result. The balance loss effectively constrains the overall intensity of the fused image, allowing the coexistence of texture and edge information.

4   Experimental Validation

4.1   Experimental Configurations

This paper details the extensive experiments conducted on the LLVIP [23] and TNO [24] datasets to comprehensively evaluate the proposed method. A large number of images in the LLVIP dataset were taken on night roads and included 12,025 training pairs and 3,463 test pairs. The proposed method is trained using the LLVIP dataset. The TNO dataset contains grayscale versions of multispectral nighttime images of various military-related scenes. EdgeFusion was evaluated in comparison with four state-of-the-art deep learning methods, namely Fusion-GAN, GANMcC, U2Fusion [21], and SeAFusion. The implementations of the four compared methods were configured according to publicly available codes and the original paper parameters.

The training parameters were set as follows: batch size of 8 for the Adam optimiser, initial learning rate of 10^{-3}, learning rate updated by multiplying the initial learning rate by $power^{(iter-1)}$, where the initial power was set to 0.75, and iter represents the current training iteration, with a total of 10 iterations. All the experiments were conducted on an NVIDIA Station server with an Intel Xeon E5 2620 v4 processor, 128 GB of memory, and four Tesla V100 GPUs, each with 32 GB of VRAM.
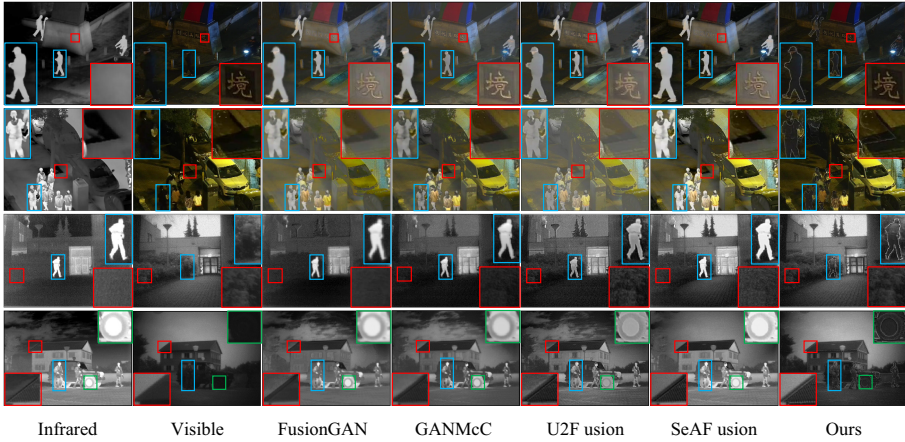
| Infrared | Visible | FusionGAN | GANMcC | U2F usion | SeAF usion | Ours |

Fig. 4. Qualitative comparison of the EdgeFusion method in this paper on the LLVIP dataset (first and second rows) versus the TNO dataset (third and fourth rows).

4.2   Fusion Metrics

The selected metrics include peak signal-to-noise ratio (PSNR), average gradient (AG), $Q^{AB/F}$, standard deviation (SD), and mutual information (MI). PSNR reflects the distortion during the fusion process; AG quantifies the gradient information of the fused image; $Q^{AB/F}$ measures the amount of edge information from source to fused image; SD reflects the distribution and contrast of the fused image; MI measures the amount of information transmitted from source to fused image. Fusion algorithms with higher PSNR, AG, $Q^{AB/F}$, SD, and MI values tend to exhibit better fusion performance.

4.3   Comparative Experiment

Qualitative Results. Two pairs of images from the LLVIP and TNO datasets were selected for fusion. Images #010011 and #190071 were chosen for the LLVIP dataset, whereas images Kapetein-1123 and Soldiers_with_jeep were selected for the TNO dataset. The results are shown in Fig. 4. The information provided by visible and infrared images in low light level is limited and the texture is weakened. However, except for our proposed method, other methods introduce global information from infrared images, which inevitably leads to coverage of the target and background details in the fused images. Examining the details within the colored box, it becomes apparent that the image generated by the edge fusion method exhibits favorable visual results. Targets and backgrounds can be distinctly differentiated, and the image overall retains a wealth of texture details, with all edges appearing sharp and clear.
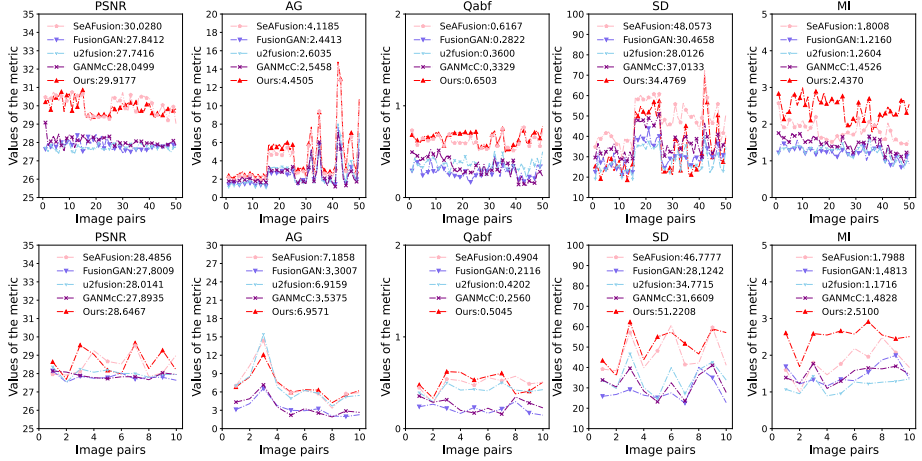
Fig. 5. A quantitative comparison of five metrics, namely PSNR, AG, Qabf, SD, and MI, was conducted on 50 images from the LLVIP dataset (row one) and 10 images from the TNO dataset (row two). The horizontal axis represents the image pairs, while the vertical axis represents the values of each fusion metric.

Quantitative Results. In this study, we conducted a quantitative comparison of five evaluation metrics on 50 image pairs from four different scenarios in the LLVIP dataset. The first row of Fig. 5 shows the results of all the methods. The method proposed in this study demonstrates significant advantages in three metrics: AG, Qabf, and MI. AG indicates that the fusion results of this method contain rich details and textures. Qabf suggests that the fusion images generated using this method contain considerable edge information. MI indicates that the method preserves most information from the source images. This is attributed to the strong edge preservation capability of the multiscale gradient preservation module. Because this method primarily selects the edge information from infrared images for fusion, it avoids the issue of completely highlighting the objects in the fused image. Therefore, it is understandable that the SD metric ranks third. In this study, we also conducted a quantitative evaluation of five metrics on a subset of 10 image pairs from the TNO dataset. The results are shown in the second row of Fig. 5. The various experimental results on different datasets indicate that EdgeFusion has a significant advantage in preserving infrared edges and enhancing texture details.

4.4   Detection Performance

This paper uses the LLVIP dataset for training with Yolov5, which has its own pedestrian detection label. The detection results for the LLVIP dataset are listed in Table 1. The average precision (mAP) metric was used to assess the detection performance, where a higher mAP value closer to 1 indicated a better pedestrian detection performance. The mAP_{iou} scores at different IoU thresholds. Additionally, mAP@[0.5:0.95] represents the average mAP across thresholds ranging from

Table 1. The pedestrian detection results on the LLVIP dataset for infrared, visible, and fused images are as follows (bold indicates the best, underline indicates the second best).

	mAP@0.5	mAP@0.6	mAP@0.7	mAP@0.8	mAP@0.9	mAP@[0.5:0.95]
Infrared	0.649	0.578	0.453	0.279	0.063	0.367
Visible	0.775	0.700	0.545	0.308	0.028	0.429
FusionGAN	0.756	0.698	0.606	0.321	0.064	0.449
GANMcC	0.786	0.717	**0.617**	0.370	0.046	0.467
U2Fusion	0.784	0.743	0.609	0.433	0.046	0.485
SeAFusion	0.769	0.729	0.597	0.379	**0.095**	0.475
Ours	**0.817**	**0.746**	0.586	**0.455**	0.057	**0.487**

Table 2. Quantitative evaluation results of the ablation study on 50 pairs of infrared and visible images from the LLVIP dataset (bold indicates the best, underline indicates the second best).

	without $\mathcal{L}_{balance}$	without MGRM	Ours
PSNR	**28.6715**	28.3948	28.6467
AG	6.7383	6.6774	**6.9571**
Qabf	**0.5167**	0.4443	0.5045
SD	44.9629	49.9748	**51.2208**
MI	1.8325	1.9944	**2.5100**

0.5 to 0.95 with a step size of 0.05. From the comprehensive point of view in the table, the method proposed in this paper has certain advantages in mAP, indicating that the method has a promoting effect on downstream tasks.

4.5   Ablation Studies

Important contributions in this paper are MGRM and balance loss. We conducted a series of ablation experiments on LLVIP and TNO datasets, and the
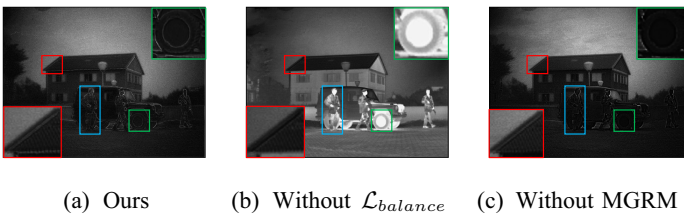

(a) Ours   (b) Without $\mathcal{L}_{balance}$ (c) Without MGRM

Fig. 6. Visualization results of ablation experiments.

results are shown in Table 2 and Fig. 6. As can be seen from Fig. 6(c), balance loss significantly inhibits the intensity of infrared information. Without MGRM, the overall image becomes darker and edge information is missing. In Fig. 6(b), the absence of balance loss makes the fused image indistishable from the infrared image. In contrast, only EdgeFusion's fused images have both the bright edges of the thermal target and the enhanced texture details.

5   Conclusion

In this study, EdgeFusion was proposed for infrared and visible images. The network innovatively utilises infrared gradient information to annotate thermal targets, effectively avoiding the problem of infrared targets becoming completely highlighted. This was achieved through the gradient preservation capability of the infrared MGRM. Additionally, a balanced loss function was designed to introduce infrared image information while avoiding excessive interference from infrared information. This allowed the fused image to retain additional textural details. Experimental results on a public dataset demonstrate that this approach is beneficial for enhancing the visibility of textures in visible images, and the fused image retains clear and rich texture details. Quantitative experiments comparing four state-of-the-art methods using five evaluation metrics further confirm that the proposed approach preserves more gradients and textures, resulting in fused images with the highest information content. The results of the object detection experiments further validated the superior performance of EdgeFusion in advanced visual tasks, such as pedestrian detection.

References

1. Dogra, A., Goyal, B., Agrawal, S.: From multi-scale decomposition to non-multi-scale decomposition methods: a comprehensive survey of image fusion techniques and its applications. IEEE Access **5**, 16040–16067 (2017)
2. Ma, Y., Chen, J., Chen, C., Fan, F., Ma, J.: Infrared and visible image fusion using total variation model. Neurocomputing **202**, 12–19 (2016)
3. Das, S., Zhang, Y.: Color night vision for navigation and surveillance. Transp. Res. Rec. **1708**(1), 40–46 (2000)
4. Li, H., Wu, X.J., Durrani, T.: NestFuse: an infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. IEEE Trans. Instrum. Meas. **69**(12), 9645–9656 (2020)
5. Ha, Q., Watanabe, K., Karasawa, T., Ushiku, Y., Harada, T.: MFNet: towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes. In: 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 5108–5115. IEEE (2017)
6. Zhou, Z., Wang, B., Li, S., Dong, M.: Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters. Inf. Fusion **30**, 15–26 (2016)
7. Li, H., Qi, X., Xie, W.: Fast infrared and visible image fusion with structural decomposition. Knowl.-Based Syst. **204**, 106182 (2020)

8. Ma, J., Zhou, Y.: Infrared and visible image fusion via gradientlet filter. Comput. Vis. Image Underst. **197**, 103016 (2020)
9. Long, Y., Jia, H., Zhong, Y., Jiang, Y., Jia, Y.: RXDNFuse: a aggregated residual dense network for infrared and visible image fusion. Inf. Fusion **69**, 128–141 (2021)
10. Li, H., Wu, X.j., Durrani, T.S.: Infrared and visible image fusion with resnet and zero-phase component analysis. Infrared Phys. Technol. **102**, 103039 (2019)
11. Li, H., Wu, X.J., Kittler, J.: RFN-Nest: an end-to-end residual fusion network for infrared and visible images. Inf. Fusion **73**, 72–86 (2021)
12. Xu, H., Wang, X., Ma, J.: DRF: disentangled representation for visible and infrared image fusion. IEEE Trans. Instrum. Meas. **70**, 1–13 (2021)
13. Ram Prabhakar, K., Sai Srikar, V., Venkatesh Babu, R.: DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 4714–4722 (2017)
14. Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., Zhang, L.: IFCNN: a general image fusion framework based on convolutional neural network. Inf. Fusion **54**, 99–118 (2020)
15. Ma, J., Tang, L., Xu, M., Zhang, H., Xiao, G.: STDFusionNet: an infrared and visible image fusion network based on salient target detection. IEEE Trans. Instrum. Meas. **70**, 1–13 (2021)
16. Ma, J., Yu, W., Liang, P., Li, C., Jiang, J.: FusionGAN: a generative adversarial network for infrared and visible image fusion. Inf. Fusion **48**, 11–26 (2019)
17. Ma, J., Xu, H., Jiang, J., Mei, X., Zhang, X.P.: DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Trans. Image Process. **29**, 4980–4995 (2020)
18. Ma, J., Zhang, H., Shao, Z., Liang, P., Xu, H.: GANMcC: a generative adversarial network with multiclassification constraints for infrared and visible image fusion. IEEE Trans. Instrum. Meas. **70**, 1–14 (2020)
19. Liu, D., Wen, B., Liu, X., Wang, Z., Huang, T.S.: When image denoising meets high-level vision tasks: a deep learning approach. arXiv preprint arXiv:1706.04284 (2017)
20. Ma, J., Ma, Y., Li, C.: Infrared and visible image fusion methods and applications: a survey. Inf. Fusion **45**, 153–178 (2019)
21. Xu, H., Ma, J., Jiang, J., Guo, X., Ling, H.: U2Fusion: a unified unsupervised image fusion network. IEEE Trans. Pattern Anal. Mach. Intell. **44**(1), 502–518 (2020)
22. Tang, L., Yuan, J., Ma, J.: Image fusion in the loop of high-level vision tasks: a semantic-aware real-time infrared and visible image fusion network. Inf. Fusion **82**, 28–42 (2022)
23. Jia, X., Zhu, C., Li, M., Tang, W., Zhou, W.: LLVIP: a visible-infrared paired dataset for low-light vision. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3496–3504 (2021)
24. Toet, A.: TNO image fusion dataset (2021). https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029