# Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint With Retinex Enhancement and Multi-Visual Weight Information

**QIAOYING PAN** [1], **LIAOYING ZHAO** [1], **SHUHAN CHEN** [2], **(Member, IEEE), AND XIAORUN LI** [2]
[1]Department of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China
[2]College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China

Corresponding author: Xiaorun Li (lxr@zju.edu.cn)

**ABSTRACT** Latent low-rank representation has been applied to multi-level image decomposition for the fusion of infrared and visible images to obtain good results. However, when the original infrared and visible images are of low quality, the visual effect of the fused images is still unsatisfactory. To combat this challenge, this paper proposes an infrared and visible image fusion method based on multi-level latent low-rank representation joint with image enhancement and multiple visual weight information. First, the source images are decomposed into detail parts - including detail images and detail matrices - and the base images respectively using multi-level latent low-rank representation. Then the nuclear norm based fusion strategy is used to fuse the detail matrices and multi-visual weights determined by the clarity, local contrast and edge-corner saliency is used to fuse the detail images. The aforementioned two fusion results are weight averaged to obtain a fused detail image. The base images are fused by an averaging strategy after Retinex-based enhancement. The final fused image is obtained by combining the fused detail image and the fused base image. Compared with other state-of-the-art fusion methods, the proposed algorithm displays better fusion performance in both subjective and objective evaluation.

**INDEX TERMS** Image fusion, latent low-rank representation, multiple visual weight, Retinex-based enhancement, infrared image, visible image.

## I. INTRODUCTION

Image fusion is an important branch of multi-sensor information fusion. It aims to integrate the image information of a certain moment under the same scene obtained by different types of sensors, which aids in describing the characteristics of the target scene more comprehensively [1]. The visible images capture scene information based on reflection mechanism. Under well-lit conditions, the obtained image has a high resolution, but it is difficult to capture all the information in the scene at night or under low visibility conditions such as fog and haze. The imaging of infrared images occurs by

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Wei.

detecting the thermal radiation of the object itself, which highlights the thermal target blocked by the object. This technique reduces interference by environmental factors, but its texture detail information is insufficient and the contrast is low [2]. Infrared and visible image fusion can fully combine the advantages of these two imaging technologies. The fused image contains highly-detailed texture information and clear infrared thermal radiation targets, which is conducive to further target detection and other work [3], [4].

In recent years, with the development of fusion technology and the demand of practical application, many infrared and visible image fusion algorithms have been proposed. The multi-scale transform method is the most widely used image fusion method. The transform domain algorithms include the

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint

IEEE*Access*

wavelet transform [4], [5], the shearlet transform [6], the curvelet transform [7], and the non-subsampled contourlet transform [8]. This type of method usually comprises three steps. First, each source image is decomposed into several levels. Then the decomposed layers are integrated by appropriate fusion rules. Finally, the fused image is obtained by the corresponding inverse transformation. In recent years, the multi-scale transform based methods are often combined with neural network to achieve better fusion performance [9], [10]. However, it is still challenging to select flexible basis functions that allow data driven choice of the best representation of source images, and the adaptive selection of decomposition levels still remains to be solved [1]. Hence, researchers attempt to find other methods to process source images without a transform, such as deep learning-based method and representation learning-based method.

Deep learning-based fusion methods specialized for infrared and visible image fusion include the convolutional neural network (CNN) [11], DenseFuse [12], the disentangled representation fusion (DRF) [13] and so on. The main drawback of deep learning-based methods is that it is difficult to train when the training data is insufficient, especially in infrared and visible image fusion tasks, and a very little attention is paid to image decomposition in the deep learning based methods [14].

Sparse representation is a commonly used method in the domain of representation learning. The image fusion method based on sparse representation does not need to transform the image to a certain frequency domain. This kind of method uses a sliding window to divide the source image into blocks, reducing the impact of image misregistration [15]. For example, Zhang *et al.* used overlapping sub-blocks of infrared and visible images for training to construct an over-complete dictionary [16]. To increase the fusion effect, sparse representation usually combines with other tools, such as shearlet transform [17] and image cartoon-texture decomposition [18].

Although sparse representation can overcome the deficiency of multi-scale transformation and achieve better fusion performance, it still suffers from the insufficient ability of detail information extraction [14]. Latent low-rank representation (LatLRR) [19] can effectively extract local and global structural information simultaneously. Li *et al.* used LatLRR for infrared and visible image fusion [20]. However, the LatLRR-based image decomposition method only extracts incomplete high-frequency information, which will lead to unsatisfied fusion results. To alleviate the issue, the basic image decomposed by LatLRR was further decomposed into high-frequency and low-frequency components via the non-subsampled shearlet transform in [21]. Li *et al.* further proposed a LatLRR based multi-level decomposition method (MDLatLRR) [14], which performs multi-level LatLRR decomposition, and can extract more details from the source image. This paper focuses on the fusion method of multi-level LatLRR decomposition.

MDLatLRR first decomposed the images by multi-level LatLRR and then averaged the low-rank images. The significance coefficient matrix was fused block by block by a weighted fusion method based on nuclear norm. Multi-level LatLRR decomposition of images can remove image noise, but it cannot effectively solve the clarity and contrast problems of blurred visible and infrared images. This results in an unsatisfactory visual effect. Based on Retinex theory, the reflection can be extracted by removing illumination to enhance the image details [22]. Recently, Retinex theory has been widely used in the enhancement of low-quality visible [23]–[25] and infrared images [26]. Although the quality of image is improved by implementing Retinex enhancement, some targets are still not clear or invisible in the enhanced image because these targets could not be captured in the source image due to occlusion or low visibility conditions. Image fusion can combine the advantages of thermal radiation information in infrared images and detailed texture information in visible images to make the missing targets to be observed or to make some vague targets clearer. The combination of image fusion and image enhancement can make full use of the advantages of the two methods to improve the fusion effect [27]. We will consider the enhancement before fusion.

In general, a good fusion algorithm depends on the image decomposition method on the one hand, and the fusion strategy on the other hand [28]. In MDLatLRR, the fusion strategy of detail parts is based on nuclear-norm, which calculates the sum of single value of each input patch to preserve 2D information from the source images. It is usually not sufficient to retain the detail information and saliency target in full measure only with nuclear norm. The importance of other visual salient features in detail image fusion has also been verified in [28]–[31]. We will further consider other visual salient features for the fusion of the detail parts.

In summary, this paper focuses on the problem of fusion for blurred visible and infrared images by proposing an infrared and visible image fusion method based on latent low-rank representation combined with image enhancement and multiple visual weights. After the multi-level LatLRR decomposition of the image, the Retinex enhancement and the image visual fusion weight calculation are combined. The innovations of the method in this paper are as follows:

(1) A novel visible and infrared image fusion framework based on multi-level latent low-rank decomposition joint with image enhancement, multi-visual information, and nuclear-norm is proposed to be suitable for the fusion of low-quality visible and infrared images.

(2) Retinex enhancement was performed on the base image, which is decomposed by MDLatLRR to improve the clarity of the final fused images.

(3) To enhance the saliency of the fused image, the fused detail image was obtained by weight summarizing the result of multi-visual fusion and nuclear norm fusion. The weights of multi-visual fusion are constructed by combining three
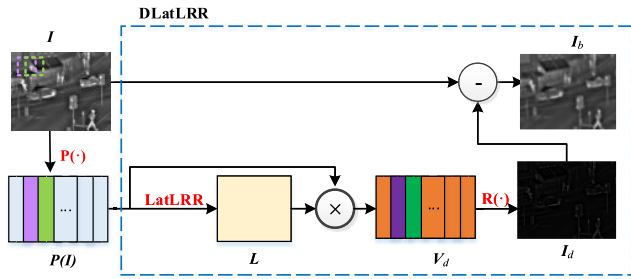
**FIGURE 1.** One level LatLRR decomposition of images.

types of visual information, which are marginal significance, local contrast, and clarity.

The rest of the paper is organized as follows. In Section II, we introduce preliminaries, including image decomposition based on LatLRR and Retinex-based image enhancement. In Section III, two strategies for the combination of MDLatLRR decomposition and image enhancement are discussed. In Section IV, the proposed image fusion method is presented in detail. The experimental result and analysis are shown in Section V. Finally, in Section VI we draw our conclusions.

## II. PRELIMINARIES
### A. IMAGE DECOMPOSITION BASED ON LATENT LOW-RANK REPRESENTATION
The main idea of the LatLRR [19] is to express the original data matrix as the sum of low-rank components (global structure), significant components (local structure), and sparse noise. The mathematical model of LatLRR can be expressed as

$$\min_{Z,L,E} \|Z\|_* + \|L\|_* + \lambda\|E\|_1$$
$$\text{s.t. } X = XZ + LX + E \tag{1}$$

where $\|\cdot\|_*$ denotes the nuclear norm which is the sum of the singular values of the matrix, $\|\cdot\|_1$ is l1-norm, and $\lambda > 0$ is a balance coefficient. $X \in \Re^{N \times M}$ denotes the observed data matrix, $Z \in \Re^{M \times M}$ is a low-rank matrix to separate data and noise, $L \in \Re^{N \times N}$ is the salient coefficients matrix and $E \in \Re^{N \times M}$ is a sparse noisy matrix. Eq.1 can be regarded as a convex optimization problem with a nuclear norm that can be solved by the inexact Augmented Lagrangian Multiplier (ALM) [19] algorithm.

When the LatLRR is used for image decomposition, the image data must be preprocessed first to obtain the observed data matrix where each column of the matrix corresponds to an image patch. Fig.1 shows the process of the single-level LatLRR decomposition of images, where $P(\cdot)$ denotes a two-stage preprocessing operator which divides the input image $I$ into many image patches by a sliding window of size $n \times n$ with an overlap stride $s$. Afterwards, these image patches are reshuffled into the observed data matrix. Stride $s$ means that the window is moved by $s$ pixels in the $x$ and $y$ direction at a time.

As Fig.1 shows, once the projection matrix $L$ is learned by LatLRR, the image $I$ is decomposed into a detail image $I_d$ and a base image $I_b$ by Eq.2 [14].

$$V_d = L \times P(I), \quad I_b = I - I_d$$
$$\text{s.t. } I_d = R(V_d) \tag{2}$$

where $R(\cdot)$ signifies the operator which reconstructs the detail image from the detail part. When restoring $V_d$, the averaging strategy is used to process the overlapping pixel.

Convenient for subsequent description, the operation of the dashed box in Fig.1 is denoted as DLatLRR. The framework of the multi-level version of DLatLRR (MDLatLRR) can be modified as Fig.2 since it is easy to verify that $P(I_b) = P(I) - V_d$. In order to distinguish the MDLatLRR in [14], the procedure of image multi-level LatLRR of images shown by Fig.2 is named as IMDLatLRR. As can be seen from Fig.2, image $I$ is decomposed by $r$-level LatLRR, and output $r$ base images $I_b^k$, $r$ detail images $I_d^k$ and $r$ detailed matrices $V_d^k$, $k = 1, 2, \cdots, r$.

### B. RETINEX-BASED IMAGE ENHANCEMENT
The basic assumption of Retinex theory is that the original image $I(x, y)$ can be decomposed into the product of the illuminance component and the reflection component as follows:

$$I(x, y) = L(x, y) \times R(x, y) \tag{3}$$

where $L(x, y)$ is the illuminance component, which determines the dynamic range of image gray scale transformation and $R(x, y)$ is the reflection component which represents intrinsic features in real-world scenarios.

The purpose of Retinex-based image enhancement is to eliminate the influence of uneven illumination and to represent intrinsic features of the image. According to single-scale Retinex, the reflection component can be obtained as follows:

$$\log R(x, y) = \log I(x, y) - \log L(x, y)$$
$$L(x, y) = G(x, y) * I(x, y) \tag{4}$$

where $*$ is the convolution operation and $G(x, y)$ is a filtering function.

## III. IMAGE ENHANCEMENT AND DECOMPOSITION
Image enhancement can improve the clarity and contrast of images, thus improving the visual effect of fusion results. After an image is decomposed by IMDLatLRR, the detail image is usually clearer and image blurriness caused by low illumination is mainly focused on the base image. In this sense, image enhancement can be conducted on the base image after IMDLatLRR. This means there are two strategies to combine IMDLatLRR decomposition and image enhancement: (1) Enhance the source image and decompose the enhanced image by IMDLatLRR; (2) Decompose the source image by IMDLatLRR and enhance the base image
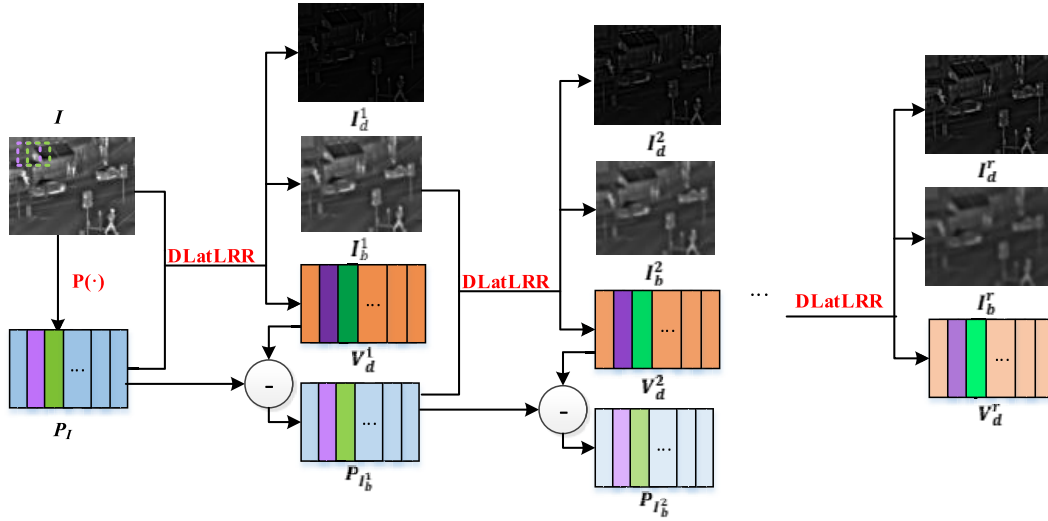
Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint

**IEEE** *Access*



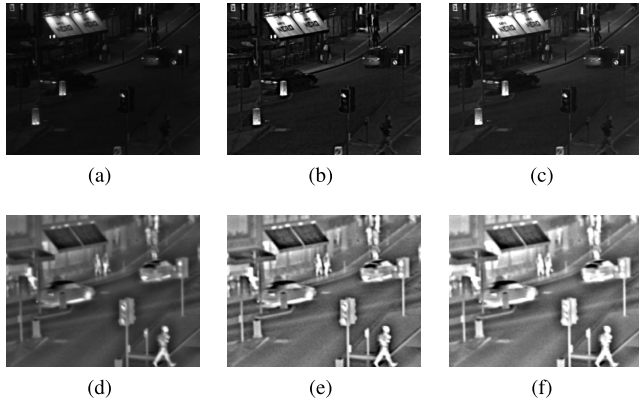**FIGURE 2.** Multi-level LatLRR decomposition of images (IMDLatLRR).



**FIGURE 3.** Result of image reconstruction after different processing. (a) Visible image (b) Reconstructed visible image 1 (c) Reconstructed visible image 2 (d) Infrared image (e) Reconstructed infrared image 1 (f) Reconstructed infrared image 2.

after decomposition. Fig.3 shows the result of image reconstruction after different processes, where Fig.3(b) is reconstructed after Retinex-based image enhancement and then IMDLatLRR, Fig.3(c) is reconstructed after IMDLatLRR and then base image enhancement. By inspecting the results in Fig.3, it is clear that the process order of IMDLatLRR and Retinex-based image enhancement has little effect on the results of image reconstruction. We also did an experiment to compare the fusion result using these two preprocessing methods and the results show that the second strategy is slightly superior. For lack of space, this article focuses on the second strategy.

## IV. PROPOSED FUSION METHOD
### A. THE FRAMEWORK OF THE FUSION METHOD
The framework of our fusion algorithm is presented in Fig.4. Input infrared image $I_{IR}$ and visible image $I_{VIS}$, IMDLatLRR is first carried out for $I_{IR}$ and $I_{VIS}$ respectively to output detail

matrices $V_{dIR}^{1:r}$ and $V_{dVIS}^{1:r}$, detail images $I_{dIR}^{1:r}$ and $I_{dVIS}^{1:r}$, and two base image $I_{bIR}^{r}$ and $I_{bVIS}^{r}$.

In our fusion method, $I_{bIR}^{r}$ and $I_{bVIS}^{r}$ are Retinex enhanced respectively and a weighted average strategy is utilized for the enhanced images $\hat{I}_{bIR}^{r}$ and $\hat{I}_{bVIS}^{r}$ to obtain the fused base image $I_{bf}$.

For each pair of detail matrices, the nuclear-norm based fusion strategy [14] is used to fuse these matrices column by column and reconstructed to obtain the fused detail image $I_{df1}$. Summarize $r$ detail images $I_{dIR}^{1:r}$ to obtain one detail image $I_{dIR}^{r}$, and summarize $r$ detail images $I_{dVIS}^{1:r}$ to obtain one detail image $I_{dVIS}^{r}$. Both $I_{dIR}^{r}$ and $I_{dVIS}^{r}$ contain more obvious structure and features. Then weights based on multiple visual salient features were performed on $I_{dIR}^{r}$ and $I_{dVIS}^{r}$ to obtain the fused detail image $I_{df2}$. The final fused detail image is obtained by $I_{df} = wI_{df1} + (1 - w)I_{df2}$, where $w$ is a constant.

Once the fused detail image and base image are obtained, the fused image is obtained by $I_f = I_{bf} + I_{df}$. In the next subsections, the details of the fusion strategies will be presented.

### B. FUSION OF BASE IMAGES BASED ON RETINEX ENHANCEMENT
Given the two base images $I_{bIR}^{r}$ and $I_{bVIS}^{r}$, firstly calculate the Gaussian mask $G$, then, according to Eq.(4), the enhanced base images $\hat{I}_{bIR}^{r}$ and $\hat{I}_{bVIS}^{r}$ are obtained by

$$\log \hat{I}_{bVIS}^{r} = \log I_{bVIS}^{r} - \log \left(G * I_{bVIS}^{r}\right)$$
$$\log \hat{I}_{bIR}^{r} = \log I_{bIR}^{r} - \log \left(G * I_{bIR}^{r}\right) \tag{5}$$

Because the base image mainly contains the background information of the source image, the simple weighted average fusion is usually used to achieve a better fusion effect. The fused base image is obtained by

$$I_{bf} = w_{bIR}\hat{I}_{bIR}^{r} + w_{bVIS}\hat{I}_{bVIS}^{r} \tag{6}$$

**IEEE** *Access*

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint
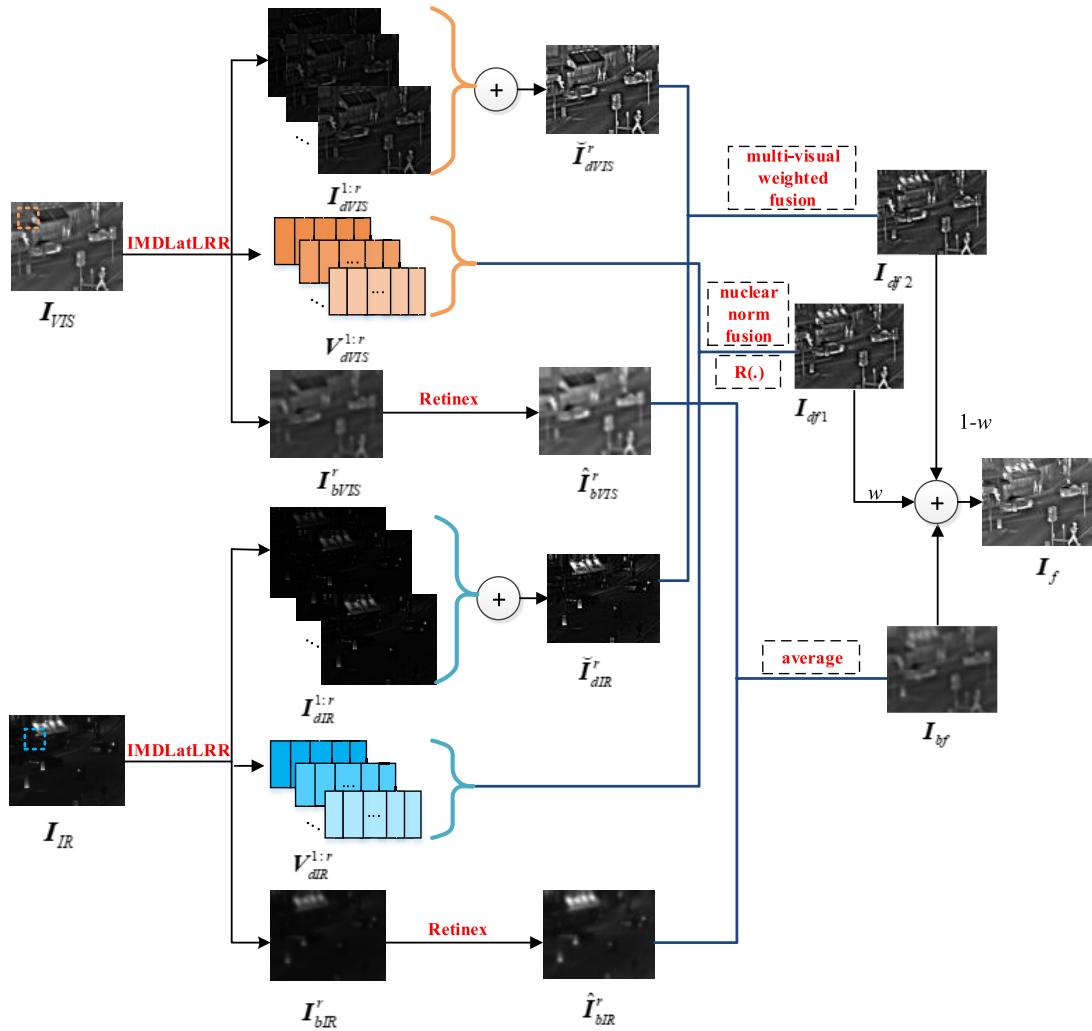


**FIGURE 4.** The framework of the proposed fusion method.

where $w_{bIR}$ and $w_{bVIS}$ denote the corresponding weights of the two enhanced base images.

## C. FUSION OF DETAIL PARTS BASED ON NUCLEAR-NORM AND MULTI-VISUAL WEIGHT INFORMATION

The detail image contains more visual salient features than the base image. The nuclear-norm based fusion strategy for the detail images has been verified effective in [14]. Furthermore, the importance of visual salient features in the detail image fusion has also been verified in [28]–[31]. Better fusion effects can be obtained by simultaneously considering nuclear-norm and multiple visual salient features. Three key indicators that reflect the visual effects of structures, namely clarity, contrast and edge-corners saliency of images, are used to construct multiple visual fusion weights.

### 1) Image clarity

The human eye is more sensitive to image clarity. Several focus measures were studied in [32] as the measures of image

clarity, and the improved Laplace energy sum was verified to show better performance. It can be formulated as

$$\nabla^2_{ML}I(x, y) = |2I(x, y) - I(x - 1, y) - I(x + 1, y)| \\ + |2I(x, y) - I(x, y - 1) - I(x, y + 1)| \quad (7)$$

For the sake of convenience, denote $I_{xy}$ as $I(x, y)$ in the following. The clarity of images is defined as

$$\mathrm{AF}\left(I_{xy}\right) = \sum_{m=-M_1}^{M_1} \sum_{n=-N_1}^{N_1} \nabla^2_{ML}I(x + m, y + n) \quad (8)$$

where $N_1$ and $M_1$ are positive integers.

### 2) Local contrast of image

Compared with the changes of a single pixel, it is easier for human eyes to detect the changes of local pixels. Both local image gradient and local image contrast are key values that assess the spatial details of an image, therefore, the magnitude

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint

**IEEE** *Access*

of local image gradients with the local image contrast are combined to calculate local contrast LC as follows:

$$LC\left(I_{xy}\right) = \beta \left( \max_{I_{pq} \in \aleph(I_{xy})} I_{pq} - \min_{I_{pq} \in \aleph(I_{xy})} I_{pq} \right)$$
$$+ (1 - \beta) \left( \max_{I_{pq} \in \aleph(I_{xy})} \|\nabla I_{pq}\| \right) \quad (9)$$

where $\aleph\left(I_{xy}\right)$ is an $s * s$ neighborhood around $I_{xy}$, $\nabla I_{pq}$ is the spatial gradient of $I_{pq}$, and $\beta$ is a constant.

3) The edge-corner saliency of the image

The structure tensor can reflect the structure and spatial information of the image. Thus, the smooth region, the edge region, and the corner region of the image can be distinguished by using the structure tensor [33].

The linear structure tensor matrix of each position in the image is defined as:

$$J_{hxy} = \begin{bmatrix} G_{xy} & F_{xy} \\ F_{xy} & H_{xy} \end{bmatrix} \quad (10)$$

where $G_{xy} = K_h * \left(\frac{\partial I}{\partial x}\right)^2$, $F_{xy} = K_h * \left(\frac{\partial I}{\partial x}\frac{\partial I}{\partial y}\right)$, $H_{xy} = K_h * \left(\frac{\partial I}{\partial y}\right)^2$, $K_h$ is the Gaussian kernel with standard deviation $h$, and $*$ denotes convolution operation.

Matrix $J_{hxy}$ has two non-negative eigenvalues used to represent spatial information. The calculation formula is as follows:

$$\lambda_{xy1} = \frac{G_{xy} + H_{xy} + \sqrt{\left(G_{xy} - H_{xy}\right)^2 + 4F_{xy}^2}}{2}$$
$$\lambda_{xy2} = \frac{G_{xy} + H_{xy} - \sqrt{\left(G_{xy} - H_{xy}\right)^2 + 4F_{xy}^2}}{2} \quad (11)$$

Construct the edge saliency matrix $M$ and the corner saliency matrix $N$, where

$$M_{xy} = \left|\lambda_{xy1} - \lambda_{xy2}\right|$$
$$N_{xy} = \left|\lambda_{xy1} + \lambda_{xy2}\right| \quad (12)$$

Normalize the matrix $M$ and $N$ to obtain $\widehat{M}$ and $\widehat{N}$ respectively. Then the saliency of the edge-corner is obtained by combining $\widehat{M}$ and $\widehat{N}$ linear as follows:

$$EA(I) = k\widehat{M} + (1 - k)\widehat{N} \quad (13)$$

where $k \in [0, 1]$ is a constant.

4) Multi-visual fusion weight

The fusion weight plays a key role in the image fusion effect. Define the single weight model as follows:

$$N_{AFIRxy}^r = AF\left(\breve{I}_{dIRxy}^r\right) / \left(AF\left(\breve{I}_{dIRxy}^r\right) + AF\left(\breve{I}_{dVISxy}^r\right)\right)$$
$$N_{EAIRxy}^r = EA\left(\breve{I}_{dIRxy}^r\right) / \left(EA\left(\breve{I}_{dIRxy}^r\right) + EA\left(\breve{I}_{dVISxy}^r\right)\right)$$
$$N_{LCIRxy}^r = LC\left(\breve{I}_{dIRxy}^r\right) / \left(LC\left(\breve{I}_{dIRxy}^r\right) + LC\left(\breve{I}_{dVISxy}^r\right)\right)$$
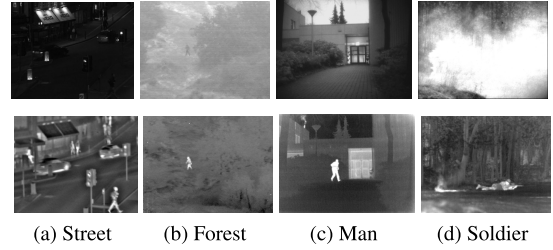$$(14)$$



**FIGURE 5.** Four pairs of source images. The top row contains visible images, and the second row contains infrared images.

Finally, the fusion weights constructed by AF, EA, and LC are defined as:

$$w_{dIRxy}^r = \frac{1}{3}\left(N_{AFIRxy}^r + N_{EAIRxy}^r + N_{LCIRxy}^r\right)$$
$$w_{dVISxy}^r = 1 - w_{dIRxy}^r \quad (15)$$

5) Fusion and reconstruction

The final fused detail image is obtained by

$$I_{df} = wI_{df1} + (1 - w)I_{df2} \quad (16)$$

where $w \in [0, 1]$ is a constant, $I_{df1}$ is the result of fusion and reconstruction for the detail matrices as referred to MDLatLRR [14], and the value of $I_{df2}$ in position $(x, y)$ is

$$I_{df2\ xy} = w_{dVIS}^r(x, y)\breve{I}_{dVISxy}^r + w_{dIR}^r(x, y)\breve{I}_{dIRxy}^r \quad (17)$$

## V. EXPERIMENT RESULT AND ANALYSIS

### A. EXPERIMENTAL SETTING

To verify the advantages of the proposed fusion algorithm, we adopted the dataset from [34]. The test dataset contains 21 aligned infrared and visible image pairs. A sample of these infrared and visible images is shown in Fig.5. The selected four groups of source images have different resolution and the infrared and visible images of the same scene have the same resolution. The scene features of each group of images are different. Among them, the group "Street" is the road scene at night, including vehicles, street lamps, pedestrians, etc. Compared with other groups, the group "Forest" has a lower resolution and relatively singular scene information. The illumination of the "Man" images is uneven and the edge illumination is weak. The visible images of the "Soldier" group have poor visibility due to the smoke.

All the parameters needed in IMDLatLRR are the same as those used in [14]. Namely, a sliding window of size $16 * 16$ with stride equal to 1 is used to divide the source images into image patches, the decomposition level varies between 1 and 4 (meaning $r = 1, 2, 3, 4$), the nuclear norm is used in the fusion of detail matrices, and $w_{bIR} = w_{bVIS} = 0.5$. According to [35], the value of $k$ in Eq.13 is set as 0.5 in our experiment. There are other parameters in the proposed algorithm. A group of experiments are conducted to evaluate the robustness of the parameters used in the algorithm.

In addition, ablation experiment is conducted to analyze the effectiveness and contributions of the Retinex
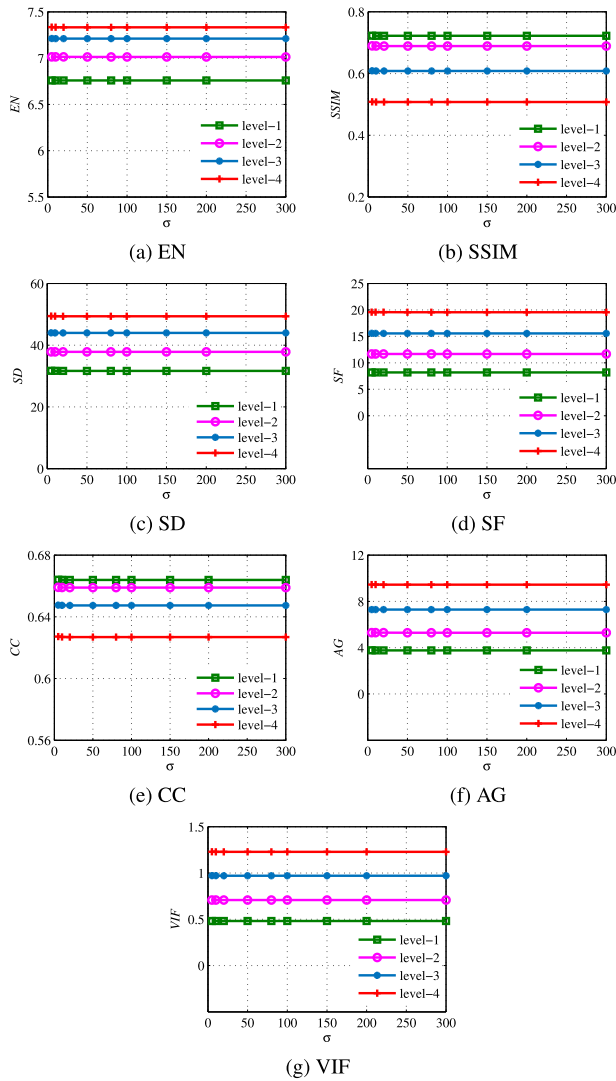
**IEEE** *Access*

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint



**FIGURE 6.** The fusion performance of different projection $\sigma$.



**FIGURE 7.** The fusion performance of different projection $s$.

enhancement of base images and the multi-visual fusion of detail images. Furthermore, 9 recent and classical fusion methods are chosen to conduct the same experiment for comparison purposes. They include the nonsubsampled contourlet transform method (NSCT) [8], the convolutional neural networks based method (CNN) [11], disentangled representation (DRF) [13], MDLatLRR [14], the visual saliency map and weighted least square optimization method (VSM-WLS) [29], convolutional sparse representation method (CSR) [36], the curvelet transform method (CVT) [37], the parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform (NSST-PAPCNN) [38], and the gradient transfer fusion method (GTF) [39].

For the purpose of a quantitative comparison between our method and the other selected comparison methods, 7 quality metrics are utilized. These metrics are entropy(En) [40], structural similarity (SSIM) [41], standard deviation (SD) [42], space frequency(SF) [43], correlation
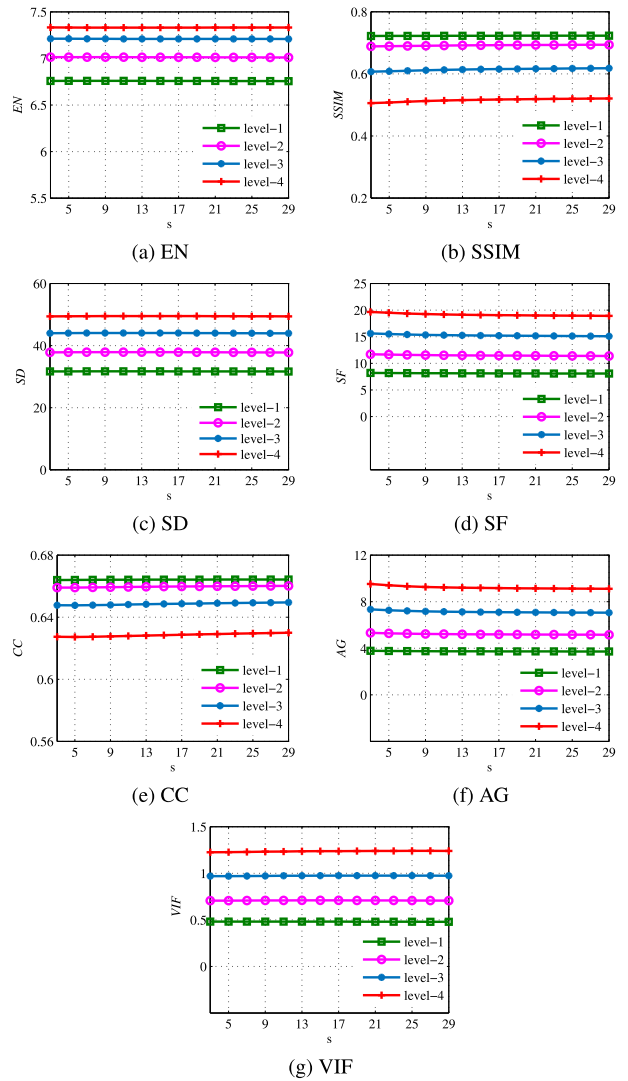
coefficient(CC) [44], average gradient(AG) [45], and visual information fidelity (VIF) [46]. The fusion result is better with the increasing magnitude of these metric values.

All the experiments are implemented in MATLAB R2017b on 3.6 GHz Intel(R) Core(TM) i7-7700 CPU with 16 GB RAM.

### B. VERIFY THE PARAMETER ROBUSTNESS OF THE ALGORITHM

The Gaussian filtering approach is used in Retinex-based enhancement. In addition, there are four parameters that may affect the algorithm's performance, including the scale $\sigma$ used in the Retinex-based enhancement of the base image, the size of neighborhood (denoted as $s$) used to calculate the LC, the size of the Gaussian filter window (denoted as $h$) used to calculate EA, and the fusion weight ($w$) in Eq.16.

Four group experiments are conducted to evaluate the robustness of the four parameters of the algorithm. In the first
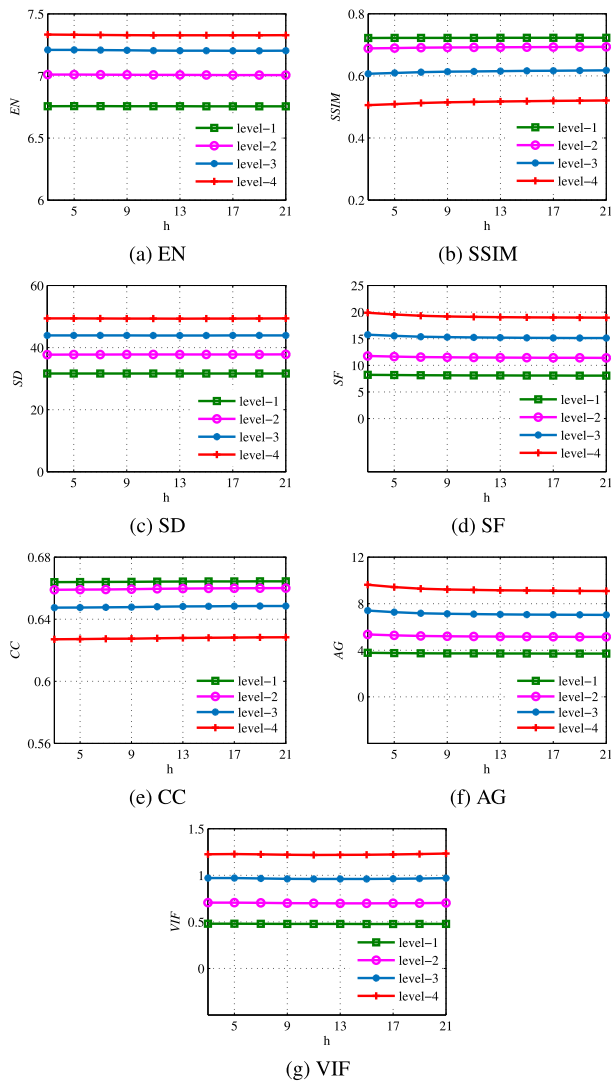
Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint

**IEEE** *Access*



**FIGURE 8.** The fusion performance of different projection *h*.



**FIGURE 9.** The fusion performance of different projection *w*.

experiment, the parameters are fixed as $w = 0.5$, $h = 5$, $s = 5$, and $\sigma$ is set as [5, 10, 20, 50, 80, 100, 150, 200, 300]. In the second experiment, $w = 0.5$, $\sigma = 80$, $s = 5$, and $h$ is set as [3:2:21]. In the third experiment, $w = 0.5$, $\sigma = 80$, $h = 5$, and $s$ is set as [3:2:29]. In the fourth experiment, $h = 5$, $s = 5$, $\sigma = 80$, and $w$ is set as [0:0.1:1]. The four pairs of source images shown in Fig.5 are used in these experiments and the average values of each metric are shown in Figs. 6-9, respectively.

By examining the results in Fig.6, each quality metric is almost a line for all four levels. In fact, the variance for each of the quality metrics of each level is less than 0.001. The results indicate that the influence caused by the value of the Gaussian filter parameter in base image enhancement to fusion results can be almost ignored.

By inspecting the results in Fig.7, except for CC, SSIM, and VIF, other metrics slightly decrease with the increase of $s$, especially for AG and SF when the decomposition
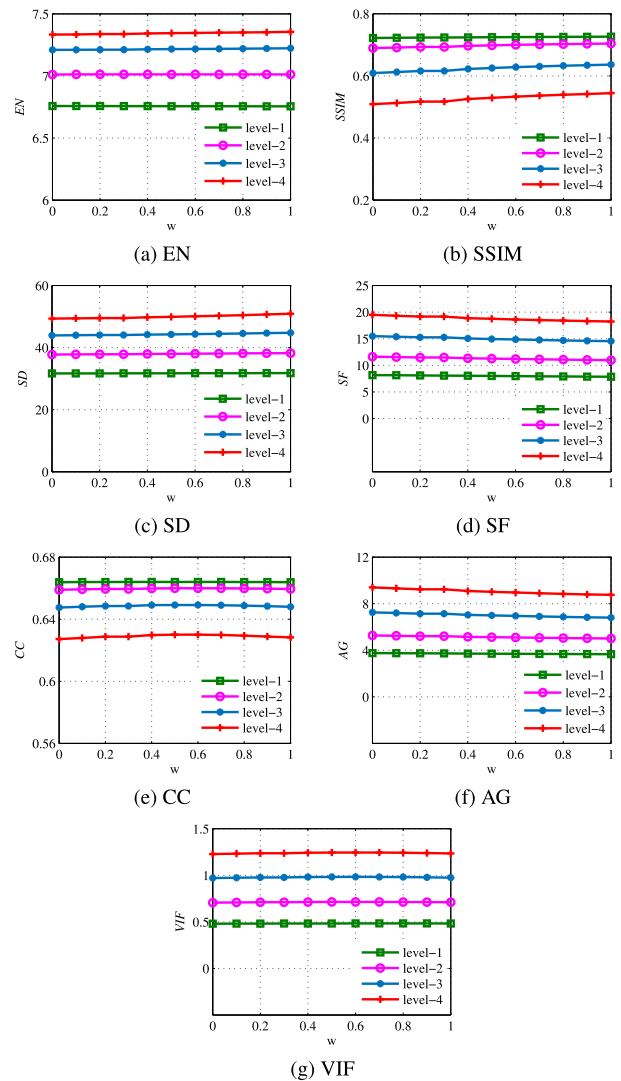
level equals 3 or 4. According to Fig.7, the suitable size of neighborhood used to calculate the LC is $3 * 3$ or $5 * 5$.

The results in Fig.8 also show that, with the exceptions of CC, SSIM, and VIF, other metrics slightly decrease with the increase of $h$, especially for AG and SF when the decomposition level equals 3 or 4. This is because with the increase of $h$, the role of the filter will be weakened. According to Fig.8, the most suitable size of the Gaussian filter window used to calculate EA is 3 or 5.

As shown in Fig.9, with the increase of $w$, SSIM and SD increased and reached the optimal value when $w = 1$, while SF and AG decreased and reached the optimal value when $w = 0$, and the other three indicators are very slightly increased. It shows that $w = 0.5$ is the most appropriate by considering all the indicators comprehensively.

As noted in Figs.6-9, with the exceptions of SSIM and CC, the other metrics increased with the growing number of decomposition levels. This is because the detail information

**IEEE**Access

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint

**TABLE 1.** The result of ablation experiment.

| Methods | Level | EN | SSIM | SD | SF | CC | AG | VIF |
|---|---|---|---|---|---|---|---|---|
| MDLatLRR | level-1 | 6.218 | 0.744 | 23.435 | 6.911 | 0.497 | 3.317 | 0.275 |
| | level-2 | 6.463 | 0.726 | 27.765 | 10.398 | 0.487 | 4.891 | 0.423 |
| | level-3 | 6.766 | 0.663 | 34.255 | 14.766 | 0.470 | 7.110 | 0.640 |
| | level-4 | 6.991 | 0.574 | 41.131 | 19.219 | 0.452 | 9.442 | 0.871 |
| MDLatLRR+Retinex | level-1 | 6.558 | 0.732 | 27.814 | 8.087 | 0.505 | 3.929 | 0.378 |
| | level-2 | 6.810 | 0.702 | 33.406 | 11.669 | 0.503 | 5.560 | 0.551 |
| | level-3 | 7.026 | 0.628 | 39.729 | 15.853 | 0.494 | 7.721 | 0.762 |
| | level-4 | 7.121 | 0.537 | 45.486 | 20.039 | 0.479 | 9.909 | 0.959 |
| Proposed | level-1 | 6.558 | 0.734 | 27.844 | 8.087 | 0.505 | 3.931 | 0.381 |
| | level-2 | 6.813 | 0.707 | 33.506 | 11.669 | 0.503 | 5.580 | 0.561 |
| | level-3 | 7.076 | 0.658 | 39.329 | 15.253 | 0.499 | 7.811 | 0.758 |
| | level-4 | 7.141 | 0.567 | 45.886 | 20.049 | 0.487 | 9.911 | 0.967 |

is not fully extracted by IMDLatLRR with the shallow decomposition level (i.e. level 1), while the base parts contain more source image information, and the simple average weight of the base image ensures that more structural information of the source image is preserved. When the decomposition level increases, the detail parts will contain more base information (e.g. luminance, contour). The fusion strategy of the detail image will cause the change of image structure information. The relation between the variation trend of SSIM and the decomposition order is consistent with the results found in [14].

### C. ABLATION EXPERIMENT
We conduct ablation experiments to explore the effectiveness of the two main components (basic image enhancement and multi-visual weighted fusion) in the proposed method. Then, the proposed method is implemented in two stages. In first stage, only the basic image enhancement is combined with MDLatLRR (MDLatLRR + Retinex), and the multi-visual weighted fusion is then considered in the second stage. In Tab. 1, compared with MDLatLRR at the same level, only SSIM is slightly reduced and all other metrics are obviously improved both in MDLatLRR + Retinex and the proposed method. This indicates that combining the Retinex enhancement of base image and multi-visual fusion of detail images can improve the fusion image contrast, highlight the image features and more according to human vision on the basis of retaining the structure of the source image. Furthermore, compared with MDLatLRR + Retinex, in the proposed method, some metrics are almost no change or very little improved when decomposition level is 1 or 2, and most metrics are improved when level is 3, while all metrics are slightly improved when decomposition level is 4. This indicates that with the increasing number of decomposition levels (between 1 to 4), more contributions are made by the multi-visual fusion of detail images because the detail becomes clearer and the salient features are better enhanced.

### D. ALGORITHM COMPARISON EXPERIMENT
The parameters used in this experiment are set as follows: $\sigma = 80$, $h = 5$, $s = 5$, $w = 0.5$.
  1) Subjective evaluation

The fused results of four pairs of low-quality source images, which are illuminated poorly or with unclear thermal targets, are shown in Figs.10-13. These results are obtained by 9 existing fusion methods and our algorithm. The MDLatLRR uses $L16$, nuclear norm, and the decomposition level is 4. For the sliding window technique, the stride is set to 1 and 4.

Fig.10 shows the fusion results of the source image pair of "Street". With the exception of MDLatLRR and the proposed method, the letters on the lamp signs are not clear enough in other fusion results.

Fig.11 shows the fusion results of the source image pair of "Forest". As we can see, it is obvious that the proposed method obtained the best result. When focusing on the man in the fusion images, it is obvious that GTF get the worst results, the result of VSM-WLS is not clear enough, and there are serious artifacts in the fusion result of algorithm NSST-PAPCNN.

Fig.12 shows the fusion results of the source image pair of "man". Object man are clear in the results of all methods, however, the grass in the red boxes are only clear in the result of MDLatLRR and the proposed method. Focusing on the grass, the three deep learning-based methods, namely CNN, NSST-PAPCNN and DRF, are all failed to obtain clear results, although their results are significant for man areas.

Fig.13 shows the fusion results of the source image pair of "Soldier". From Fig.13 we can see that the three deep learning-based methods (CNN, NSST-PAPCNN and DRF) fail to remove haze from visible images well, and the people lying down could hardly be seen. Moreover, the fusion results of GTF and VSM-WLS are also not clear enough. Focusing on the tree trunk (red boxes), the proposed method is obviously the clearest.

On the other hand, it can be seen from Figs.10-13 that the fused images obtained by MDLatLRR and the proposed method capture more salient features, and for the same method, the results of stride = 1 and stride = 4 are almost indistinguishable. Furthermore, compared with MDLatLRR, the visual quality of the fused images captured by the proposed method is slightly better.
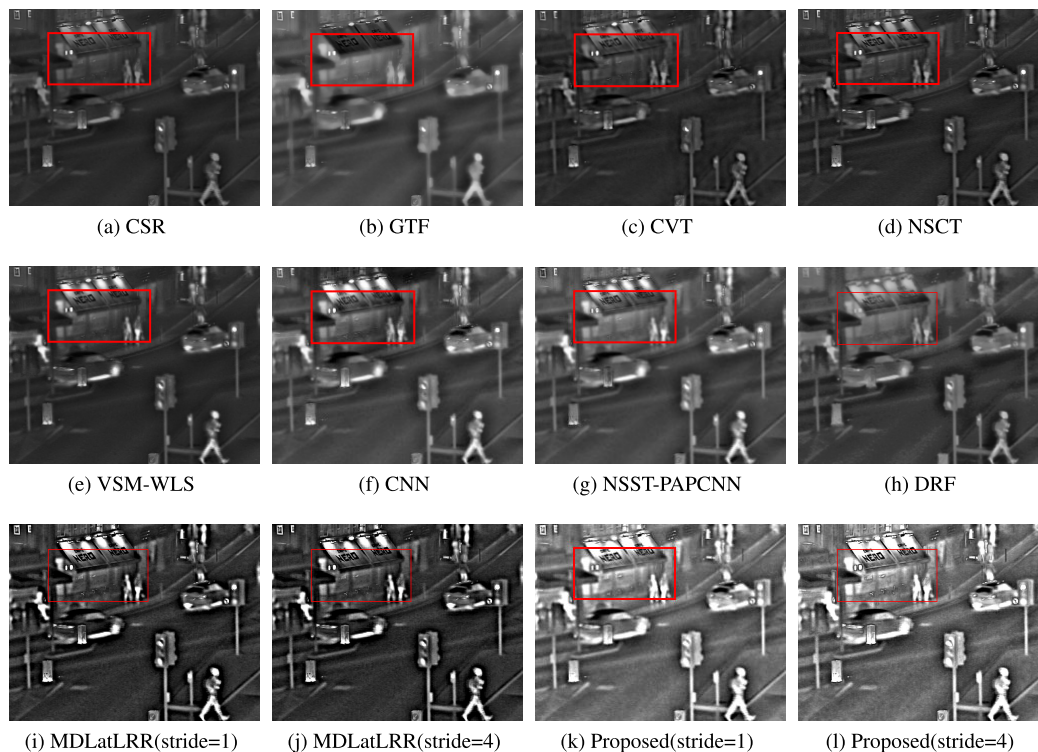  2) Evaluation metrics

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint

**IEEE** *Access*



(a) CSR     (b) GTF     (c) CVT     (d) NSCT

(e) VSM-WLS     (f) CNN     (g) NSST-PAPCNN     (h) DRF

(i) MDLatLRR(stride=1)     (j) MDLatLRR(stride=4)     (k) Proposed(stride=1)     (l) Proposed(stride=4)

**FIGURE 10.** Fusion results of "Street."



(a) CSR     (b) GTF     (c) CVT     (d) NSCT

(e) VSM-WLS     (f) CNN     (g) NSST-PAPCNN     (h) DRF

(i) MDLatLRR(stride=1)     (j) MDLatLRR(stride=4)     (k) Proposed(stride=1)     (l) Proposed(stride=4)
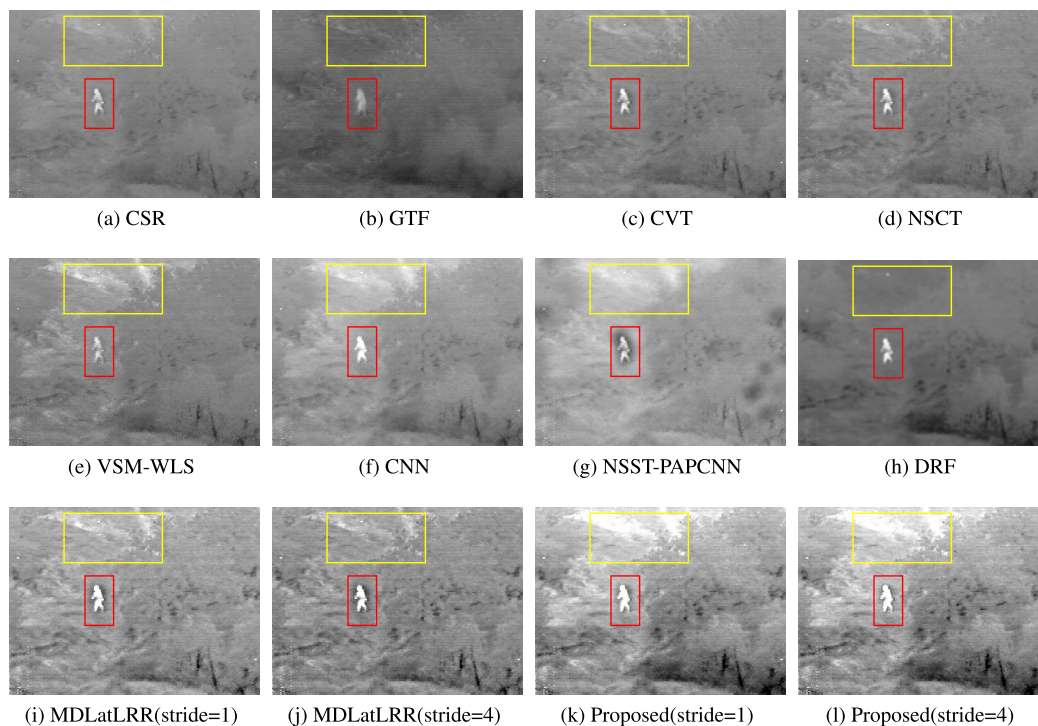
**FIGURE 11.** Fusion results of "Forest."

In order to quantitatively evaluate the performance of fusion results, 7 criteria are used to compare 9 existing fusion method with the proposed method. Tab. 2 summarizes averaged values of the 7 criteria resulting from the set of test
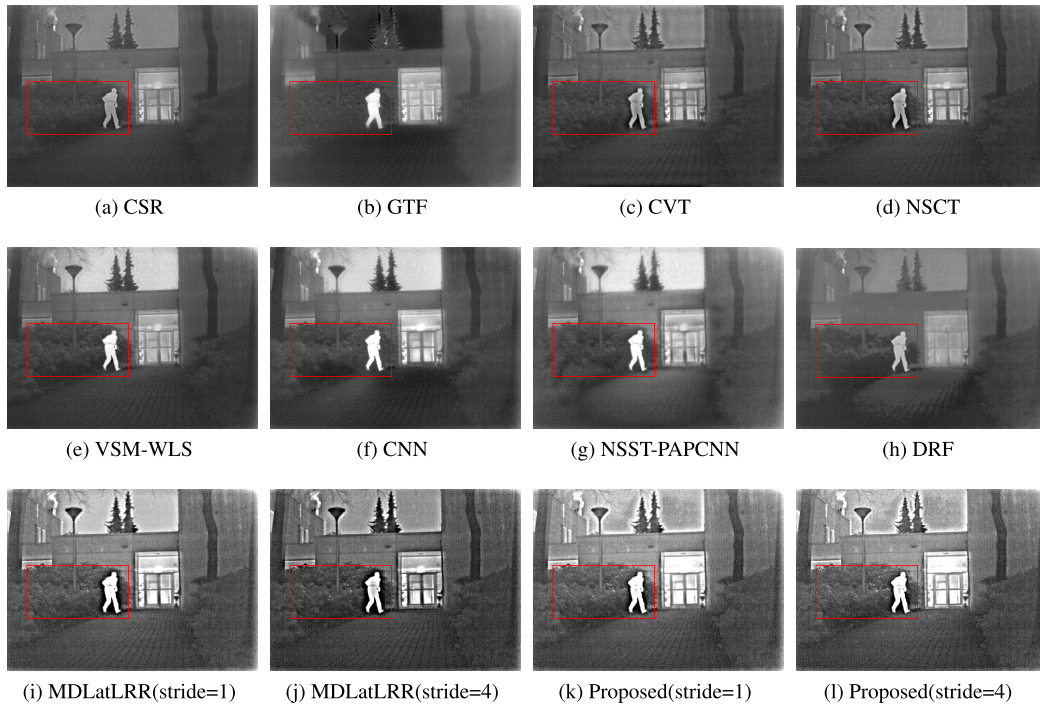
IEEE *Access*

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint



| (a) CSR | (b) GTF | (c) CVT | (d) NSCT |

| (e) VSM-WLS | (f) CNN | (g) NSST-PAPCNN | (h) DRF |

| (i) MDLatLRR(stride=1) | (j) MDLatLRR(stride=4) | (k) Proposed(stride=1) | (l) Proposed(stride=4) |

**FIGURE 12.** Fusion results of "Forest."



| (a) CSR | (b) GTF | (c) CVT | (d) NSCT |

| (e) VSM-WLS | (f) CNN | (g) NSST-PAPCNN | (h) DRF |

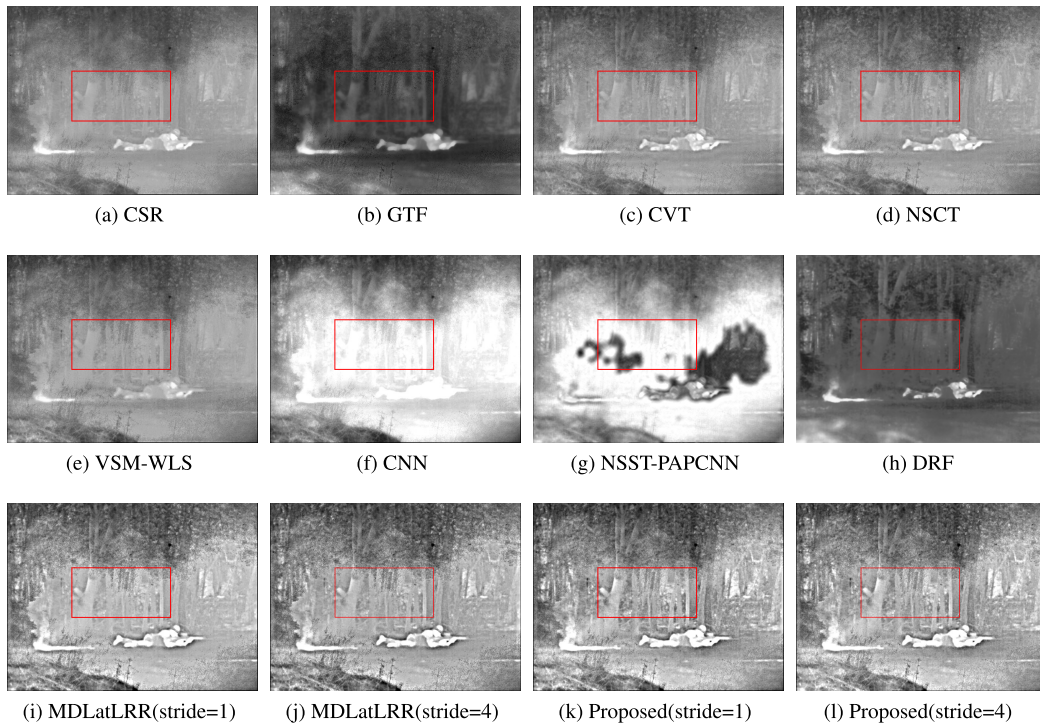| (i) MDLatLRR(stride=1) | (j) MDLatLRR(stride=4) | (k) Proposed(stride=1) | (l) Proposed(stride=4) |

**FIGURE 13.** Fusion results of "Forest."

images. The optimal and suboptimal values are marked in bold red and bold blue font with underline, respectively.

In Tab. 2, the proposed method achieves the best results in 3 criteria (En, SD, and VIF) and second-best values in SF and AG when stride = 1. For stride = 4, the proposed method achieves the best results in SF and AG and second-best values in EN and VIF. Although in CC, our method does not achieve the best value, it obtains a comparable result. The quantitative

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint

IEEE *Access*

**TABLE 2.** The average values of seven metrics for all source images.

| Methods | | EN | SSIM | SD | SF | CC | AG | VIF |
|---|---|---|---|---|---|---|---|---|
| CSR[34] | | 6.259 | **0.727** | 23.929 | 9.170 | <u>0.492</u> | 4.154 | 0.272 |
| GTF[37] | | 6.637 | 0.678 | 31.365 | 9.185 | 0.329 | 4.506 | 0.190 |
| CVT[35] | | 6.431 | 0.690 | 26.046 | 11.094 | 0.474 | 5.459 | 0.316 |
| NSCT[8] | | 6.419 | <u>0.704</u> | 26.114 | 11.124 | 0.482 | 5.442 | 0.358 |
| VSM-WLS[29] | | 6.638 | 0.702 | 33.939 | 12.755 | 0.477 | 6.313 | 0.444 |
| CNN[11] | | 7.062 | 0.686 | <u>45.541</u> | 11.343 | 0.428 | 5.539 | 0.431 |
| NSST-PAPCNN[36] | | 6.869 | 0.667 | 38.080 | 9.594 | 0.408 | 4.466 | 0.302 |
| DRF[13] | | 6.579 | 0.672 | 30.521 | 9.535 | 0.436 | 4.621 | 0.341 |
| MDLatLRR[14] | stride = 1 | 6.991 | 0.574 | 41.131 | 19.219 | 0.452 | 9.442 | 0.871 |
| | stride = 4 | 6.949 | 0.549 | 40.975 | 19.632 | **0.497** | 9.635 | 0.867 |
| Proposed | stride = 1 | **7.141** | 0.567 | **45.886** | <u>20.049</u> | 0.487 | <u>9.911</u> | **0.967** |
| | stride = 4 | <u>7.094</u> | 0.522 | 44.816 | **21.502** | 0.475 | **10.495** | <u>0.942</u> |

**TABLE 3.** Computational time comparison of various methods. (Unit: second).

| Methods | | Image Size | | | | Average Time |
|---|---|---|---|---|---|---|
| | | 632*496 | 360*270 | 620*450 | 768*576 | |
| CSR[34] | | 229.922 | 47.589 | 158.873 | 245.801 | 170.546 |
| GTF[37] | | 8.060 | 0.827 | 6.272 | 6.932 | 5.523 |
| CVT[35] | | 3.052 | 1.078 | 1.730 | 2.475 | 2.084 |
| NSCT[8] | | 5.538 | 1.535 | 4.460 | 6.780 | 4.578 |
| VSM-WLS[29] | | 5.423 | 0.608 | 2.797 | 4.710 | 3.385 |
| CNN[11] | | 71.164 | 20.296 | 66.231 | 99.847 | 64.385 |
| NSST-PAPCNN[36] | | 63.703 | 11.154 | 50.760 | 91.102 | 54.180 |
| DRF[13] | | 19.351 | 18.643 | 13.355 | 22.386 | 18.434 |
| MDLatLRR[14] | stride = 1 | 182.343 | 54.976 | 157.651 | 248.869 | 160.960 |
| | stride = 4 | 11.114 | 3.362 | 9.544 | 16.627 | 10.162 |
| Proposed | stride = 1 | 193.800 | 57.090 | 163.410 | 265.328 | 169.907 |
| | stride = 4 | 43.237 | 15.526 | 41.416 | 58.644 | 39.706 |

results show that the proposed fusion method delivers better fusion performance than the compared methods. Tab. 2 also shows that there is very little difference in the results of MDLatLRR under stride = 1 and stride = 4.

3) Computational cost

Tab. 3 depicts the average running time of four image with different size for different methods in this paper. As can be seen from Tab. 3, the GTF, CVT, NSCT and VSM-WLS methods have short running time, but their performances are quite ordinary. The CNN, NSST-PAPCNN and DRF algorithms are based on deep learning models without training. The corresponding testing phase does not require much time, the running time of the algorithm is relatively short, and the image fusion effect is limited. Compared with MDLatLRR, the proposed method takes more time because it has two more steps, namely, base image enhancement and the calculation of multi-visual weights.

Both MDLatLRR and the proposed method are image decomposition methods based on representation learning. When stride = 1, the running time of the proposed method is similar to that of the sparse representation algorithm CSR. Compared with other methods, the proposed method did not yield good performance in running time, but its fusion performance is the best. For scenarios where time is preferred, stride = 4 can be considered in the application, and the fusion performance is slightly inferior to that with stride = 1. Therefore, compared with other methods, the running time of the proposed algorithm is still competitive. The running time

of the proposed method may be reduced to some extent by using efficient programming languages like Python and C.

## VI. CONCLUSION
In this paper, we developed a new MDLatLRR-based fusion framework for fusing infrared and visible images by jointing Retinex-based enhancement and multi visual weights. Firstly, the infrared and visible images were decomposed by multi-level LatLRR to extract detail parts and base parts of the input images at several representation levels. The clarity, local contrast and edge and corner significance of the image were calculated to construct the visual weights for deal parts fusion, the final fusion results of the detail images are weighted averaged by nuclear norm fusion and multiple visual weights fusion. Retinex-based enhancements were conducted on the base parts before fusion the base image using the averaging strategy. The proposed method was evaluated both subjectively and objectively in a number of experiments. The experimental results demonstrate that compared to MDLatLRR, the effectiveness of image enhancement of base image is obvious, and with the increasing number of decomposition levels, the multi-visual fusion of detail images contributes more to the improvement of the fusion performance. The performance of the proposed method is superior to that of the compared methods. However, the proposed method also has its limitations, such as the difficulty of determining the combination of multi-visual weight coefficients and the insufficient investigation of the

selection of Retinex enhancement filters. In the future, we will conduct further research on the following aspects: 1) Try to use other metrics of multi-visual information for image fusion, such as edge information using fractional order differentiation based fusion metrics [47]. 2) Try to use optimized methods to determine the combination of different multi-visual weight coefficients. 3) To investigate the effect to the fusion result caused by different filters in the Retinex enhancement algorithm. 4) Develop fast fusion scheme for color images fusion based on MDLatLRR and Retinex enhancement.

## ACKNOWLEDGMENT

## REFERENCES

[1] J. Ma, Y. Ma, and C. Li, "Infrared and visible image fusion methods and applications: A survey," *Inf. Fusion*, vol. 45, pp. 153–178, Jan. 2019.

[2] Y. Shen, C.-H. Huang, F. Huang, J. Li, M.-J. Zhu, and S. Wang, "Infrared and visible image fusion: Review of key technologies," *Infr. Laser Eng.*, vol. 45, pp. 1–16, Mar. 2021.

[3] A. Seal, D. Bhattacharjee, M. Nasipuri, C. Gonzalo-Martin, and E. Menasalvas, "Fusion of visible and thermal images using a directed search method for face recognition," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 4, pp. 218–233, 2017.

[4] Z.-G. Jiang, D.-B. Han, J. Chen, and X.-K. Zhou, "A wavelet based algorithm for multi-focus micro-image fusion," in *Proc. ICIG*, Hong Kong, Dec. 2004, pp. 176–179.

[5] A. Seal, D. Bhattacharjee, M. Nasipuri, C. Gonzalo-Martin, and E. Menasalvas, "À-trous wavelet transform-based hybrid image fusion for face recognition using region classifiers," *Expert Syst.*, vol. 35, no. 6, Jul. 2018, Art. no. e12307.

[6] C. Deng, S. Wang, and X. Chen, "Remote sensing images fusion algorithm based on shearlet transform," in *Proc. Int. Conf. Environ. Sci. Inf. Appl. Technol.*, Wuhan, China, Jul. 2009, pp. 451–454.

[7] R. Srivastava, O. Prakash, and A. Khare, "Local energy-based multimodal medical image fusion in curvelet domain," *IET Comput. Vis.*, vol. 10, no. 6, pp. 513–527, Sep. 2016.

[8] Y.-Z. Sun, H.-F. Li, and B.-S. Li, "The nonsampled Contourlet transform of multi-scale image fusion algorithm," *J. Comput. Simul.*, vol. 28, no. 8, pp. 250–255, 2011.

[9] G. S. El-Taweel and A. K. Helmy, "Image fusion scheme based on modified dual pulse coupled neural network," *IET Image Process.*, vol. 7, no. 5, pp. 407–414, Jul. 2013.

[10] X. Huang, G. Qi, H. Wei, Y. Chai, and J. Sim, "A novel infrared and visible image information fusion method based on phase congruency and image entropy," *Entropy*, vol. 21, no. 12, p. 1135, Nov. 2019.

[11] Y. Liu, X. Chen, J. Cheng, H. Peng, and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," *Int. J. Wavelets, Multiresolution Inf. Process.*, vol. 16, no. 3, Nov. 2018, Art. no. 1850018.

[12] H. Li and X.-J. Wu, "DenseFuse: A fusion approach to infrared and visible images," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2614–2623, Dec. 2018.

[13] H. Xu, X. Wang, and J. Ma, "DRF: Disentangled representation for visible and infrared image fusion," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.

[14] H. Li, X.-J. Wu, and J. Kittler, "MDLatLRR: A novel decomposition method for infrared and visible image fusion," *IEEE Trans. Image Process.*, vol. 29, pp. 4733–4746, 2020.

[15] Z. Gao and C. Zhang, "Texture clear multi-modal image fusion with joint sparsity model," *Optik*, vol. 130, pp. 255–265, Feb. 2017.

[16] Q. Zhang, Y. Liu, R. S. Blum, J. Han, and D. Tao, "Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review," *Inf. Fusion*, vol. 40, pp. 57–75, Mar. 2018.

[17] M. Yin, P. Duan, W. Liu, and X. Liang, "A novel infrared and visible image fusion algorithm based on shift-invariant dual-tree complex shearlet transform and sparse representation," *Neurocomputing*, vol. 226, pp. 182–191, Feb. 2017.

[18] Z. Zhu, H. Yin, Y. Chai, Y. Li, and G. Qi, "A novel multi-modality image fusion method based on image decomposition and sparse representation," *Inf. Sci.*, vol. 432, pp. 516–529, Mar. 2018.

[19] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 1615–1622.

[20] H. Li and X.-J. Wu, "Infrared and visible image fusion using latent low-rank representation," 2018, *arXiv:1804.08992*.

[21] S. Yu and X. Chen, "Infrared and visible image fusion based on a latent low-rank representation nested with multiscale geometric transform," *IEEE Access*, vol. 8, pp. 110214–110226, 2020.

[22] E. Land, "An alternative technique for the computation of the designator in the retinex theory of color vision," *Proc. Nat. Acad. Sci. USA*, vol. 83, no. 10, pp. 3078–3080, May 1986.

[23] G. Qin, "The research on a new image enhancement algorithm based on Retinex theory," in *Advanced Research on Computer Education, Simulation and Modeling*, vol. 167. Berlin, Germany: Springer, Jun. 2011, pp. 336–342.

[24] S. Feng and L. Bo, "Low-light color image enhancement based on retinex," in *Proc. 5th Int. Conf. Automat., Control Robot. Eng. (CACRE)*, Dalian, China, Sep. 2020, pp. 570–574.

[25] P. Hao, S. Wang, S. Li, and M. Yang, "Low-light image enhancement based on retinex and saliency theories," in *Proc. Chin. Automat. Congr. (CAC)*, Hangzhou, China, Nov. 2019, pp. 2594–2597.

[26] J. Li, S.-J. Li, X.-H. Duan, Y. Yao, J.-Y. Li, and L.-Z. Wang, "Infrared image enhancement based on Retinex and probability nonlocal means filtering," *Acta Photonica Sinica*, vol. 49, no. 4, p. 0410003, Apr. 2020.

[27] Z.-T. Jiang, Q. Jiang, Y.-S. Huang, and S.-Q. Zhang, "Infrared and low-light-level visible light enhancement image fusion method based on latent low-rank representation and composite filtering," *Acta Photonica Sinica*, vol. 49, no. 4, Apr. 2020, Art. no. 410001.

[28] Y. Yang, Y. Zhang, S. Huang, Y. Zuo, and J. Sun, "Infrared and visible image fusion using visual saliency sparse representation and detail injection model," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–15, 2020.

[29] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infr. Phys. Technol.*, vol. 82, pp. 8–17, May 2017.

[30] S.-H. Liu, X.-F. Li, and X.-L. Zhang, "Remote sensing image fusion algorithm based on mutual-structure for joint filtering using saliency detection," *J. Electron. Imag.*, vol. 28, no. 3, p. 033007, May 2019.

[31] S. Zhang, X. Li, X. Zhang, and S. Zhang, "Infrared and visible image fusion based on saliency detection and two-scale transform decomposition," *Infr. Phys. Technol.*, vol. 114, May 2021, Art. no. 103626.

[32] W. Huang and Z.-L. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern Recognit. Lett.*, vol. 28, no. 4, pp. 493–500, Mar. 2007.

[33] J.-H. Qu, Y.-S. Li, and W.-Q. Dong, "Fusion of hyperspectral and panchromatic images using structure tensor," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Valencia, Spain, Jul. 2018, pp. 7216–7219.

[34] (2014). *TNO Image Fusion Dataset*. [Online]. Available: https://figshare.com/articles/TN_Image_Fusion_Dataset/1008029

[35] Z.-Y. He, C. Wang, L.-N. Sun, B.-Q. Zuo, G.-R. Weng, and J.-J. Hu, "A method for calculating visual saliency map," CN Patent CN 201 510 116 370.6, Jun. 10, 2015.

[36] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image fusion with convolutional sparse representation," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1882–1886, Dec. 2016.

[37] F. Nencini, A. Garzelli, S. Baronti, and L. Alparone, "Remote sensing image fusion using the curvelet transform," *Inf. Fusion*, vol. 8, no. 2, pp. 143–156, Jun. 2007.

[38] M. Yin, X. Liu, Y. Liu, and X. Chen, "Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain," *IEEE Trans. Instrum. Meas.*, vol. 68, no. 1, pp. 49–64, Jan. 2019.

[39] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, Sep. 2016.

[40] W. J. Roberts, J. A. A. Van, and F. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *J. Appl. Remote Sens.*, vol. 2, no. 1, pp. 1–28, May 2008.

Q. Pan *et al.*: Fusion of Low-Quality Visible and Infrared Images Based on Multi-Level Latent Low-Rank Representation Joint

**IEEE** *Access*

[41] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[42] Y.-J. Rao, "In-fibre Bragg grating sensors," *Meas. Sci. Technol.*, vol. 8, no. 4, p. 355, 1997.

[43] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959–2965, Dec. 1995.

[44] M. Deshmukh and U. Bhosale, "Image fusion and image quality assessment of fused images," *Int. J. Image Process.*, vol. 4, no. 5, pp. 484–508, Dec. 2010.

[45] G. Cui, H. Feng, Z. Xu, Q. Li, and Y. Chen, "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition," *Opt. Commun.*, vol. 341, pp. 199–209, Apr. 2015.

[46] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Inf. Fusion*, vol. 14, pp. 127–135, Apr. 2013.

[47] A. Sengupta, A. Seal, C. Panigrahy, O. Krejcar, and A. Yazidi, "Edge information based image fusion metrics using fractional order differentiation and sigmoidal functions," *IEEE Access*, vol. 8, pp. 88385–88398, 2020.
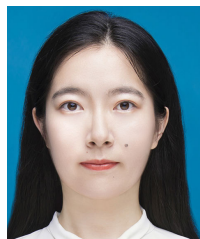
**LIAOYING ZHAO** received the B.S. and M.S. degrees from Hangzhou Dianzi University, Hangzhou, China, in 1992 and 1995, respectively, and the Ph.D. degree from Zhejiang University, Hangzhou, in 2004. Since 1995, she has been with Hangzhou Dianzi University, where she is currently a Professor with the College of Computer Science. Her research interests include hyperspectral image processing, signal and image processing, pattern recognition, and machine learning.
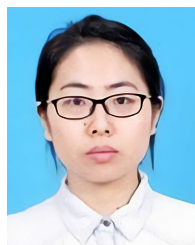
**SHUHAN CHEN** (Member, IEEE) received the B.S. degree from Ludong University, Yantai, China, in 2011, the M.S. degree from Liaoning Technical University, Huludao, China, in 2014, and the Ph.D. degree from Zhejiang University, Hangzhou, China, in 2020. She was a Visiting Faculty Research Assistant at the Remote Sensing Signal and Image Processing Laboratory (RSSIPL), Department of Computer Science and Electrical Engineering, University Maryland, Baltimore County (UMBC), Baltimore, MD, USA, from 2018 to 2020. She is currently a Postdoctoral Researcher with the Department of Electrical Engineering, Zhejiang University. Her research interests include image registration, hyperspectral anomaly detection, hyperspectral target detection, and pattern recognition.

**QIAOYING PAN** received the B.S. degree in computer science and technology from Huzhou University, Huzhou, China, in 2019. She is currently pursuing the M.S. degrees in computer technology with Hangzhou Dianzi University, Hangzhou, China. Her research interests include image processing and image fusion.

**XIAORUN LI** received the B.S. degree from the National University of Defense Technology, Changsha, China, in 1992, and the M.S. and Ph.D. degrees from Zhejiang University, Hangzhou, China, in 1995 and 2008, respectively. Since 1995, he has been with Zhejiang University, where he is currently a Professor with the College of Electrical Engineering. His research interests include hyperspectral image processing, signal and image processing, and pattern recognition. His research interests include image processing and image registration.

• • •