ORIGINAL ARTICLE

Adaptive low light visual enhancement and high-significant target detection for infrared and visible image fusion

Wenxia Yin[1] · Kangjian He[1] · Dan Xu[1] · Yingying Yue[1] · Yueying Luo[1]

Abstract

Infrared and visible image fusion aim to obtain a fused image with salient targets and preserve abundant texture detail information as much as possible, which can potentially improve the reliability of some target detection and tracking tasks. However, some visible images taken from low-illumination conditions are subjected to losing many details and cannot obtain a good fusion result. To address this issue, we proposed a novel adaptive visual enhancement and high-significant targets detection-based fusion scheme in this paper. First, a bright-pass bilateral filter and adaptive-gamma correction-based algorithm are proposed to enhance the visible image adaptively. Second, an iterative guided and infrared patch-tensor-based algorithm are proposed to extract the infrared target. Third, an efficient hybrid $\ell_1 - \ell_0$ model decomposes the infrared and visible image into base and detail layers and then fuses them by weight map strategy. The final fused image is obtained by merging the fused base layers, detail layers, and infrared targets. Qualitative and quantitative experimental results demonstrate that the proposed method is superior to 9 state-of-the-art image fusion methods as more valuable texture details and significant infrared targets are preserved. Supplemental material and codes of this work are publicly available at: https://github.com/VCMHE/BI-Fusion.

Keywords Image fusion · Adaptive low light visual enhancement · High-significant target detection · Hybrid decomposition

1 Introduction

As an information-enhanced technique, image fusion plays a vital role in computer vision [1, 2]. Specifically, image fusion aims to extract and integrate feature information from different source images to generate a single fused image, providing more comprehensive information for the human visual system and applications [3, 4]. Based on different types of source images, image fusion can be mainly divided into five categories [5]: multi-focus image fusion [6], infrared and visible image fusion [7], multi-exposure image fusion [8], multi-modal medical image fusion [9], and remote sensing image fusion [10]. Among these, infrared and visible image fusion is one of the essential branches since the high complementarity between source images [11].

Generally speaking, visible images contain high resolution and abundant texture details, but they are susceptible to the environment, such as low illumination environment and foggy atmosphere, resulting in the ability of depicts target drops significantly [12] (Fig. 1a). Compared with visible images, infrared images are advantageous in representing salient targets because of the thermal radiation [13]. However, infrared images have shortcomings such as poor texture and low resolution (Fig. 1b). Therefore, the fusion of infrared and visible images can obtain more comprehensive scene information, as shown in Fig. 1c. The fused image contains both significant infrared targets and preserves abundant visible detail information, potentially improving the recognizability of targets under various environments [14].

Infrared and visible image fusion methods can be divided into pixel-level, feature-level, and decision-level-based methods [15]. Some traditional pixel-level-based methods can fuse infrared and visible images, but minimum artifacts still exist [16]. Currently, with the development of multiscale theory, many corresponding algorithms have been proposed, such as TE-MST [17], discrete wavelet transform (DWT) [18], MGFF [19], and ADF [20]. Meanwhile, some

✉ Kangjian He
  hekj@ynu.edu.cn

✉ Dan Xu
  danxu@ynu.edu.cn

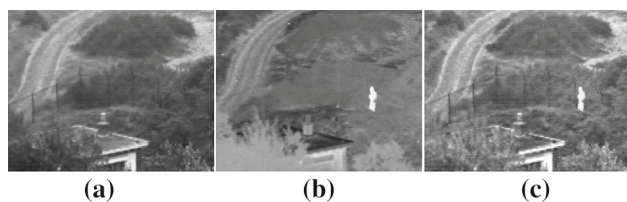[1] School of Information Science and Engineering, Yunnan University, Kunming 650091, China

Fig. 1 **a** Visible image; **b** infrared image; **c** fusion result of our method

sub-space-based methods have also been proposed, such as FPDE [21], PCA [22], and ICA [23].

Some feature-based methods are also applied in infrared and visible fusion tasks. Li et al. proposed a novel fusion strategy (IVFusion) [24], which obtained the feature weight map by calculating the structural similarity between the pre-fusion image and the source images. In our previous work, we extracted salient features in infrared images through a significant target analysis method to highlight the saliency of fused targets [25]. Other feature exact filters such as CBF [26] and GFF [27], are also applied in infrared and visible image fusion tasks.

The decision-level-based method can facilitate the process of designing fusion rules. Zhao et al. proposed an infrared and visible image fusion method based on the Bayesian model [28], which can automatically obtain the adaptive fusion weight. Panigrahy et al. [29]. proposed a parameter adaptive unit-link dual-channel PCNN model to automatically measure the activity level and demonstrated the proposed model is more effective than the corresponding manual settings. Meanwhile, with the development of deep learning theory, many fusion methods based on neural networks have been proposed in recent years. Li et al. [30] proposed a novel end-to-end residual fusion network (named RFN-Nest) in which the detail-preserving and feature-enhancing loss functions are carefully designed to preserve the detail of the fused image. Ma et al. [31] proposed FusionGAN, which first applied GAN to infrared and visible image fusion tasks, avoiding the difficulty of manually designing fusion rules. Li et al. [32]. proposed a novel coupled GAN structure to conduct the fusion task and they use the pre-fused image as the label to guide the training process. Tan et al. [33]. designed a flexible spatiotemporal fusion model based on conditional GAN, which can alleviate the time dependence between references and prediction images. Based on GAN and dense blocks, Fu et al. [34] proposed a perceptual Fusion-GAN network. By concatenating visible images at each layer and designing structural similarity and gradient loss functions, the fused image is consistent with perception. These deep learning-based methods can also automatically extract deep features. However, challenges still exist, such as lacking training data and ground truth.

Infrared and visible image fusion aim to obtain a fused image with salient targets and preserve rich texture detail

information. However, the quality of infrared and visible images collected by sensors is usually different, especially those visible images taken from low-illumination conditions are subjected to losing many details and cannot obtain a good fusion result. Moreover, some current infrared and visible image fusion algorithms mainly focus on the decomposition mode and fusion rules, such as MGFF [19], IVFusion [24], and our previous work [25], so the robustness of the low-illumination visible images needs to be improved.

Based on the above analysis, a novel adaptive visual enhancement and high-significant target detection-based infrared and visible image fusion scheme is proposed. First, the illumination component is estimated by the fast bright-pass bilateral filtering, and then enhance the visible image by an adaptive gamma enhancement method. Second, the infrared patch-tensor model (IPT) is constructed to extract the coarse infrared targets. Then the accurate salient regions map is generated by the iterative optimization method based on the guided filter. Third, decomposing the infrared and visible images into base and detail layers by the efficient hybrid $\ell_1 - \ell_0$ model, the base layers are fused by the visual saliency map strategy, and the detail layers are fused by the visual difference feature map strategy. The final fused image is obtained by merging the fused base layer, detail layers, and infrared targets.

The main contributions of our work can be summarized as follows:

(1) A visible image enhancement algorithm (named BFAC) is proposed. BFAC is based on bright-pass bilateral filter and adaptive-gamma correction. Unlike some previous fusion algorithms that directly fuse infrared and visible images, BFAC is applied to preprocess the visible image before fusion. Qualitative and quantitative experimental results demonstrate that BFAC can effectively enhance the hidden details in visible images while avoiding the overexposure problem, leading to the final fusion results containing rich scene information and consistent with the human visual system.

(2) An infrared targets extraction algorithm (named IGPT) is proposed. The algorithm first obtains coarse saliency regions by calculating the tensor rank through the IPT model and then generates the accurate saliency region map through iterative guided filtering and convergence conditions. Compared with previous saliency detection networks, this algorithm can extract infrared targets accurately, thereby enhancing the saliency of infrared targets in the final fused image.

(3) Based on BFAC and IGPT, we propose a novel infrared and visible image fusion scheme based on the hybrid $\ell_1 - \ell_0$ decomposition model. The qualitative and quantitative analysis demonstrates that the final fusion results are superior to some state-of-the-art methods with richer

texture details and salient infrared targets, which can potentially improve the recognizability of the target under complex environments.

The rest of this paper is organized as follows. In Sect. 2, corresponding preliminaries and methods are discussed. Section 3 presents the proposed method in detail. The experimental results and discussions are provided in Sect. 4. Section 5 draws a conclusion of this work.

2 Preliminaries

This section briefly introduces the preliminary theories and methods, including the Retinx model and Infrared-Patch-Tensor model (IPT).

2.1 Retinx model

The Retinex model is widely applied in image enhancement tasks [35, 36], as it is consistent with the human visual system. The idea of Retinex-based methods can be generally summarized as three steps. First, decomposing the observed image into the illumination component and reflectance component. Second, enhancing the illumination component. Then, the final enhanced image can be obtained by combing the enhanced illumination and reflectance component.

Bright-pass bilateral filtering (BPBF) [37] belongs to the Retinex theory and can be used to estimate the illumination component L_s of a source image VI_s. Which can be mathematically expressed as:

$$L_s(i) = \frac{1}{\eta_i} \sum_{j \in \Omega} \omega(j) \Phi(VI_s(i-j) - VI_s(i)) VI_s(i-j),$$

$$(1)$$

where

$$\eta_i = \sum_{j \in \Omega} \omega(j) \Phi(VI_s(i-j) - VI_s(i)). \tag{2}$$

Ω is the neighborhood of the performed aggregation.

The spatial kernel ω is:

$$\omega(i) = \exp\left(-\frac{||i||^2}{2\theta^2}\right), \tag{3}$$

The range kernel Φ in (1) is:

$$\Phi(t) = \begin{cases} \exp\left(-\frac{t^2}{2\sigma^2}\right) & \text{if } t \geq 0, \\ 0 & \text{otherwise,} \end{cases} \tag{4}$$

where θ and σ are standard deviations.

Based on BPBF, a Fourier approximation algorithm called FBPBF is proposed in [38], which replaces the range kernel Φ with a K-term Fourier approximation. As a result, the FBPBF can effectively estimate the illumination component with a low computing cost.

2.2 Infrared-patch-tensor model

The infrared patch-tensor (IPT) model converted the detection of infrared targets as a matrix decomposition and optimization problem, which assumed the background is low-rank and the salient regions are sparse. Accurately approximating the background tensor rank is a key in the IPT model.

In [39], kong et al. proposed a novel IPT model base on nonconvex tensor fibered rank approximation to detect the salient regions of infrared images. The proposed model can be mathematically defined as:

$$\min_{\mathcal{B}, \mathcal{T}} ||\mathcal{B}||_{LogTFNN} + \lambda ||\mathcal{T} \odot \mathcal{W}||_1 + \beta ||\mathcal{B}||_{HTV},$$

$$\text{s.t.} ||\mathcal{I} - \mathcal{B} - \mathcal{T}||_F \leq \delta \tag{5}$$

where $\mathcal{I}$ is the infrared patch tensor, $\mathcal{B}$ is the background patch tensor, and $\mathcal{T}$ is the target patch tensor. In (5), the infrared patch images are converted to tensors to improve the data processing efficiency. Then, the tensor fibered nuclear norm based on Log operation ($LogTFNN$) is used to approximate the tensor fibered rank. In order to suppress the sparse interference of corners and edges, the prior information is added by the local structure tensor $\mathcal{W}$, $\odot$ is the Hadamard product. Factor λ is used to balance the sparse and low-rank components. Besides, a novel hypertotal variation regularization term (HTV) [40] is also applied to constrain background noise and β is the corresponding regularization factor. The model in (5) can be further expanded as:

$$\min_{\mathcal{B}, \mathcal{T}} \sum_{k=1}^{3} \alpha_k \mathcal{H}_k(\mathcal{X}_k) + \lambda ||\mathcal{T} \odot \mathcal{W}||_1 + \beta ||\mathcal{Z}||_{HTV}$$

$$+ \frac{1}{2\mu} ||\mathcal{I} - \mathcal{B} - \mathcal{T}||_F^2,$$

$$\text{s.t.} \begin{cases} \mathcal{B} = \mathcal{X}_k, \ k = 1, 2, 3 \\ \mathcal{Z} = \mathcal{B} \end{cases} \tag{6}$$

where mode—mode-kt-SVD[41] is used to calculate the tensor rank, the variable k ranges from 1 to 3, α_k is a weighting factor and satisfies the condition $\sum_{k=1}^{3} \alpha_k = 1$, $\mathcal{H}_k(\mathcal{X}_k) = ||\mathcal{B}||_{\log-\text{mode}-k-\text{TFNN}}$, $\mathcal{Z}$ is the auxiliary variable, and μ is the penalty factor.

In order to solve this optimization problem, the alternating direction method of multipliers (ADMM) [42] framework
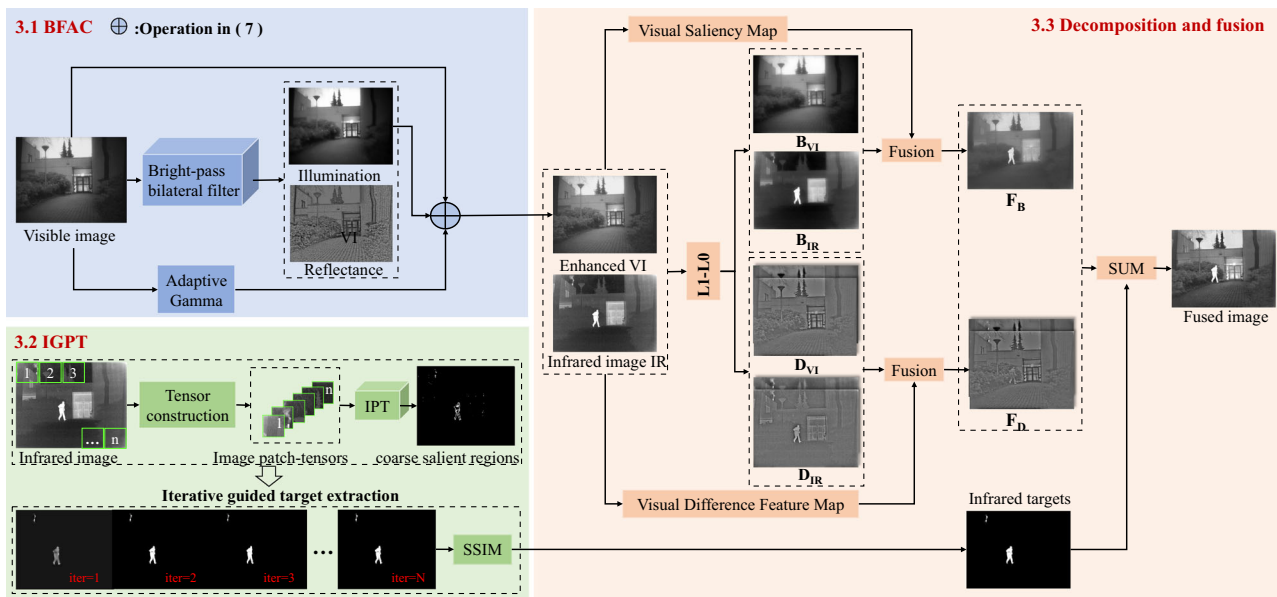
Fig. 2 The framework of the proposed method

is applied, which converts problem (6) to solving five sub-problems iteratively, then the coarse background and salient targets can be obtained by reconstructing $\mathcal{B}$ and $\mathcal{T}$.

3 Methodology

The framework of the proposed method is shown in Fig. 2 which consists of three sub-modules: visible image adaptive enhancement module, infrared target extraction module, decomposition, and fusion module.

First, the illumination component is estimated by the fast bright-pass bilateral filtering, and then we correct the brightness of the visible image by calculating the adaptive gamma value so as to obtain the enhanced visible image with an excellent visual effect. Second, the infrared patch-tensor model is constructed to extract the coarse infrared targets; then, the accurate target map is generated through iterative guided filtering and convergence conditions. Third, decomposing the infrared and visible images into base and detail layers by the efficient hybrid $\ell_1-\ell_0$ model, the base layers are fused by the visual saliency map, and the detail layers are fused by the visual difference feature map. The final fused image is obtained by merging the fused base layer, detail layers, and infrared targets.

3.1 Bright-pass bilateral filter and adaptive-gamma correction-based image enhancement (BFAC)

As mentioned in the preliminaries section, given an observed image VI_s, the FBPBF algorithm [38] can estimate the illumination component L_s by a Fourier approximation of

bright-pass bilateral filtering. Then multiply the gamma correction with the illumination component to obtain the enhanced image VI. This process can be presented in (7).

$$VI = \frac{VI_s}{L_s} \times 255 \left(\frac{L_s}{255} \right)^{\mathrm{T}},$$

(7)

where $\mathrm{T} = 1/\gamma$, in the above-mentioned FBPBF algorithm, the γ is set to 2.2, enhancing the image uniformly. However, for the infrared and visible image fusion task, the characteristics of the collected visible images are diverse. It is impractical to enhance all the visible images with a constant coefficient, as shown in (b2), (b3), (b4), (b5), and (b6) in Fig. 3.

Overexposure problems arise in FBPBF. Therefore, we proposed an efficient algorithm to achieve adaptive enhancement of visible images.

In (7), the illumination component L_s is enhanced by the gamma correction function:

$$F_\gamma(x) = (x)^{\frac{1}{\gamma}},$$

(8)

where x is the normalized pixel value in L_s which ranges from 0 to 1, and the range of F_γ is (0, 1) too. Figure 4 (p1) shows the plot of the gamma correction function $F_\gamma(x)$ with the different values of γ, when γ is increased, the value of F_γ gradually increases, and the enhancement effect gradually strengthens. So, we can control the degree of luminance by adaptively adjusting γ.

Specifically, we define I_{Mean} as the normalized mean value of pixels. Given a low-illumination image with small I_{Mean}, γ should be larger to make a higher level of enhancement. Given a normal-illumination image with large value of
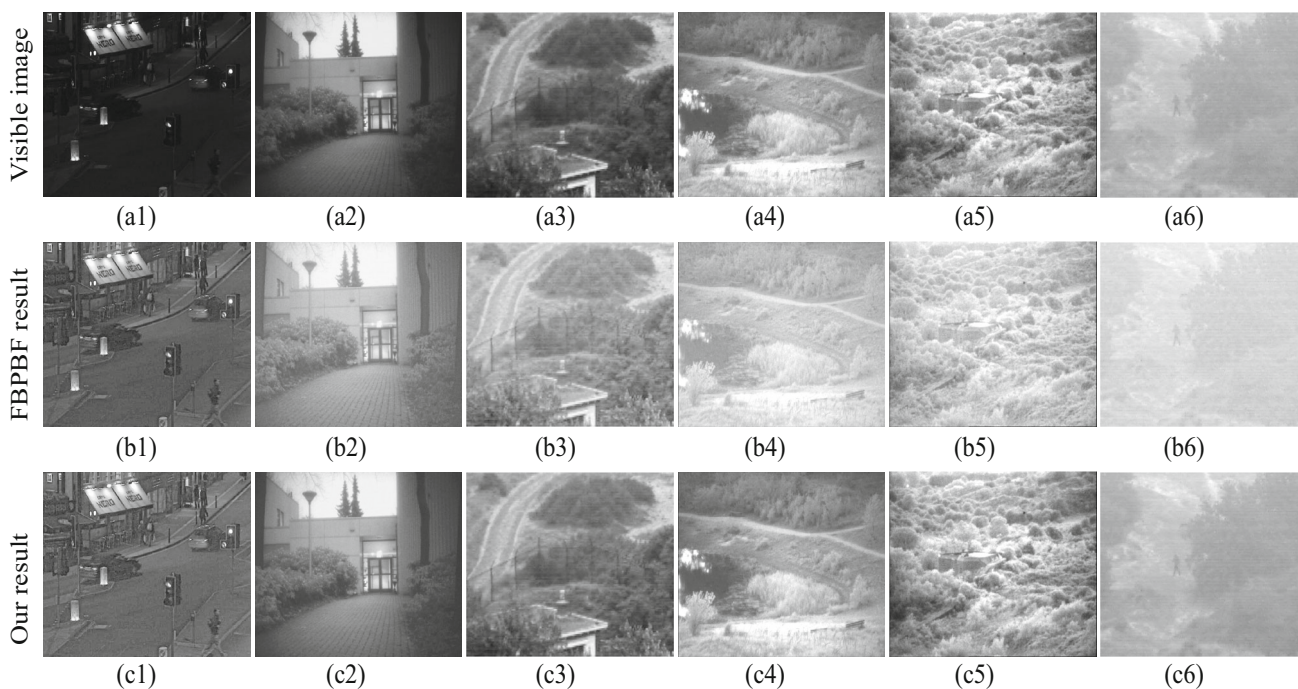
Fig. 3 The result of visible image enhancement. The first row is the original visible images; the second row is the enhanced results of FBPBF [38]; the third row is the enhanced results of the proposed method (BFAC)
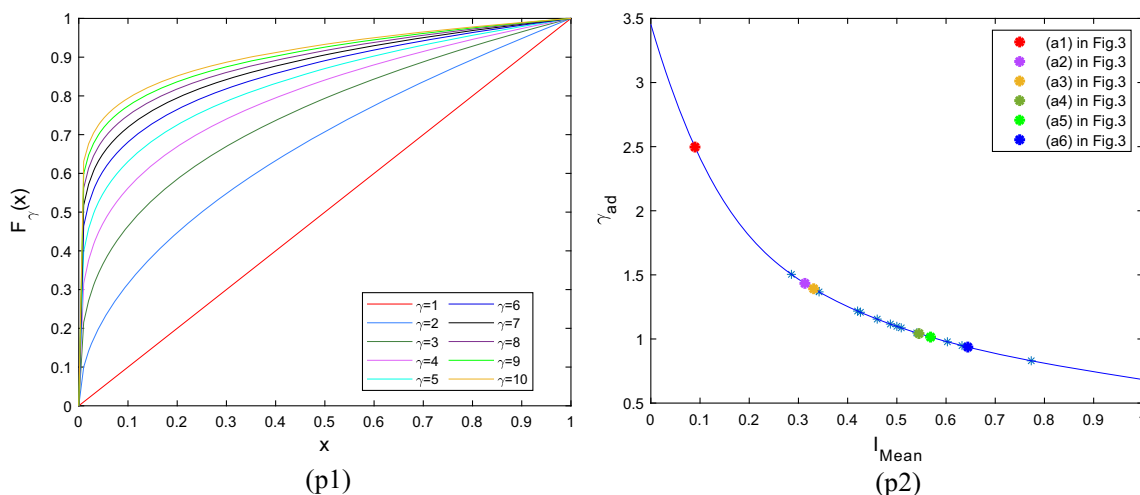

Fig. 4 (p1): The plot of $F_\gamma(x)$ with different γ. when x remains constant, the value of γ continues to increase, the value of $F_\lambda(x)$ increases too. (p2): The relationship between I_{Mean} and γ_{ad}. Where the highlighted dots represent the visible images (a1), (a2), (a3), (a4), (a5), (a6) in Fig. 3

I_{Mean}, γ should be smaller to avoid overexposure problems. On the other hand, according to Weber-Fechner law [43], the brightness perception of the human visual system is logarithmically related to illuminance. Therefore, we define an optimization function in (9) to obtain the adaptive γ_{ad}, Then obtain the adaptive enhanced image according to (7).

$$\gamma_{ad} = ln\left(1 + \frac{1}{I_{Mean}}\right). \quad (9)$$

In this paper, we select 21 pairs of visible images captured in different scenes to verify the effectiveness of the proposed algorithm. The relationship between I_{Mean} and γ_{ad} is illustrated in Fig. 4 (p2), when I_{Mean} increased, γ_{ad} gradually decreased, the enhancement effect gradually weakens, which means our algorithm can adaptively enhance visible images according to the actual situation.

Figure 3 shows the enhanced results by different methods. Figure 3 (a1)–(a6) are the visible images (I_{Mean} from low

to high), Fig. 3 (b1)–(b6), (c1)–(c6) illustrate the enhanced results of the FBPBF algorithm and the proposed method. We can observe that the FBPBF cannot enhance the visible images adaptively, resulting in overexposure (such as Fg.3 (b2)–(b6)). On the contrary, our method has two distinct advantages. First, for the low-illumination images (Fg.3 (a1)–(a3)), our method can enhance the detailed information and maintain the naturalness of the images. Second, for the normal-illumination visible image (Fig. 3 (a4)–(a6)), our results maintain consistent properties with the original visible images, avoiding the distortion of detail and contrast compared to the FBPBF algorithm (Fig. 4).

Figure 5 shows the quantitative comparison results on 21 pairs of visible images. Lightness order error (LOE) can measure the lightness distortion of enhanced results [44]. The smaller the value of LOE, the more natural the enhanced result is. VIFP is consistent with the human visual system and is used to measure the distortion of visual information [45]. The larger the value, the better visual information contains.

From the statistical results, our method can achieve the best average value on LOE and VIFP, which demonstrates our results can enhance the detailed information from low-illumination images while maintaining the naturalness of enhanced images, achieving the adaptive enhancement of visible images. The total calculation process can be expressed by Algorithm 1.

Algorithm 1: main steps of the proposed BFAC

Parameter descriptions:

VI_s is the original visible image

VI is the adaptively enhanced visible image.

Input: VI_s **Output:** VI

1: obtain the normalized illumination component L_s of VI_s by (1);

2: obtain the mean value I_{Mean} of VI_s ;

3: calculate the adaptive γ_{ad} by (10);

4: obtain the adaptively enhanced visible image VI by (8);

5: **return** VI

3.2 Iterative guided and infrared patch-tensor based target extraction (IGPT)

As highlighted by the rectangles in Fig. 6, infrared images can efficiently represent salient targets compared with visible images. Therefore, the preservation of infrared targets generally needs to be considered during the fusion process. Based on this motivation, an iterative guided and infrared patch-tensor based target extraction algorithm (IGPT) is proposed.

As discussed in the section of preliminaries, we first applied the infrared patch-tensor model (IPT) to coarsely extract infrared targets as shown in the fourth row of Fig. 6. It can be seen that, problems such as isolated pixels and discontinuous edges exist in the results of IPT model.

In order to address the above problems, a threshold strategy [46] is devised to remove those small and insignificant pixels. Specifically, when the area of the pixel region is smaller than the threshold, these pixels are reversed. The threshold is a user-defined value that can be determined according to different tasks, and we analyzed it in Sect. 4.6. In this paper, we set the threshold to $0.0002 \times H \times W$, where H and W are the height and weight of the infrared image, respectively. In addition, considering the excellent structure transferring property of the guided filter [47], we further proposed an iterative guided scheme to preserve as much edge details as possible of infrared targets. The total calculation process of IGPT can be summarized in Algorithm 2.

Algorithm 2: main steps of the proposed IGPT

Parameter descriptions:

IR is the infrared image.

ADMM is the alternating direction method of multipliers

T_C is the coarse extraction result.

$GF(A, B)$ is the guided filtering operation, where A is the guidance, B is the filtering input.

SSIM is the structural similarity index measure and can be mathematically defined in (16).

T_{IR} is the final infrared target map.

Input: IR **Output:** T_{IR}

1: Traverse the IR through a sliding window technique, then construct the patch tensor $\mathcal{T}$;

2: Iteratively solve the optimization problem in (6) by the ADMM framework;

3: Obtain the coarse infrared target map T_C ;

4: Remove the isolated pixels in T_C by the threshold strategy;

5: $T_{IR}^1 = GF(IR, T_C)$;

6: **For** iter=2 to N **Do**

7: **While** not converged **Do**

8: $T_{IR}^{iter+1} = GF(IR, T_{IR}^{iter})$;

9: Check the convergence conditions:

10: $SSIM(T^{iter}, T^{iter+1}) \to 1$

11: **End While**

12: **End For**

In the iterative guided scheme, the dynamic optimization model between the guidance IR and the infrared target map T_{IRi}^n in n-th iteration is defined as follows:

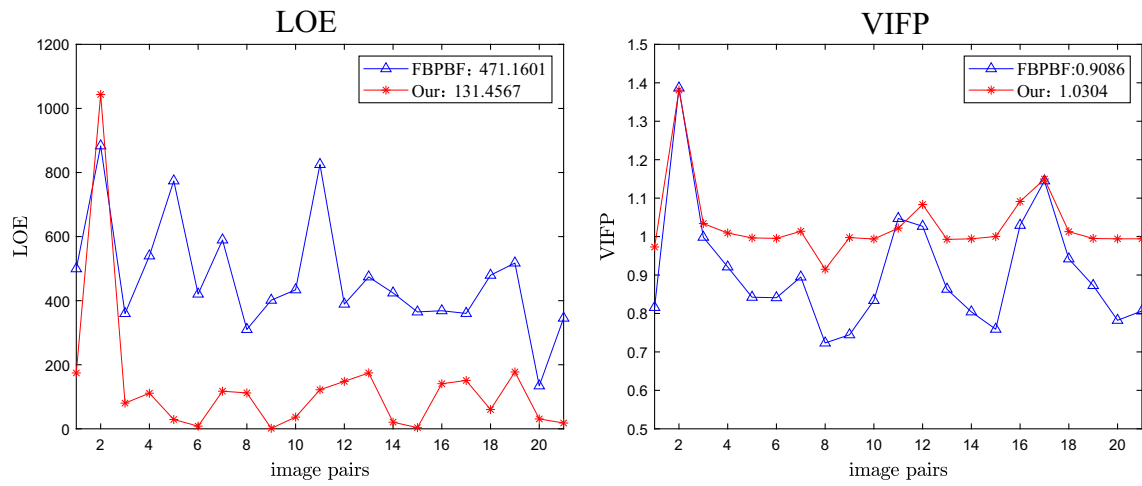$$T_{IRi}^n = a_k^n IR_i + b_k^n \ \forall i \in \omega_k, \tag{10}$$

Fig. 5 Quantitative comparison results on 21 pairs of visible images
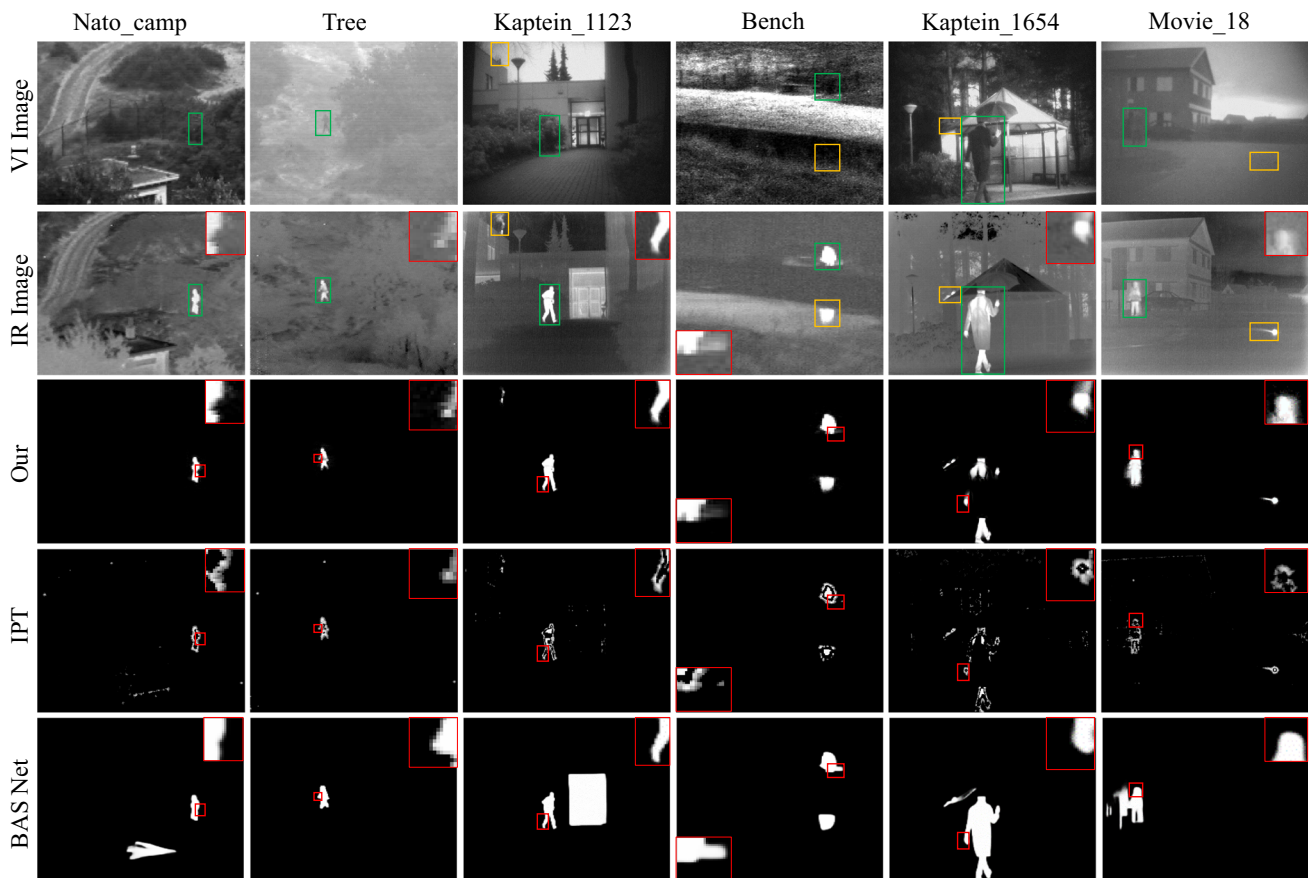

Fig. 6 Extracted infrared targets. The first row is the visible images, the second row is the infrared images, the third row is the results of IGPT, the fourth row is the result of IPT [39], and the last row is the result of BASNet [49]

W. Yin et al.

where i is the index of pixels, n is the number of iterations, ω_k is a square window with radius r, a_k^n and b_k^n are coefficients in n-th iteration which can be obtained by solving the following cost function:

$$E(a_k^n, b_k^n) = \sum_{i \in \omega_k} ((a_k^n IR_i + b_k^n - T_{IRi}^{n-1})^2 + \varepsilon a_k^n) \qquad (11)$$

where T_{IRi}^{n-1} is the filtering input, ε is a regularization parameter, and the solutions of (a_k^n, b_k^n) is presented as (12) and (13) which is based on the previous output T_{IR}^{n-1}.

$$a_k^n = \frac{\frac{1}{|\omega|} \sum_{i \in \omega_k} IR_i T_{IRi}^{n-1} - \mu_k \overline{T_{IRi}^{n-1}}}{\sigma_k^2 + \varepsilon}, \qquad (12)$$

$$b_k^n = \overline{T_{IRi}^{n-1}} - a_k^n \mu_k, \qquad (13)$$

where $|\omega|$ is the number of pixels in ω_k, $\overline{T_{IRi}^{n-1}}$ is the mean of T_{IRi}^{n-1} in ω_k, μ_k and σ_k^2 are the mean and variance of IR in ω_k.

By analyzing the above definition, it can be concluded that if a pixel is in a "High variance" area, it means this pixel belongs to the edge area, then we have $\sigma_k^2 >> \varepsilon$, so a_k^n will become far from 0. And the edge structure of the infrared target in IR will gradually transfer to the T_{IR}^n with increasing n. If a pixel is in the "Flat patch" area, it means pixels in this area are almost constant, then we have $\sigma_k^2 << \varepsilon$, so a_k^n is approximate to 0 (more details can be referred to [47]), then based on (10) and (13) the infrared target map T_{IR}^n can be obtained by:

$$T_{IRi}^n \approx \overline{T_{IRi}^{n-1}}, \ \ \mathrm{if} a_k^n \approx 0. \qquad (14)$$

According to (14), it can be further deduced that:

$$\left. \begin{array}{l} T_{IR}^n \approx \overline{T_{IR}^{n-1}} \\ T_{IR}^{n-1} \approx \overline{T_{IR}^{n-2}} \end{array} \right\} \Rightarrow T_{IR}^n \approx T_{IR}^{n-1}, \ \ \mathrm{if} \ n \text{is large enough.} \qquad (15)$$

which means the edge details of the target in IR image have been transferred to the infrared target map T_{IR}^n as much as possible, and the final infrared target map T_{IR} can be obtained.

As shown in Fig. 7 (a2)–(b4). As the number of iterations increases, Targets become more and more salient, and the edges also become sharper. In order to find the convergence condition, we define an evaluation function as (16).

$$\mathrm{SSIM}_{x,y} = \sum_{x,y} \frac{2\mu_x \mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1} \cdot \frac{2\sigma_x \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \cdot \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3}, \qquad (16)$$

where SSIM (structural similarity index measure)[48] is used to measure the structural similarity between images x and y, μ is the mean value, σ_x and σ_y are the standard deviations of x and y, σ_{xy} is the covariance, c_1, c_2 and c_3 are stable factors. A lager $SSIM$ indicates the more similar the two images are in terms of brightness, structure, and contrast information. The curve of $SSIM$ with different iterations is shown in Fig. 7. It can be seen that as iter increases, $SSIM$ increases too, and then begins to converge. We conduct lots of experiments with different infrared images and found that the curve tends to converge with 50 iterations.

Figure 6 shows the infrared target map obtained by different methods. Compare with IPT [39] and BASNet [49], our method has two distinct advantages. First, the proposed method can accurately obtain the infrared target. Such as in Kaptein_1123 (The third column in Fig. 6) the result obtained by our method contains both the chimney and the man. Second, as the iterative guided scheme is adopted, our method is superior in edge retention. As highlighted by the red rectangles, the results obtained by our method have a better visual effect as continuous and fine boundaries are preserved, which can increase the target saliency of the fused images effectively.

3.3 Decomposition and fusion

In this section, we will discuss how the infrared and visible images are decomposed and the process of obtaining the final fused image. For convenience, the adaptive enhanced visible image is denoted by VI, the infrared image is denoted by IR, the decomposed base layers of visible and infrared image are denoted by B_{VI} and B_{IR}, respectively, the detail layers of the visible and infrared images are denoted by D_{VI}^i and D_{IR}^i, respectively, where i is the scales ranges from 1 to 2 in this paper. The whole process of decomposition and fusion can be summarized as the following four steps:

Step1: Decompose the VI and IR into base layers and detail layers $(B_{VI}, B_{IR}, D_{VI}^1, D_{VI}^2, D_{IR}^1, D_{IR}^2)$ by the hybrid $\ell_1-\ell_0$ decomposition model.

The hybrid $\ell_1-\ell_0$ decomposition model can be defined as:

$$\begin{cases} b^1 = \mathrm{model}_{\ell_1-\ell_0}(s) \\ d^1 = s - b^1 \\ b^2 = \mathrm{model}_{\ell_1}(b^1) \\ d^2 = b^1 - b^2 \end{cases}, \qquad (17)$$

This model can decompose the image into one base layer b^2 and two detail layers d^1 and d^2, s represents the source image, $\mathrm{model}_{\ell_1-\ell_0}(\cdot)$ represents the optimization in (18).

$$\min_b \sum_{i=1}^N \{(s_i - b_i)^2 + \lambda_1 \sum_{j=\{x,y\}} |\partial_j b_i| + \lambda_2 \sum_{j=\{x,y\}} F(|\partial_j d_i|)\}, \qquad (18)$$
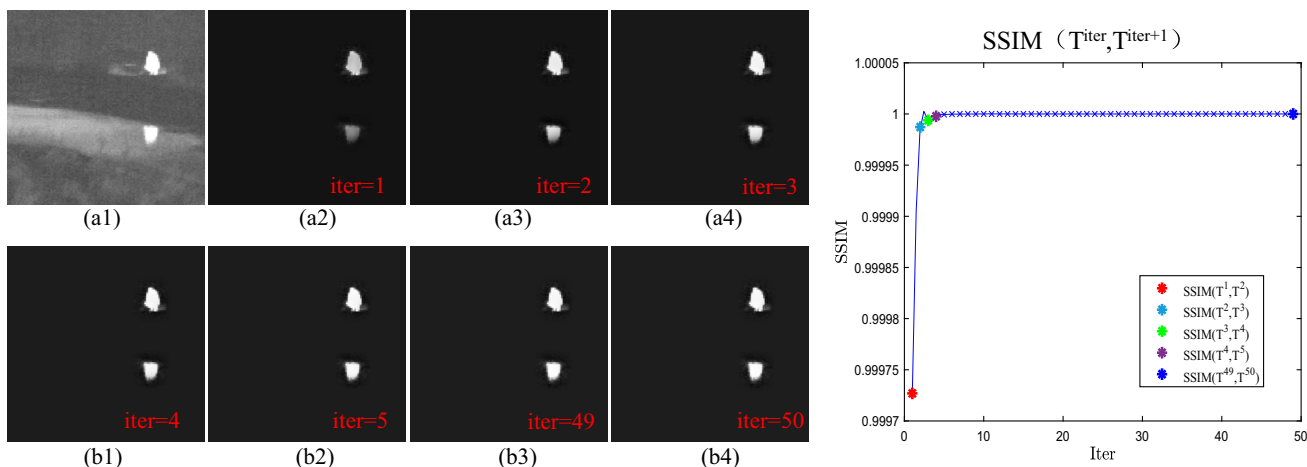
Fig. 7 (**a1**): Infrared image; (**a2**)–(**b4**): finely extracted target map T^{iter} with different number of iterations. The scatter plot represents the structural similarity of two images with adjacent iterations, and the highlighted dots represent the target map T^{iter} on the left side; as iter increases, $SSIM$ increases too and then begins to converge

where i is the pixel index and N is the number of pixels in the image, ∂_j is the partial derivative in x and y directions. The first term $(s_i - b_i)^2$ controls the base layer close to the source image, the second term indicates the ℓ_1 gradient prior term imposed on the base layer, and the third term means the ℓ_0 gradient prior term imposed on the detail layer by a function $F(x)$.

$$F(x) = \begin{cases} 1, & x \neq 0 \\ 0, & else \end{cases}. \tag{19}$$

$model_{\ell_1}(\cdot)$ is the simplification of $model_{\ell_1 - \ell_0}(\cdot)$ which can be represented by (20).

$$\min_{b^2} \sum_{i=1}^{N} \{(b_i^1 - b_i^2)^2 + \lambda_3 \sum_{j=\{x,y\}} |\partial_j b_i^2|\}, \tag{20}$$

In the hybrid decomposed model, two main parameters (λ_1 and λ_2) would affect the decomposed result, and we set λ_1 to 0.3 and λ_2 to 0.003, according to the previously published study [50].

Step 2: Fusion of base layers.

To obtain the fused image with a natural visual effect consistent with the human visual system. The base layer of the fused image should keep a proper contrast. However, in the traditional multi-scale-based fusion algorithms, most adopt the "average" fusion rule [17, 51], which will cause energy loss and result in a poor fusion result. In this paper, we adopt a fusion strategy based on the visual saliency map [52] to solve the above-mentioned problems.

The visual saliency map can be defined as follows:

$$V = |I_q - I_1| + |I_q - I_2| + \cdots + |I_q - I_n|, \tag{21}$$

where I_q is the intensity of a pixel q in the image I, the number of pixels ranges from 1 to n. Then, V is normalized in [0, 1]. The visual saliency map can retain the mainly interesting regions of images.

The fusion weight is obtained based on the visual saliency map V, and the base layer of the fused image is obtained by:

$$B_F = \frac{(1 + V_{IR} - V_{VI})B_{IR} + (1 + V_{VI} - V_{IR})B_{VI}}{2}. \tag{22}$$

Step 3: Fusion of detail layers.

In Sect. 3.2, an infrared target extraction algorithm (IGPT) is proposed, which can maintain the salient infrared targets in the final fused image. Therefore, the key to designing the detail layer fusion rules should be to preserve more valuable visible detail information while preserving the basic saliency of infrared features. Based on the above analysis, we proposed a fusion strategy based on the visual different feature map [53], which is defined as follows:

$$R^i = \begin{cases} |D_{IR}^i| - |D_{VI}^i|, & |D_{IR}^i| - |D_{VI}^i| > 0 \\ 0, & otherwise, \end{cases} \tag{23}$$

where D_{IR}^i, D_{VI}^i is the detail layers obtained by (17), i is the scale ranges from 1 to 2, then R^i is normalized in [0, 1]. And the fusion weight of detail layers is obtained by (24), $Gaus(\cdot)$ is the Gaussian filter used to reduce noise pixels. The fused detail layers can be obtained by (25).

$$C^i = Gaus(R^i), \tag{24}$$

$$D_F^i = C^i D_{IR}^i + (1 - C^i) D_{VI}^i \ i = 1, 2. \tag{25}$$

Step 4: Obtain the Final fused image.

After obtaining the fused base layer and detail layer, the pre-fused image can be reconstructed by

$$F_{pre} = B_F + D_F^1 + D_F^2 \tag{26}$$

And the final fused image F is obtained by

$$F = (1 - T_{IR}) \times F_{pre} + T_{IR} \times IR \tag{27}$$

where T_{IR} is the infrared target map obtained by the proposed IGPT method described in Algorithm 2.

4 Experiments and analysis

4.1 Experimental configuration

The experimental images consist of 21 pairs of infrared and visible images taken from the classical TNO Image Fusion Dataset (TNO). The proposed method is compared with nine related state-of-the-art fusion methods, including IVFusion [24], RFN-NEST [30], Bayesian [28], perceptual FusionGAN (abbreviated as PerGAN) [34], FusionGAN [31], MGFF [19], FPDE [21], ADF [20] and CBF [26]. The relevant parameters of the algorithms above are set according to the original papers. All experiments are implemented in MATLAB R2019b on a 3.40 GHz Intel(R) Core (TM) i7–6700 CPU with 8.00 GB RAM.

In the quantitative comparisons section, five evaluation metrics are adopted: Mean value (ME), pixel-based visual information fidelity (VIFP) [45], Gradient-based fusion metric ($Q^{AB/F}$) [54], Chen-Blum metric (Q_{CB}) [55], and Chen-Varshney metric (Q_{CV}) [56]. These metrics are used to evaluate the performance of the fusion algorithm objectively. This paper first enhances the visible image and then fuses it with the infrared image. So we use the adaptively enhanced visible image, infrared image, and fused image to calculate these evaluation indicators.

Among these indicators, ME calculates the arithmetic mean of all pixels, representing the average brightness that human eyes can perceive. VIFP measures the visual fidelity between the fused image and the source image. $Q^{AB/F}$ is used to estimate the edge information in the fused image. The greater the value, the more edge information from the source image remains. Q_{CB} is designed based on the human visual system and can closely match human perceptual evaluations. For the above indicators, a greater value means a better quality of the fused image. As for Q_{CV}, a human perception inspired indicators that measure the visual difference between the source image and fused image, the smaller the Q_{CV} value, the better the fusion performance.

4.2 Visual comparison and analysis

We first exhibit some intuitive experimental results on three pairs of classic images ('Road', 'Kaptein_1654', and 'Nato_camp') as shown in Figs. 8, 9, and 10, respectively. The results of the other six groups are shown in Fig. 11. From these results, it can be seen that all the algorithms have their peculiarities, and our method has three distinct advantages. First, the proposed method can depict scene information as much as possible, especially for those images captured in low illumination conditions; our method can effectively enhance the hidden details. Second, the proposed can also maintain the saliency of the infrared targets, which are beneficial to some object tracking tasks. Third, our method can obtain natural fusion images suitable for the human visual system.

Specifically, Fig. 8 shows the experimental results of 'Road', where the visible image is taken in a low illumination condition. The infrared targets in the fusion results of RFN-Nest, Bayesian, FusionGAN, MGFF, FPDE, and ADF are not significant enough. At the same time, although the fusion results of IVFusion, PerGAN, and CBF maintain the high contrast of the infrared targets, some valuable visible details are lost (such as the billboard and traffic signs). Compared with these algorithms, our method can highlight the targets and contain more visible details, providing comprehensive scene information.

Moreover, for some video images, our method also performs well. Figure 10 is one frame of the 'Nato_camp' video sequence. The infrared image conveys the message that a man is coming near to something, and the visible image depicts rich scene information (such as the trees and fence). Our method is advantageous as more natural details are contained, which can improve the reliability of some tracking tasks. Fusion results of other groups are exhibited in Fig. 11, where the same phenomena are shown (As magnified by the rectangles).

4.3 Quantitative comparisons

As mentioned in the experimental configuration section, five well-known evaluation metrics (ME, VIFP, $Q^{AB/F}$, Q_{CV} Q_{CB}) are adopted to check the fusion performance objectively. Among these metrics, the larger values of ME, VIFP, $Q^{AB/F}$ and Q_{CB}, the better quality of the fused image. As for Q_{CV} which measure the difference between the source image and the fused image, the smaller the value, the better performance of the fused result.

The detailed scores for each metric on 'Road', 'Kaptein_1654', and 'Nato_camp' are tabulated in Tables 1, 2, and 3, respectively. The best value of each metric is shown in bold, and the second best is underlined and in italic. In most cases, our results rank in a leading position.
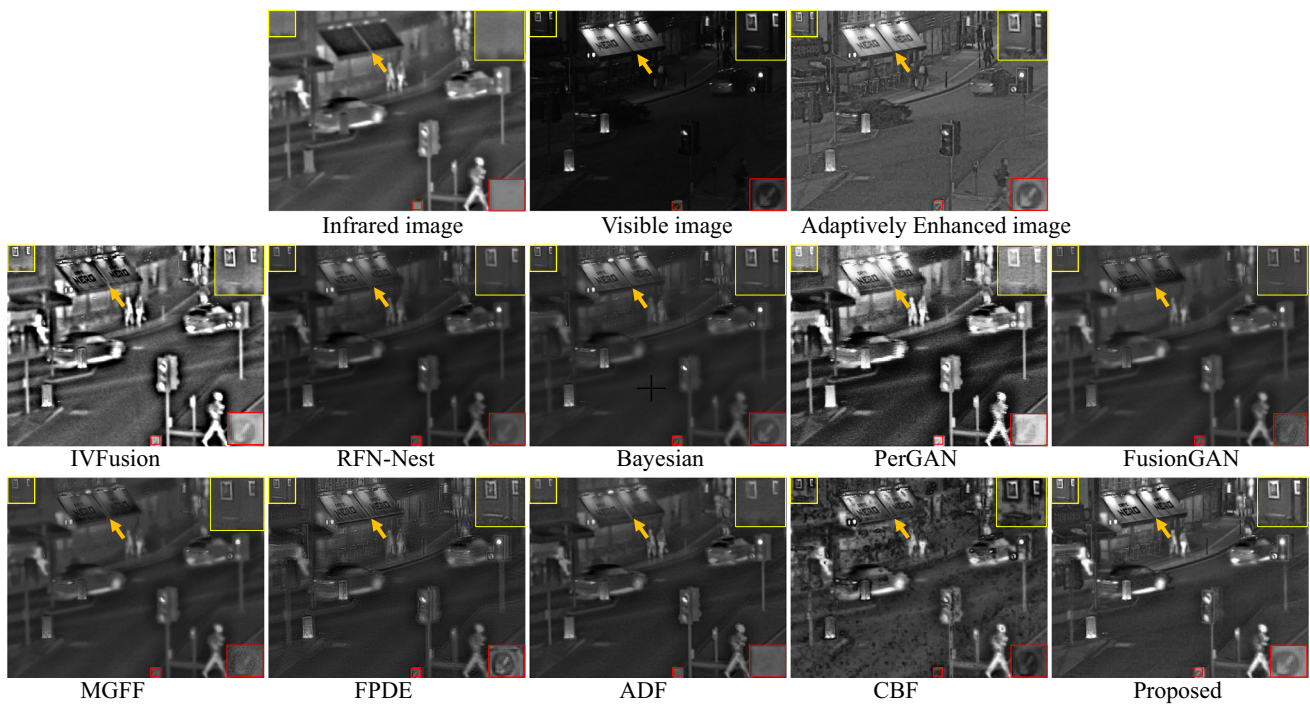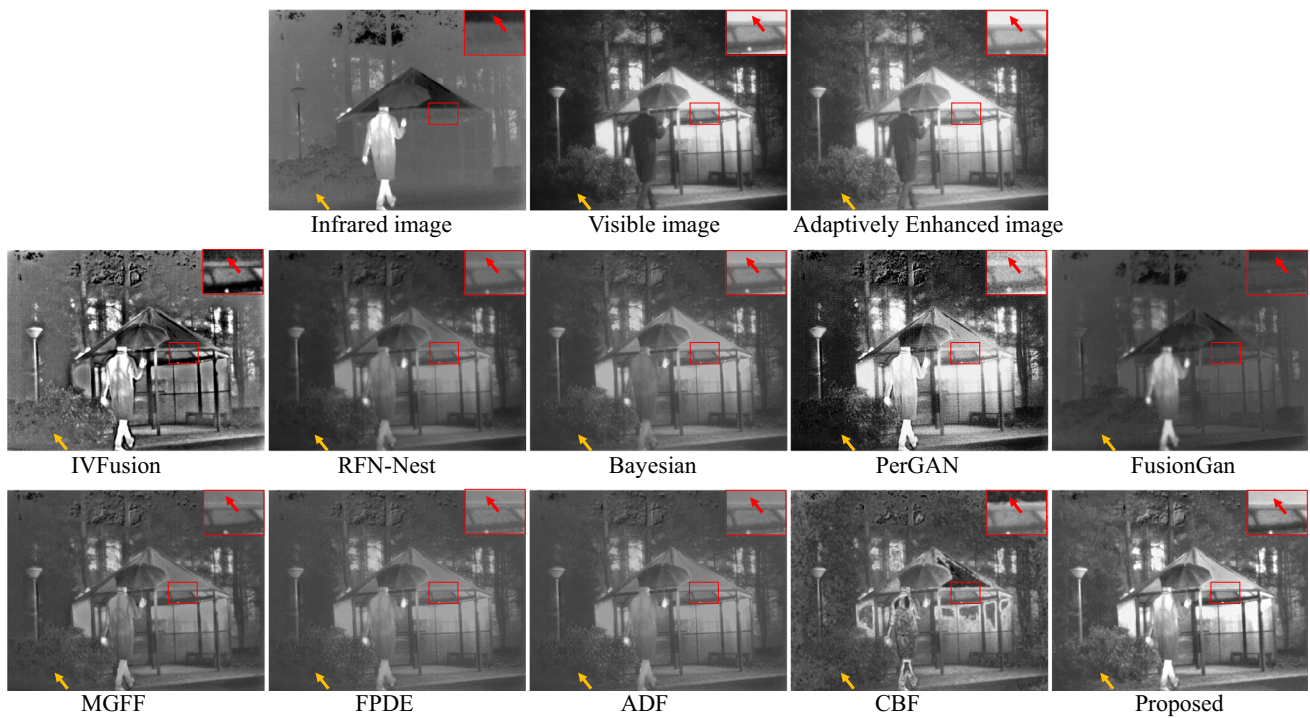
Fig. 8 Fusion results of the "Road" by different methods

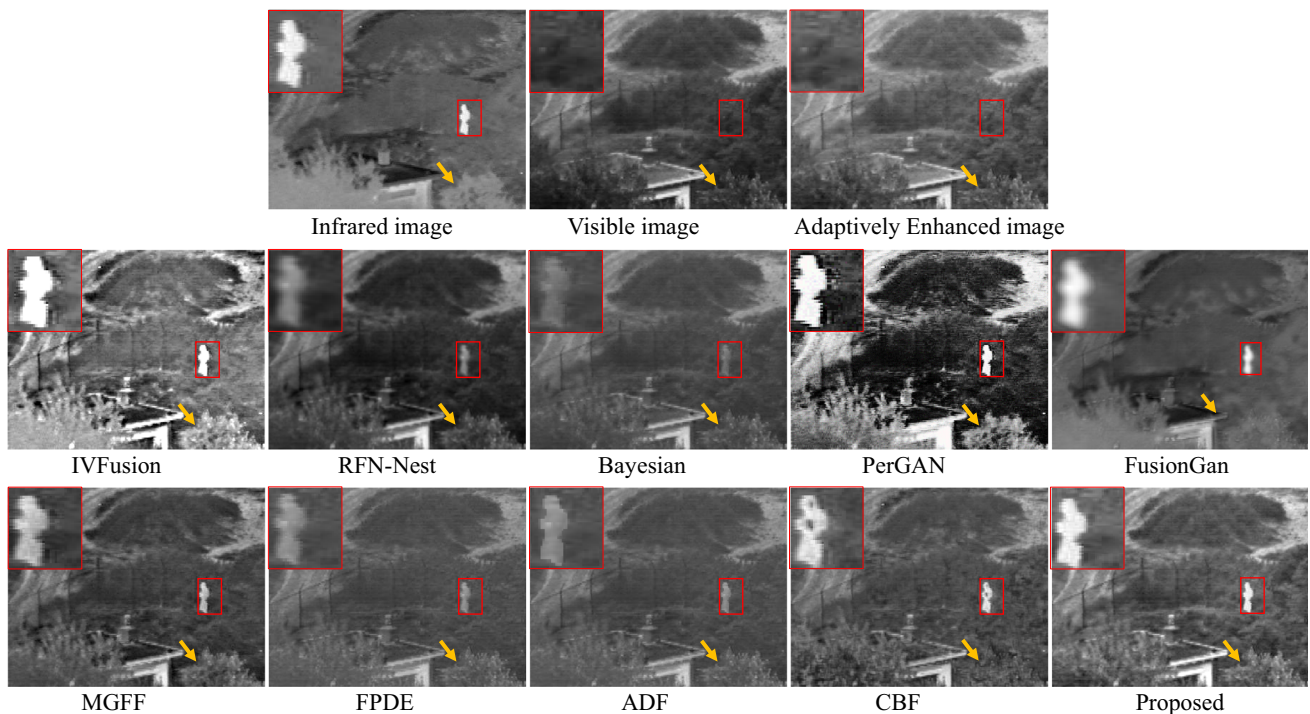**Fig. 9** Fusion results of the "Kaptein_1654" by different methods

Fig. 10 Fusion results of the "Nato_camp" by different methods

The results of quantitative comparisons on 21 pairs of infrared and visible images are summarized in Fig. 12, and the average values are given in Table 4. From the statistical results. Our method achieves the best average score on ME, VIFP, $Q^{AB/F}$, Q_{CV} and Q_{CB}. These results demonstrate that our method can reduce image distortion as much as possible while enhancing visible details and highlighting infrared targets, making the fusion results consistent with the human visual system.

4.4 Ablation study

Infrared and visible fusion aims to generate a fused image with prominent infrared targets and abundant scene texture information, so the BFAC and IGPT algorithms are proposed. The BFAC is used to enhance details in the visible images adaptively, and IGPT is used to highlight the salient infrared targets. An ablation study is conducted to verify the effectiveness of the proposed methods shown in Fig. 13 and Table 5, where 'a' is the fusion method of removing the proposed adaptive visible image enhancement module (BFAC) and infrared target extraction (IGPT) module. 'b' is the fusion method of removing the BFAC module.'c' is the fusion method of removing the IGPT module. 'd' is the proposed method.

As seen in Fig. 13, when removing the BFAC module, the detailed information in the grass is reduced compared with the proposed method (as highlighted by the yellow arrows).

When removing the IGPT module, the saliency of the infrared target is reduced compared with the proposed method (as highlighted by the red rectangles). The objective indicators of the fused images are shown in Table 5. The proposed method achieves the best value on ME, VIFP, $Q^{AB/F}$ and Q_{CB}. The results of the total test images are summarized in Fig. 14, and the average values are shown in Table 6. From the statistical results, our method achieves the best score in metrics: ME, VIFP, $Q^{AB/F}$, Q_{CV} and Q_{CB}, which demonstrated the analysis objectively.

4.5 Parameter analysis

In Sect. 3.2.1, we applied a threshold strategy to remove some isolated pixels in the coarse result. Specifically, when the area of the pixel region is smaller than the threshold, these pixels are reversed. The threshold th can be presented by

$$th = t \times H \times W \qquad (28)$$

where H and W are the height and weight of the infrared image, respectively. t is a user-defined value. In this paper, we conduct a series of experiments to obtain the proper t value and find the range of t is from 0 to 0.01 in our experimental dataset. Due to the space constraint, only two sets of infrared images are exhibited for the analysis. In Fig. 15, when t is equal to 0, all pixels are retained, as the t value continues to increase, it can be seen that the isolated pixels gradually
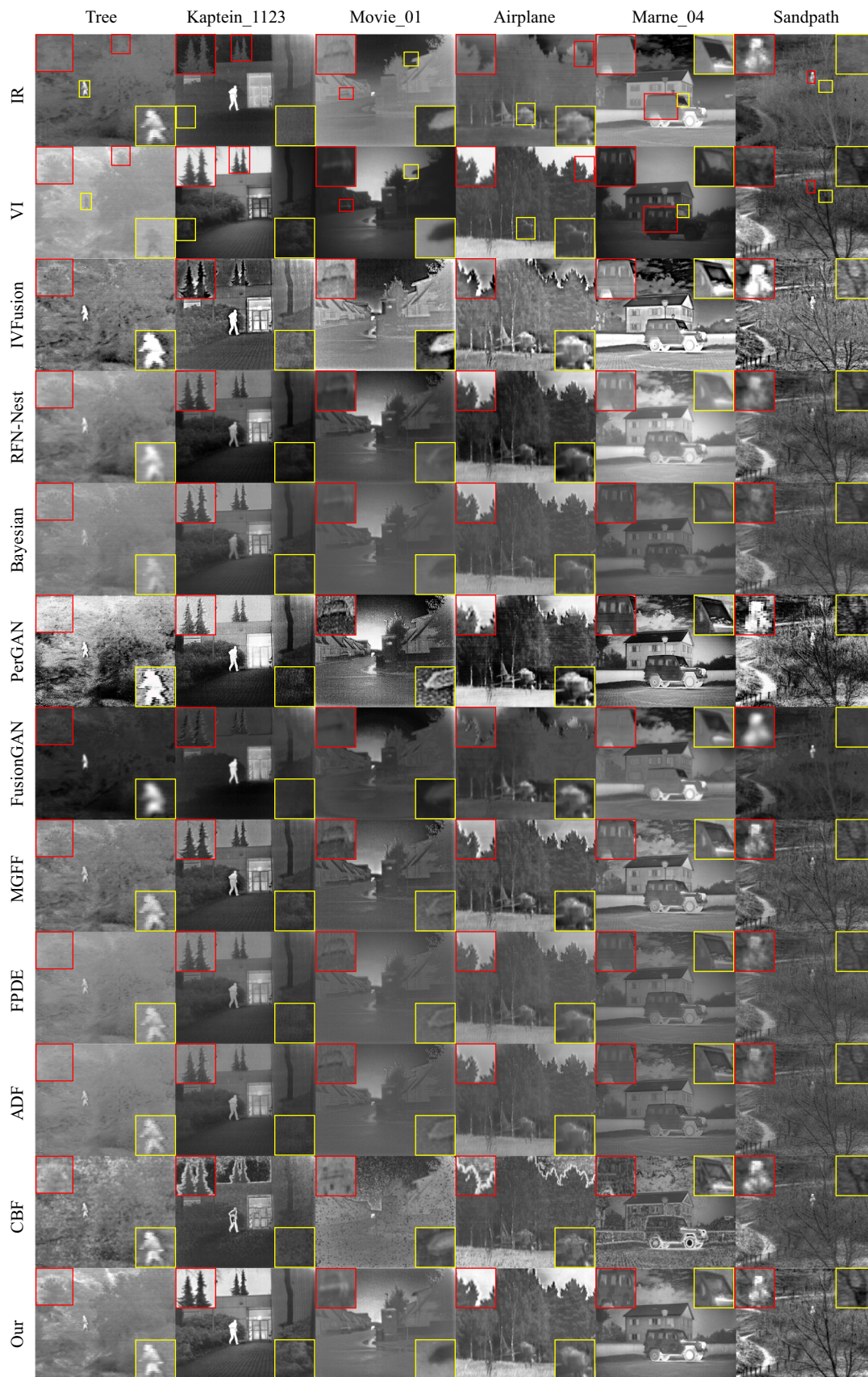
Fig. 11 Fusion results of six pairs of infrared and visible images by different methods

Table 1 Quantitative comparison results on 'Road.'

Method	ME	VIFP	$Q^{AB/F}$	Q_{CV}	Q_{CB}
Proposed	_86.91_	0.3746	**0.6042**	**478.32**	**0.5999**
IVFusion	**93.34**	**0.4345**	0.2752	2319.48	0.4180
RFNNest	42.41	0.2862	0.2129	837.71	0.4538
Bayesian	47.33	0.3179	0.3408	_513.07_	0.3947
PerGAN	80.49	_0.3802_	0.2537	2522.86	0.3915
FusionGan	55.28	0.2588	0.2411	1467.66	0.3923
MGFF	45.17	0.3418	0.4060	1038.32	_0.5118_
FPDE	52.40	0.2149	0.2673	844.36	0.4345
ADF	51.78	0.3020	0.4206	1013.37	0.4849
CBF	65.16	0.2077	_0.4638_	836.87	0.4473

Table 2 Quantitative comparison results on 'Kaptein_1654'

Method	ME	VIFP	$Q^{AB/F}$	Q_{CV}	Q_{CB}
Proposed	**111.91**	**0.4123**	**0.5688**	**216.51**	**0.5168**
IVFusion	_108.68_	0.3346	0.3110	392.92	0.4612
RFNNest	72.58	0.2613	0.2959	364.03	0.4159
Bayesian	89.25	_0.3357_	_0.4902_	420.83	0.4796
PerGAN	93.12	0.2905	0.2259	353.10	_0.4854_
FusionGan	73.15	0.1765	0.1467	439.18	0.3148
MGFF	86.49	0.2933	0.4502	_271.91_	0.4540
FPDE	93.27	0.2700	0.4286	334.50	0.4024
ADF	92.86	0.2793	0.4534	320.33	0.4454
CBF	94.36	0.1902	0.4323	319.05	0.4601

Table 3 Quantitative comparison results on 'Nato_camp'

Method	ME	VIFP	$Q^{AB/F}$	Q_{CV}	Q_{CB}
Proposed	_105.40_	**0.3802**	**0.5433**	434.25	**0.6216**
IVFusion	**126.60**	0.2469	0.2804	795.25	0.5304
RFNNest	66.65	0.2470	0.3968	431.12	0.5769
Bayesian	88.90	_0.2939_	_0.4567_	430.18	_0.6174_
PerGAN	95.28	0.2221	0.1918	1115.18	0.4304
FusionGan	83.98	0.1766	0.2420	760.20	0.4256
MGFF	82.50	0.2367	0.4173	453.13	0.5589
FPDE	91.59	0.2266	0.4160	_399.95_	0.5603
ADF	91.14	0.2539	0.4503	**382.91**	0.5755
CBF	89.89	0.1605	0.3925	510.61	0.5191

decreased. Particularly, when t is set to 0.0002, all the isolated pixels are revised, and the extracted infrared targets tend to stabilize until t is enlarged to 0.01. Therefore, we set the threshold th to $0.0002 \times H \times W$. It is worth noting that, th is mainly used to delete some isolated pixels in images, and it is a user-defined value that can be determined according to different tasks.

5 Conclusion

In this paper, a novel adaptive visual enhancement and salient targets analysis-based fusion scheme is proposed. First, in order to address the problem that visible images are easily affected by illumination, a visual adaptive enhancement algorithm for visible images (named BFAC) is proposed. Based on the bright-pass bilateral filter, the algorithm corrects the brightness of the visible image by calculating the adaptive
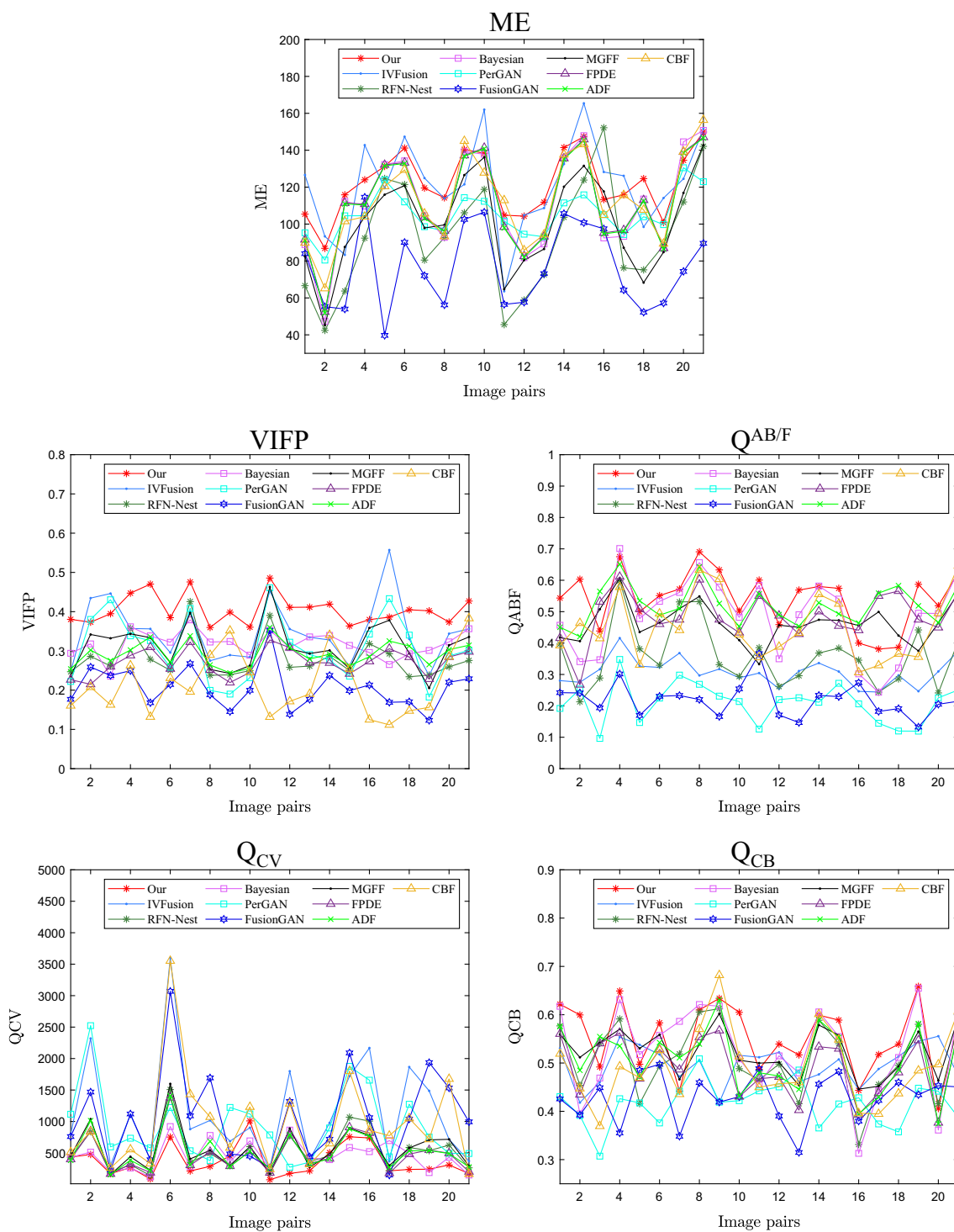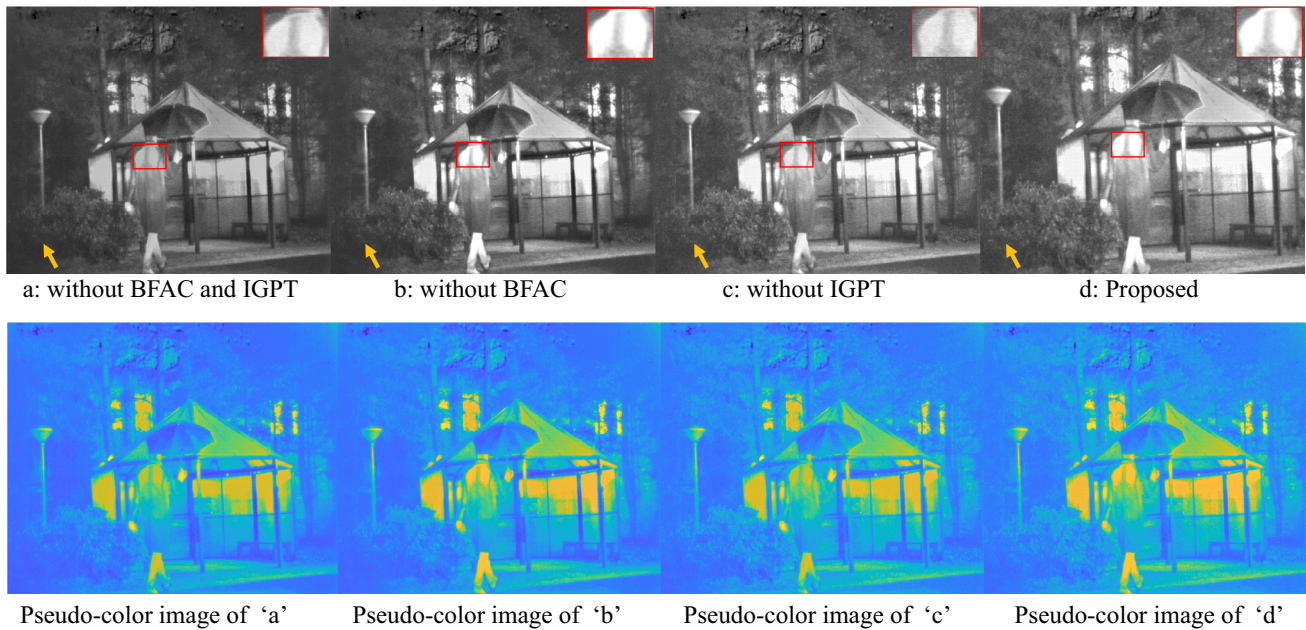
Fig. 12 Comparison of our method with nine state-of-the-art methods on 21 infrared and visible image pairs

Table 4 Average quantitative comparison results on 21 testing image pairs

Method	ME	VIFP	$Q^{AB/F}$	Q_{CV}	Q_{CB}
Proposed	**122.27**	**0.4082**	**0.5433**	**374.09**	**0.5537**
IVFusion	_121.58_	_0.3485_	0.3068	1148.0	0.4966
RFNNest	93.34	0.2809	0.3598	534.25	0.5078
Bayesian	111.27	0.3176	0.4840	_453.58_	_0.5252_
PerGan	105.50	0.3075	0.2092	919.91	0.4160
FusionGan	76.37	0.2064	0.2189	1061.6	0.4283
MGFF	100.82	0.3103	0.4633	563.57	0.5233
FPDE	111.62	0.2707	0.4879	489.42	0.4923
ADF	111.11	0.2928	_0.5209_	518.58	0.5073
CBF	112.84	0.2160	0.4540	947.61	0.4933

a: without BFAC and IGPT   b: without BFAC   c: without IGPT   d: Proposed

Pseudo-color image of 'a'   Pseudo-color image of 'b'   Pseudo-color image of 'c'   Pseudo-color image of 'd'

Fig. 13 Effectiveness illustration of BFAC and IGPT

Table 5 Quantitative comparison results of the different ablation experiments on 'Kaptein_1654'

Method	ME	VIFP	$Q^{AB/F}$	Q_{CV}	Q_{CB}
a: Without two modules	98.45	0.4005	0.5337	_209.35_	0.4918
b: Without BFAC	98.82	0.4057	0.5403	**198.91**	0.4920
c: Without IGPT	_111.6_	_0.4069_	_0.5617_	230.48	_0.5163_
d: Proposed	**111.9**	**0.4123**	**0.5688**	216.51	**0.5168**

gamma value to obtain the enhanced visible image with an excellent visual effect. Secondly, an infrared target extraction algorithm (named IGPT) is proposed. The algorithm obtains the infrared target by calculating the tensor rank, iteratively guided filtering and solving the convergence condition. Compared with the previous IPT model and BASNet, our IGPT has more accurate results as significant infrared targets and continuous edges are contained. Third, an efficient hybrid $\ell_1-\ell_0$ model decomposes the infrared and visible image into base and detail layers, then fuses them by weight strategy. The final fused image is obtained by merging the fused base layers, detail layers, and infrared targets.

Qualitative and quantitative experimental results demonstrate that the proposed method is superior to 9 state-of-the-art image fusion methods as more valuable texture details and significant infrared targets are preserved. However, in
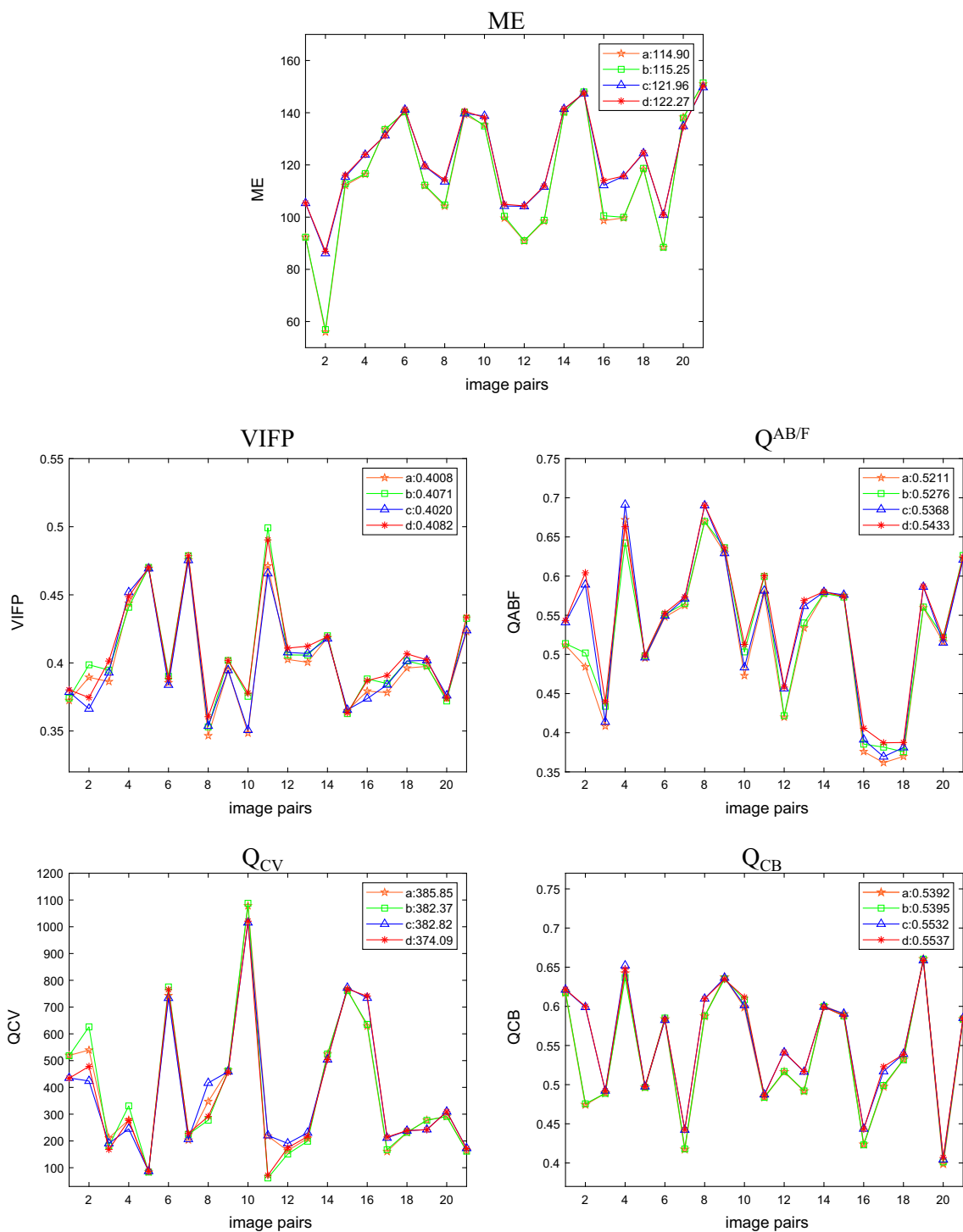
Fig. 14 Quantitative comparison of the different ablation experiments on 21 pairs of infrared and visible images

Table 6 Average quantitative comparison results of the different ablation experiments on 21 pairs of infrared and visible images

Method	ME	VIFP	$Q^{AB/F}$	Q_{CV}	Q_{CB}
a: without two modules	114.9	0.4008	0.5211	385.85	0.5392
b: without BFAC	115.2	*0.4071*	0.5276	*382.37*	0.5395
c: without IGPT	*121.9*	0.4020	*0.5368*	382.82	*0.5532*
d: Proposed	**122.2**	**0.4082**	**0.5433**	**374.09**	**0.5537**
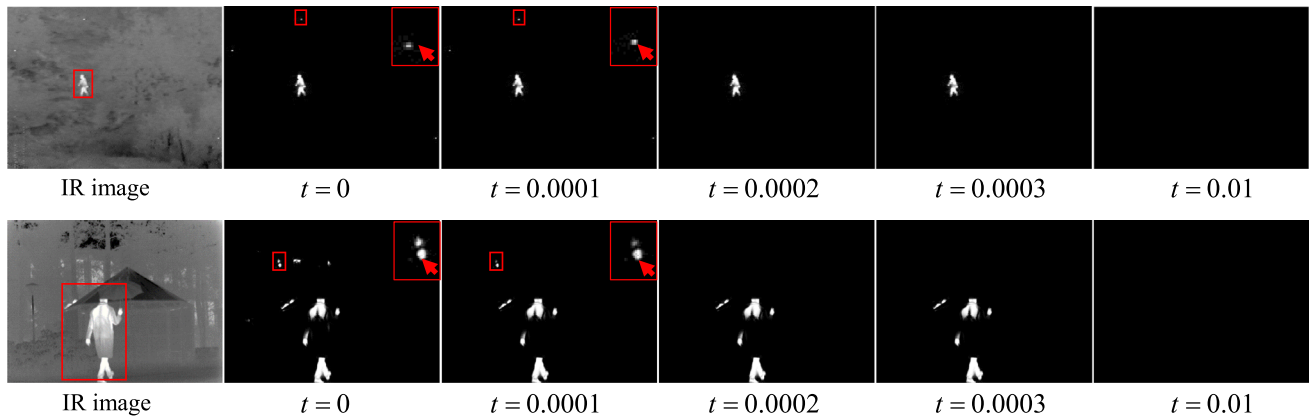
Fig. 15 The changes of isolated pixel points when t gradually increases

the stage of the infrared target extraction, an ADMM framework is applied to iterative solve the optimization model, resulting in more computing resources and the average running time of our method is about 7 s. In our future studies, we will focus on some light-weight networks and efficient optimization algorithms to make the proposed method more suitable for some real-time industrial applications.

Declarations

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of the article.

References

1. Lu, R., Gao, F., Yang, X., Fan, J., Li, D.: A novel infrared and visible image fusion method based on multi-level saliency integration. Vis. Comput. **1**, 1–15 (2022)
2. Wang, X., Hua, Z., Li, J.: Cross-UNet: dual-branch infrared and visible image fusion framework based on cross-convolution and attention mechanism. Vis. Comput. **1**, 1–18 (2022)
3. Liu, J., Jiang, Z., Wu, G., Liu, R., Fan, X.: A unified image fusion framework with flexible bilevel paradigm integration. Vis. Comput. **1**, 1–18 (2022)
4. Ma, J., Ma, Y., Li, C.: Infrared and visible image fusion methods and applications: a survey. Inf. Fusion. **45**, 153–178 (2019). https://doi.org/10.1016/j.inffus.2018.02.004
5. Zhang, X., Ye, P., Xiao, G.: VIFB: A visible and infrared image fusion benchmark. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14–19, 2020. pp. 468–478. Computer Vision Foundation/IEEE (2020)
6. Jagtap, N.S., Thepade, S.D.: High-quality image multi-focus fusion to address ringing and blurring artifacts without loss of information. Vis. Comput. **1**, 1–19 (2021)
7. Aghamaleki, J.A., Ghorbani, A.: Image fusion using dual tree discrete wavelet transform and weights optimization. Vis. Comput. (2022). https://doi.org/10.1007/s00371-021-02396-9
8. Wang, C., He, C., Xu, M.: Fast exposure fusion of detail enhancement for brightest and darkest regions. Vis. Comput. **37**, 1233–1243 (2021)

9. He, K., Gong, J., Xie, L., Zhang, X., Xu, D.: Regions preserving edge enhancement for multisensor-based medical image fusion. IEEE Trans. Instrum. Meas. **70**, 1–13 (2021). https://doi.org/10.1109/TIM.2021.3066467

10. Chen, Y., Shi, K., Ge, Y., Zhou, Y.: Spatiotemporal remote sensing image fusion using multiscale two-stream convolutional neural networks. IEEE Trans. Geosci. Remote. Sens. **60**, 1–12 (2022). https://doi.org/10.1109/TGRS.2021.3069116

11. Luo, Y., He, K., Xu, D., Yin, W., Liu, W.: Infrared and visible image fusion based on visibility enhancement and hybrid multi-scale decomposition. Optik **1**, 168914 (2022)

12. Luo, Y., He, K., Xu, D., Yin, W.: Infrared and visible image fusion based on visibility enhancement and norm optimization low-rank representation. J. Electron. Imaging **31**, 013032 (2022)

13. Yin, W., He, K., Xu, D., Luo, Y., Gong, J.: Adaptive enhanced infrared and visible image fusion using hybrid decomposition and coupled dictionary. Neural Comput. Appl. **1**, 1–19 (2022)

14. Soroush, R., Baleghi, Y.: NIR/RGB image fusion for scene classification using deep neural networks. Vis. Comput. **1**, 1–15 (2022)

15. Piella, G.: A general framework for multiresolution image fusion: from pixels to regions. Inf. Fusion. **4**, 259–280 (2003). https://doi.org/10.1016/S1566-2535(03)00046-0

16. Jin, X., Jiang, Q., Yao, S., Zhou, D., Nie, R., Hai, J., He, K.: A survey of infrared and visual image fusion methods. Infrar. Phys. Technol. **85**, 478–501 (2017)

17. Chen, J., Li, X., Luo, L., Mei, X., Ma, J.: Infrared and visible image fusion based on target-enhanced multiscale transform decomposition. Inf. Sci. **508**, 64–78 (2020). https://doi.org/10.1016/j.ins.2019.08.066

18. Zhan, L., Zhuang, Y., Huang, L.: Infrared and visible images fusion method based on discrete wavelet transform. J. Comput. **28**, 57–71 (2017)

19. Bavirisetti, D.P., Xiao, G., Zhao, J., Dhuli, R., Liu, G.: Multi-scale guided image and video fusion: a fast and efficient approach. Circuits Syst. Signal Process. **38**, 5576–5605 (2019). https://doi.org/10.1007/s00034-019-01131-z

20. Bavirisetti, D.P., Dhuli, R.: Fusion of infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform. IEEE Sens. J. **16**, 203–209 (2015)

21. Bavirisetti, D.P., Xiao, G., Liu, G.: Multi-sensor image fusion based on fourth order partial differential equations. In: 20th International Conference on Information Fusion, FUSION 2017, Xi'an, China, July 10–13, 2017, pp. 1–9. IEEE (2017)

22. Bashir, R., Junejo, R., Qadri, N.N., Fleury, M., Qadri, M.Y.: SWT and PCA image fusion methods for multi-modal imagery. Multim. Tools Appl. **78**, 1235–1263 (2019). https://doi.org/10.1007/s11042-018-6229-5

23. Mitianoudis, N., Stathaki, T.: Pixel-based and region-based image fusion schemes using ICA bases. Inf. Fusion. **8**, 131–142 (2007). https://doi.org/10.1016/j.inffus.2005.09.001

24. Li, G., Lin, Y., Qu, X.: An infrared and visible image fusion method based on multi-scale transformation and norm optimization. Inf. Fusion. **71**, 109–129 (2021)

25. Yin, W., He, K., Xu, D., Luo, Y., Gong, J.: Significant target analysis and detail preserving based infrared and visible image fusion. Infrar. Phys. Technol. 104041 (2022)

26. Kumar, B.S.: Image fusion based on pixel significance using cross bilateral filter. SIViP **9**, 1193–1204 (2015)

27. Li, S., Kang, X., Hu, J.: Image fusion with guided filtering. IEEE Trans. Image Process. **22**, 2864–2875 (2013)

28. Zhao, Z., Xu, S., Zhang, C., Liu, J., Zhang, J.: Bayesian fusion for infrared and visible images. Signal Process. **177**, 107734 (2020). https://doi.org/10.1016/j.sigpro.2020.107734

29. Panigrahy, C., Seal, A., Mahato, N.K.: Parameter adaptive unit-linking dual-channel PCNN based infrared and visible image fusion. Neurocomputing **514**, 21–38 (2022)

30. 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). https://doi.org/10.1016/j.inffus.2021.02.023

31. 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). https://doi.org/10.1016/j.inffus.2018.09.004

32. Li, Q., Lu, L., Li, Z., Wu, W., Liu, Z., Jeon, G., Yang, X.: Coupled GAN with relativistic discriminators for infrared and visible images fusion. IEEE Sens. J. **21**, 7458–7467 (2019)

33. Tan, Z., Gao, M., Li, X., Jiang, L.: A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network. IEEE Trans. Geosci. Remote Sens. **60**, 1–13 (2021)

34. Fu, Y., Wu, X.-J., Durrani, T.S.: Image fusion based on generative adversarial network consistent with perception. Inf. Fusion. **72**, 110–125 (2021). https://doi.org/10.1016/j.inffus.2021.02.019

35. Maurya, L., Mahapatra, P.K., Kumar, A.: A social spider optimized image fusion approach for contrast enhancement and brightness preservation. Appl. Soft Comput. **52**, 575–592 (2017). https://doi.org/10.1016/j.asoc.2016.10.012

36. Xu, J., Hou, Y., Ren, D., Liu, L., Zhu, F., Yu, M., Wang, H., Shao, L.: STAR: A Structure and Texture Aware Retinex Model. IEEE Trans. Image Process. **29**, 5022–5037 (2020). https://doi.org/10.1109/TIP.2020.2974060

37. Tomasi, C., Manduchi, R.: Bilateral Filtering for Gray and Color Images. In: Proceedings of the Sixth International Conference on Computer Vision (ICCV-98), Bombay, India, January 4–7, 1998. pp. 839–846. IEEE Computer Society (1998)

38. Ghosh, S., Chaudhury, K.N.: Fast Bright-Pass Bilateral Filtering for Low-Light Enhancement. In: 2019 IEEE International Conference on Image Processing, ICIP 2019, Taipei, Taiwan, September 22–25, 2019. pp. 205–209. IEEE (2019)

39. Kong, X., Yang, C., Cao, S., Li, C., Peng, Z.: Infrared small target detection via nonconvex tensor fibered rank approximation. IEEE Trans. Geosci. Remote Sens. **60**, 1–21 (2021)

40. Chen, Y., Li, J., Zhou, Y.: Hyperspectral image denoising by total variation-regularized bilinear factorization. Signal Process. **174**, 107645 (2020)

41. Zheng, Y.-B., Huang, T.-Z., Zhao, X.-L., Jiang, T.-X., Ma, T.-H., Ji, T.-Y.: Mixed noise removal in hyperspectral image via low-fibered-rank regularization. IEEE Trans. Geosci. Remote Sens. **58**, 734–749 (2019)

42. Eckstein, J., Bertsekas, D.P.: On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992)

43. Dehaene, S.: The neural basis of the Weber-Fechner law: a logarithmic mental number line. Trends Cogn. Sci. **7**, 145–147 (2003)

44. Ying, Z., Li, G., Ren, Y., Wang, R., Wang, W.: A New Image Contrast Enhancement Algorithm Using Exposure Fusion Framework. In: Felsberg, M., Heyden, A., and Krüger, N. (eds.) Computer Analysis of Images and Patterns - 17th International Conference, CAIP 2017, Ystad, Sweden, August 22–24, 2017, Proceedings, Part II. pp. 36–46. Springer (2017)

45. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. **15**, 430–444 (2006). https://doi.org/10.1109/TIP.2005.859378

46. Liu, Y., Chen, X., Peng, H., Wang, Z.: Multi-focus image fusion with a deep convolutional neural network. Inf. Fus. **36**, 191–207 (2017)

47. He, K., Sun, J., Tang, X.: Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1397–1409 (2012)

48. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)

49. Qin, X., Zhang, Z.V., Huang, C., Gao, C., Dehghan, M., Jägersand, M.: BASNet: Boundary-Aware Salient Object Detection. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16–20, 2019. pp. 7479–7489. Computer Vision Foundation/IEEE (2019)

50. Liang, Z., Xu, J., Zhang, D., Cao, Z., Zhang, L.: A Hybrid l1-l0 Layer Decomposition Model for Tone Mapping. In: 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18–22, 2018. pp. 4758–4766. Computer Vision Foundation / IEEE Computer Society (2018)

51. Liu, Y., Liu, S., Wang, Z.: A general framework for image fusion based on multi-scale transform and sparse representation. Inf. Fus. **24**, 147–164 (2015). https://doi.org/10.1016/j.inffus.2014.09.004

52. Ma, J., Zhou, Z., Wang, B., Zong, H.: Infrared and visible image fusion based on visual saliency map and weighted least square optimization. Infrar. Phys. Technol. **82**, 8–17 (2017)

53. 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). https://doi.org/10.1016/j.inffus.2015.11.003

54. Xydeas, C., Petrovic, V.: Objective image fusion performance measure. Electron. Lett. **36**, 308–309 (2000)

55. Chen, Y., Blum, R.S.: A new automated quality assessment algorithm for image fusion. Image Vis. Comput. **27**, 1421–1432 (2009). https://doi.org/10.1016/j.imavis.2007.12.002

56. Chen, H., Varshney, P.K.: A human perception inspired quality metric for image fusion based on regional information. Inf. Fusion. **8**, 193–207 (2007). https://doi.org/10.1016/j.inffus.2005.10.001

Wenxia Yin received her B.S. degree in computer science and technology from Huazhong Agricultural University, Wuhan, China, in 2020. She is currently pursuing her master's degree at Department of Computer Science and Engineering, Yunnan University, Kunming, China. Her research focuses on infrared and visible image fusion.
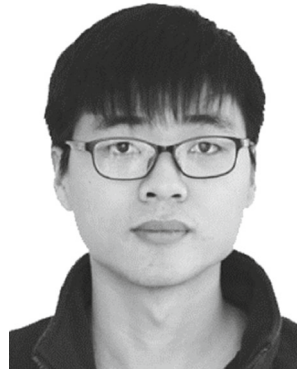
Kangjian He received the Ph.D. degree from the Yunnan University, Kunming, China, in 2019. He is currently a Postdoctoral Research Fellow with the School of Information Science and Engineering, Yunnan University. He has authored and co-authored over 50 papers in refereed international journals and conferences, and his current research interests include image processing, neural network theory and applications.

Dan Xu received the M.Sc. and Ph.D. degrees in computer science and technology from Zhejiang University, Hangzhou, China, in 1993 and 1999, respectively. She is currently a Professor with the School of Information Science and Engineering, Yunnan University, Kunming, China. Her research interests include different aspects of image-based modeling and rendering, image processing and understanding, nonphotorealistic rendering, and 3-D reconstruction. She is a senior member of China Computer Federation.

Yingying Yue received the M.Sc. degree in computer science and technology from Yunnan University, Kunming, China, in 2010. Currently, she is pursuing her Ph.D. at Department of Computer Science and Engineering, Yunnan University, Kunming, China. Her research focuses on multi-target detection and tracking in complex scenes.
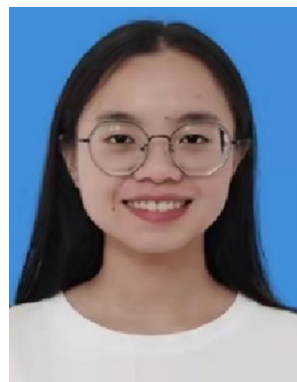
Yueying Luo is pursuing her M.D. at Department of Computer Science and Engineering, Yunnan University, Kunming, China. Her research focuses on image processing and fusion algorithm.