frontiers | Frontiers in Physics

TYPE Original Research
PUBLISHED 08 July 2025
DOI 10.3389/fphy.2025.1514476

OPEN ACCESS

EDITED BY
Xinzhong Li,
Henan University of Science and Technology, China

REVIEWED BY
Ruimei Zhang,
Sichuan University, China
Peter Yuen,
Cranfield University, United Kingdom

*CORRESPONDENCE
Bo Ye,
boye@kust.edu.cn

RECEIVED 21 October 2024
ACCEPTED 15 May 2025
PUBLISHED 08 July 2025

CITATION
Zou Y, Wu J, Ye B, Cao H, Feng J, Wan Z and Yin S (2025) Infrared and visible image fusion based on multi-scale transform and sparse low-rank representation.
Front. Phys. 13:1514476.
doi: 10.3389/fphy.2025.1514476

Infrared and visible image fusion based on multi-scale transform and sparse low-rank representation

Yangkun Zou[1,2], Jiande Wu[1,3], Bo Ye[4,5]*, Honggui Cao[4,5], Jiqi Feng[6], Zijie Wan[4,5] and Shaoda Yin[4,5]

[1]School of Information Science and Engineering, Yunnan University, Kunming, China, [2]Faculty of Civil Aviation and Aeronautics, Kunming University of Science and Technology, Kunming, China, [3]Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University, Kunming, China, [4]Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming, China, [5]Yunnan Key Laboratory of Intelligent Control and Application, Kunming University of Science and Technology, Kunming, China, [6]Guangxi Huasheng new material Co., Ltd., Fangchenggang, China

Infrared and visible image sensors are wildly used and show strong complementary properties, the fusion of infrared and visible images can adapt to a wider range of applications. In order to improve the fusion of infrared and visible images, a novel and effective fusion method is proposed based on multi-scale transform and sparse low-rank representation in this paper. Visible and infrared images are first decomposed to obtain their low-pass and high-pass bands by Laplacian pyramid (LP). Second, low-pass bands are represented with some sparse and low-rank coefficients. In order to improve the computational efficiency and learn a universal dictionary, low-pass bands are separated into several image patches using a sliding window prior to sparse and low rank representation. The low-pass and high-pass bands are then fused by particular fusion rules. The max-absolute rule is used to fuse the high-pass bands, and max-L1 norm rule is utilized to fuse the low-pass bands. Finally, an inverse LP is performed to acquire the fused image. We conduct experiments on three datasets and use 13 metrics to thoroughly and impartially validate our method. The results demonstrate that the proposed fusion framework can effectively preserve the characteristics of source images, and exhibits superior stability across various image pairs and metrics.

KEYWORDS

image fusion, multi-scale transform, sparse representation, low-rank representation, infrared image, visible image

1 Introduction

Due to the use of new application scenarios and the increasing demand for image sensors in fields such as transportation, security, and military, single image sensors are becoming less effective. Because source images from multiple sensors contain more detailed information and can meet more complex requirements, it is more effective to collect information about a particular situation using various image sensors. Infrared and visible image sensors are wildly used and show strong complementary properties. Visible image sensors have high spatial resolution, rich details, and contrast between light and dark, but

10.3389/fphy.2025.1514476

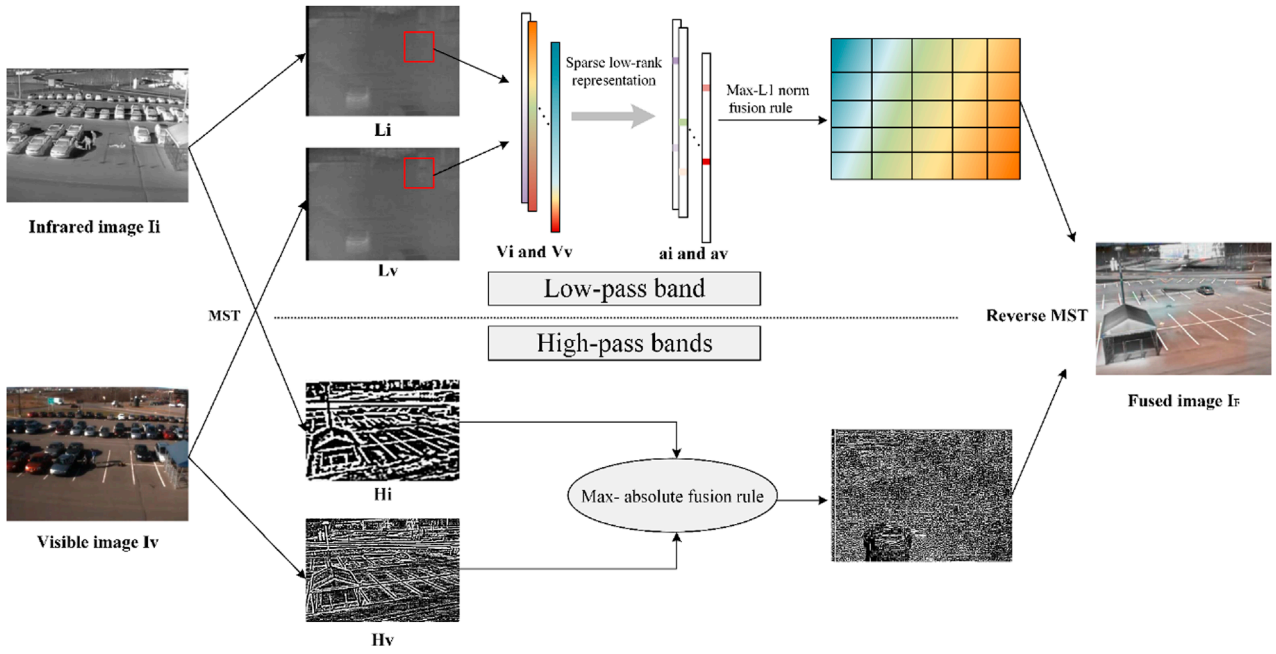

FIGURE 1
The schematic diagram of the proposed fusion method.

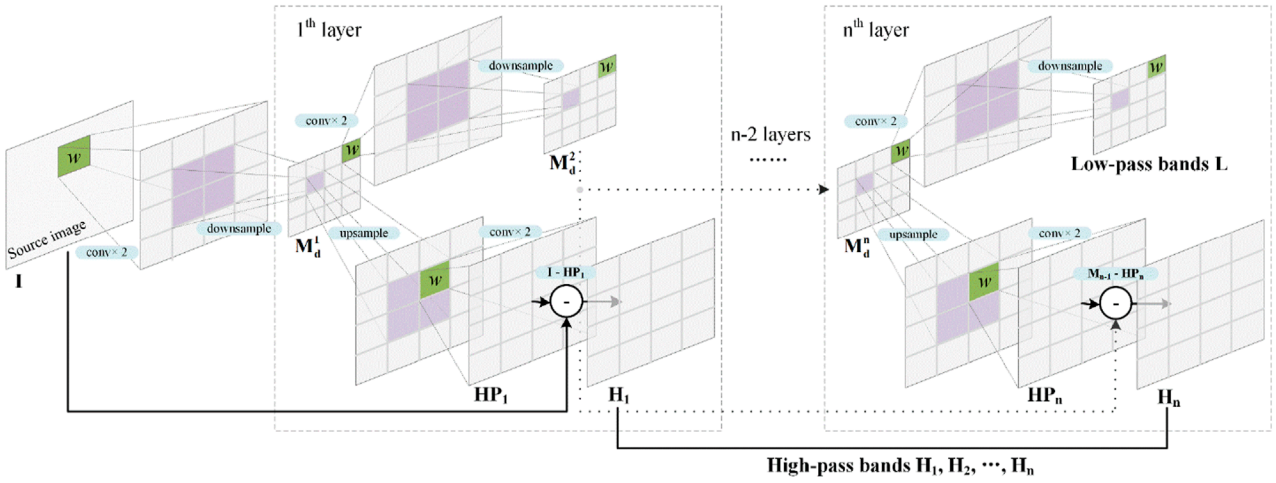

FIGURE 2
Diagram of image decomposition.

they are susceptible to adverse environments like low lighting and fog. Infrared image sensors are less affected by environments but have low resolution and poor texture. Therefore, the combination of infrared and visible sensors can adapt to a wider range of applications.

Visible and infrared image fusion has attracted considerable attention. So far, fusion methods can be divided into conventional methods and deep learning-based methods on whether deep learning is utilized [1]. Conventional methods include spatial domain-based methods and transform domain-based methods. Spatial domain-based methods fuse images directly on pixels or image patches/blocks. However, these methods are susceptible to noise and it is easy to introduce artifacts. Zhang [2] proposed an effective infrared and visual image fusion algorithm through infrared feature extraction and visual information preservation, while Xiao [3] presented a spatial domain-based image fusion method based on fourth order partial differential equations and principal component analysis, Ma [4] proposed a novel fusion algorithm based on gradient transfer and total variation minimization; the gradient calculation is related to a specific matric. Transform domain-based methods firstly transform source images into another domain, and then fusion rules are used to fuse the

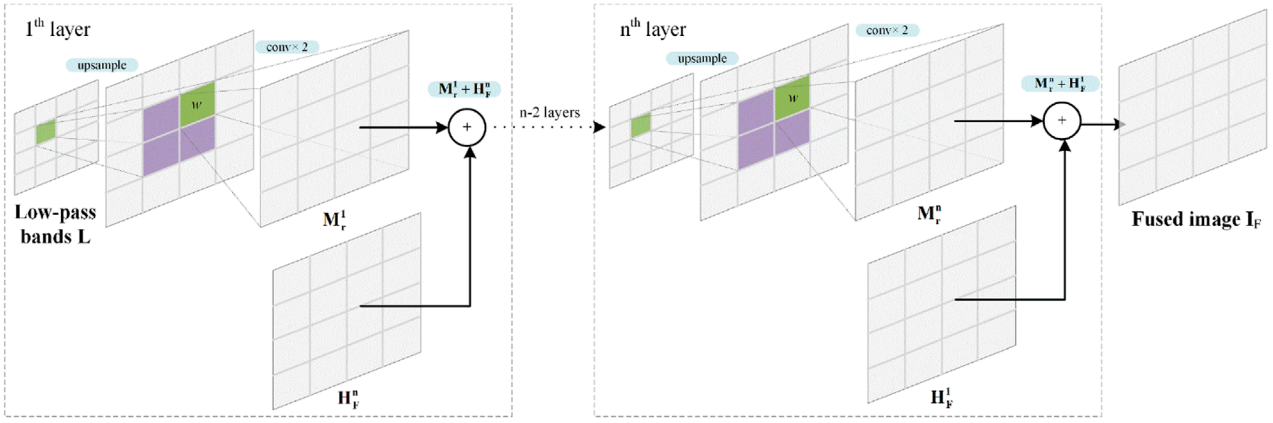

FIGURE 3
Diagram of fused image reconstruction.

TABLE 1 The 13 metrics and their meaning in the following experiments. "↓" means that smaller value denotes better results.

Category	Metrics	Meaning	Category	Metrics	Meaning
Information theory-based metrics	CE↓	Cross entropy	Image feature-based metrics	AG	Average gradient
	EN	Entropy		SD	Standard deviation
	MI	Mutual information		EI	Edge intensity
	PSNR	Peak signal-to-noise ratio		$Q^{AB/F}$	Edge based similarity measurement
				Spatial frequency	
Structural similarity-based metrics	RMSE↓	Root mean squared error	Human visual perception	Q^{CB}	Chen-Blum metric
	SSIM	Structural similarity index measure		Q^{CV}↓	Chen-Varshney metric

source images. After the fusion is completed, the fused results are transformed back to image form. Many good results have been achieved with the transform domain-based method. Li [5] proposed a latent low-rank representation image fusion method that received very good results although it is time-intensive. Bavirisetti [6] achieved image fusion based on saliency detection and two scale image decomposition; they also studied anisotropic diffusion and Karhunen - Loeve transform for image fusion [7]. Kumar [8] used a cross-bilateral filter to extract the image details and achieved image fusion based on weighted average. Zhou [9] and Li [10] both studied the guided filter to achieve image fusion. Bavirisetti [11] used multi-scale image decomposition, structure transferring property, visual saliency detection, and weight map construction to fuse images. Naidu [12] studied multi-resolution singular value decomposition to fuse images. Qi [13] proposed a saliency-based decomposition strategy for infrared and visible image fusion. Considering the advantages in image information capture, some researchers have studied many hybrid methods by combining domain transform and some emerging image information processing algorithms. Zhou [14] studied a hybrid image fusion method through a hybrid multi-scale decomposition with Gaussian and bilateral filters. Liu [15] studied the effectiveness of image fusion by combing different transform methods and sparse representation; the results demonstrated that the combination of multi-scale transform (MST) and sparse representation achieved better results. Ma [16] used a rolling guidance filter and Gaussian filter to decompose source images, and an improved visual saliency map was proposed to fuse images. In order to eliminate the modal differences between infrared and visible images, Chen [17] utilized feature-based decomposition and domain normalization to improve the image fusion quality. Li [18] used rolling guidance filtering and a gradient saliency map to address the issue of brightness and detail information loss in infrared and visible image fusion.

As artificial intelligence continues to develop quickly, many deep learning methods are being examined for their use in visible and infrared image fusion. Liu [19] used convolution neural networks and Gaussian pyramid decomposition to compute a wight map—the wight map was used to fuse the Laplacian pyramid decomposed source images. Li [20] used a deep learning network to extract multi-layer features of detailed parts of source images and fused images

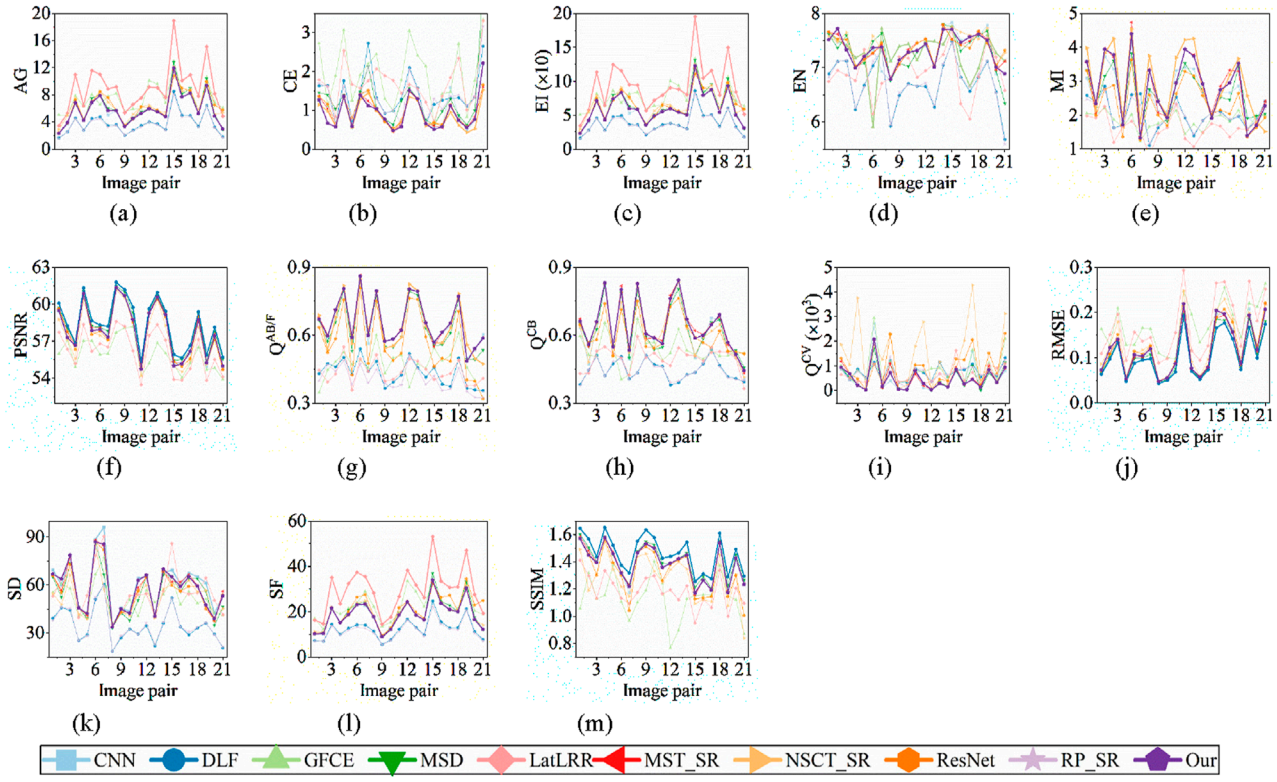

FIGURE 4
Metrics of all 21 image pairs. The data of the first 9 methods are from the workbench in reference [22]. The best result and our result are bolded, and there is one bolded result if our result is the best. **(a)** AG **(b)** CE **(c)** EI **(d)** EN **(e)** MI **(f)** PSNR **(g)** $Q^{AB/F}$ **(h)** Q^{CB} **(i)** Q^{CV} **(j)** RMSE **(k)** SD **(l)** SF **(m)** SSIM.

by l1-norm and weighted-average strategy. Li [21] also studied the performance of ResNet and zero-phase component analysis. Zhang [1] comprehensively reviewed the current deep learning-based image fusion algorithms and established a benchmark. He also compared learning methods and non-learning methods and concluded that the performances of deep learning-based image fusion algorithms do not show superiority over non-learning algorithms [22]. Wei [23] proposed an attention-based dual-branch feature decomposition fusion network.

As discussed above, sparse representation has achieved good results in visible and infrared image fusion. Recent studies demonstrate that combining sparse and low-rank constraints can enhance the information capture of images [24,25], because low-rank constraints can recover some structural information of source images. In this paper, we aim at establishing a new visible and infrared image fusion method using sparse and low-rank representation. Considering the good performance of Laplacian pyramid (LP), source images are firstly transformed into another domain to obtain their low-pass and high-pass bands. Secondly, low-pass bands are represented as some sparse and low-rank coefficients. Then, specific fusion rules are adopted to fuse the low-pass and high-pass bands. Finally, an inverse LP is conducted to obtain the fused image.

The contributions of this paper include 1) combining sparse constraint and low-rank constraint to establish a sparse low-rank representation model to improve the performance of extracting the complex structural features; 2) precisely designing a solution strategy for the sparse low-rank representation model, which is essential for image fusion with high qualities; and 3) establishing a new infrared and visible image fusion method based on Laplacian pyramid and sparse low-rank representation, the experimental results of which proved the advantages in fusion quality and improved runtime performance compared to methods with similar fusion quality.

The rest of this paper is organized as follows. Section 2 first presents the sparse low-rank representation model. Our fusion method is presented in Section 3. In Section 4, the experimental results are listed and analyzed. Section 5 summarizes the conclusions of this paper.

2 Sparse low-rank representation model

2.1 Basic theory of sparse representation

Given a signal vector $\mathbf{V}$, sparse representation of $\mathbf{V}$ can be formulated as

$$\min_{\boldsymbol{a},\boldsymbol{D}} \|\boldsymbol{a}\|_0 subject\ to\ \|\mathbf{V}-\boldsymbol{D}\boldsymbol{a}\|_2 \le \varepsilon, \tag{1}$$

where $\boldsymbol{a}$ is the sparse coefficient of source image $\mathbf{V}$, $\boldsymbol{D}$ is an unknown dictionary matrix to be learnt, $\|\cdot\|_0$ and $\|\cdot\|_2$ are L0 norm and L2

TABLE 2 Different methods' average results of metrics on VIFB dataset. Red, green, and blue denote the best, second best, and third best results, numbers in the second column brackets are the statistical result of the best, second best, and third best results correspondingly. Some red, green, and blue results are ranked according to the third decimal place. Considering the size of the table, all results are only reserved for two decimal places. Considering the size of table, all results are only reserved for two decimal places. The data of the first 20 methods are from the workbench in [22].

Method	statistic	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	Q^{CB}	Q^{CV}	RMSE	SF	SSIM	SD
ADF	(0,0,0)	4.58 ± 0.39	1.46 ± 0.79	46.53 ± 15.30	6.79 ± 0.39	1.92 ± 0.45	58.41 ± 2.04	0.52 ± 0.09	0.47 ± 0.06	778 ± 345	0.104 ± 0.05	14.13 ± 4.42	1.40 ± 0.15	35.19 ± 10.61
CBF	(0,0,2)	**7.15** ± 0.30	0.99 ± 0.39	**74.59** ± 22.00	7.32 ± 0.30	2.16 ± 0.60	57.60 ± 2.07	0.58 ± 0.13	0.53 ± 0.08	1,575 ± 1,279	0.126 ± 0.06	20.38 ± 5.95	1.17 ± 0.20	48.54 ± 11.92
CNN	(1,0,2)	5.81 ± 0.32	1.03 ± 0.46	60.24 ± 23.30	7.32 ± 0.32	2.65 ± 0.73	57.93 ± 2.17	**0.66** ± 0.11	0.62 ± 0.12	**513** ± 642	0.118 ± 0.06	18.81 ± 6.42	1.39 ± 0.14	**60.08** ± 16.21
DLF	(3,0,0)	3.83 ± 0.46	1.41 ± 0.56	38.57 ± 15.70	6.72 ± 0.46	2.03 ± 0.46	**58.44** ± 2.06	0.43 ± 0.06	0.45 ± 0.05	760 ± 315	**0.104** ± 0.05	12.49 ± 4.67	**1.46** ± 0.13	34.72 ± 10.90
FPDE	(0,0,0)	4.54 ± 0.43	1.37 ± 0.66	46.02 ± 18.01	6.77 ± 0.43	1.92 ± 0.48	58.40 ± 2.06	0.48 ± 0.10	0.46 ± 0.05	780 ± 320	0.105 ± 0.05	13.47 ± 5.02	1.39 ± 0.20	34.93 ± 10.73
GFCE	(0,3,0)	**7.50** ± 0.41	1.93 ± 0.78	**77.40** ± 20.23	7.27 ± 0.41	1.84 ± 0.26	55.94 ± 1.30	0.47 ± 0.07	0.54 ± 0.06	899 ± 659	0.173 ± 0.05	**22.46** ± 6.15	1.13 ± 0.16	51.56 ± 10.23
GFF	(0,0,0)	5.33 ± 0.45	1.19 ± 0.54	55.20 ± 21.20	7.21 ± 0.45	2.64 ± 0.64	58.10 ± 2.05	0.62 ± 0.16	0.62 ± 0.12	882 ± 826	0.112 ± 0.05	17.27 ± 6.59	1.40 ± 0.13	50.06 ± 11.80
GTF	(0,0,0)	4.30 ± 0.66	1.29 ± 0.47	43.66 ± 17.70	6.51 ± 0.66	1.99 ± 0.74	57.86 ± 2.00	0.44 ± 0.10	0.41 ± 0.08	2,138 ± 1,292	0.118 ± 0.05	14.74 ± 5.62	1.37 ± 0.15	35.13 ± 19.90
HMSD_ GF	(0,0,1)	6.25 ± 0.35	1.16 ± 0.76	65.03 ± 25.60	7.27 ± 0.35	2.47 ± 0.56	57.94 ± 2.05	0.62 ± 0.11	0.60 ± 0.09	533 ± 525	19.904 ± 0.05	19.90 ± 6.66	1.39 ± 0.13	**57.62** ± 14.67
MSD	(0,1,1)	6.13 ± 0.37	1.26 ± 0.78	63.49 ± 25.80	7.30 ± 0.37	2.62 ± 0.69	58.17 ± 2.07	0.64 ± 0.09	**0.62** ± 0.11	**511** ± 474	0.110 ± 0.05	19.66 ± 7.04	1.41 ± 0.13	54.92 ± 14.32
IFEVIP	(0,0,0)	4.98 ± 0.42	1.34 ± 0.54	51.78 ± 21.30	6.95 ± 0.42	2.25 ± 0.73	57.17 ± 2.03	0.49 ± 0.05	0.46 ± 0.07	574 ± 341	0.138 ± 0.06	15.85 ± 5.69	1.39 ± 0.12	48.49 ± 13.29
LatLRR	(3,0,0)	**8.96** ± 0.42	1.68 ± 0.58	**92.81** ± 36.02	6.91 ± 0.42	1.65 ± 0.39	56.18 ± 1.69	0.44 ± 0.11	0.50 ± 0.05	697 ± 469	0.169 ± 0.07	**29.54** ± 10.08	1.18 ± 0.12	57.13 ± 15.03
MGF	(0,0,0)	5.84 ± 0.27	1.30 ± 0.37	60.61 ± 22.81	7.11 ± 0.27	1.77 ± 0.96	58.21 ± 2.10	0.57 ± 0.09	0.54 ± 0.12	677 ± 522	0.109 ± 0.06	17.92 ± 6.29	1.41 ± 0.14	44.29 ± 15.25
MST- SR	(0,2,3)	5.85 ± 0.42	**0.96** ± 0.78	60.78 ± 23.80	**7.34** ± 0.42	**2.81** ± 0.41	57.95 ± 2.06	**0.66** ± 0.09	**0.65** ± 0.05	523 ± 403	0.117 ± 0.05	18.81 ± 6.21	1.39 ± 0.13	57.31 ± 12.61

(Continued on the following page)

10.3389/fphy.2025.1514476

TABLE 2 (*Continued*) Different methods' average results of metrics on VIFB dataset. Red, green, and blue denote the best, second best, and third best results, numbers in the second column brackets are the statistical result of the best, second best, and third best results correspondingly. Some red, green, and blue results are ranked according to the third decimal place. Considering the size of the table, all results are only reserved for two decimal places. Considering the size of table, all results are only reserved for two decimal places. The data of the first 20 methods are from the workbench in [22].

Method	statistic	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	Q^{CB}	Q^{CV}	RMSE	SF	SSIM	SD
MSVD	(0,0,3)	3.55 ± 0.46	1.46 ± 0.58	36.20 ± 16.60	6.71 ± 0.46	1.96 ± 0.46	**58.42** ± 2.05	0.33 ± 0.12	0.43 ± 0.05	809 ± 242	**0.104** ± 0.05	12.53 ± 5.26	**1.43** ± 0.15	34.37 ± 11.02
NSCT-SR	(2,1,0)	6.50 ± 0.28	**0.90** ± 0.36	67.96 ± 22.10	**7.40** ± 0.28	**2.99** ± 1.01	57.44 ± 2.14	0.65 ± 0.05	0.62 ± 0.12	1,447 ± 1,254	0.131 ± 0.06	19.39 ± 6.38	1.28 ± 0.17	52.48 ± 11.99
ResNet	(1,2,0)	3.67 ± 0.45	1.36 ± 0.59	37.26 ± 15.10	6.74 ± 0.45	1.99 ± 0.39	**58.44** ± 2.05	0.41 ± 0.14	0.45 ± 0.05	725 ± 316	**0.104** ± 0.05	11.74 ± 4.35	**1.46** ± 0.13	34.94 ± 10.84
TIF	(0,0,0)	5.56 ± 0.25	1.37 ± 0.37	57.84 ± 21.70	7.08 ± 0.25	1.77 ± 0.69	58.23 ± 2.16	0.58 ± 0.11	0.55 ± 0.08	613 ± 729	0.109 ± 0.06	17.74 ± 6.94	1.40 ± 0.18	42.64 ± 14.38
VSMWLS	(0,0,0)	5.61 ± 0.41	1.41 ± 0.79	57.25 ± 21.30	7.03 ± 0.41	2.04 ± 0.38	58.19 ± 2.05	0.55 ± 0.09	0.50 ± 0.06	755 ± 453	0.109 ± 0.05	17.66 ± 5.56	1.42 ± 0.14	46.25 ± 11.33
RP-SR	(0,1,1)	6.36 ± 0.44	0.99 ± 0.89	65.22 ± 22.70	**7.35** ± 0.44	2.34 ± 0.45	57.78 ± 2.02	0.57 ± 0.09	0.61 ± 0.06	889 ± 634	0.122 ± 0.05	**21.17** ± 6.30	1.33 ± 0.13	55.81 ± 14.83
Our	(4,2,0)	5.86 ± 0.22	**0.94** ± 0.40	60.84 ± 22.69	7.32 ± 0.26	**2.99** ± 0.94	57.85 ± 2.09	**0.68** ± 0.10	**0.65** ± 0.11	**495** ± 371	0.119 ± 0.06	18.91 ± 6.38	1.39 ± 0.14	**58.10** ± 14.27

10.3389/fphy.2025.1514476

norm correspondingly, and $\varepsilon > 0$ is the error tolerance and is related to noise. The dictionary $\boldsymbol{D}$ can be efficiently solved by K-SVD, any source images can be sparsely represented with $\boldsymbol{D}$.

As the L0 norm is nonconvex and Equation 1 is a NP-hard problem, some research proved that the results of L1 norm are equal with L0 norm if the sparsity of optimized $\boldsymbol{a}$ is near its true value [26]. Hence, the L1 norm is used rather than the L0 norm to establish the spare representation model in this paper.

2.2 Sparse low-rank representation

In actual practice, the noise item ε in Equation 1 is hard to know in advance, so we change Equation 1 with L1 norm as the following form:

$$\min_{\boldsymbol{a}} \ \|\boldsymbol{a}\|_1 + \lambda\|\boldsymbol{E}\|_{2,1} \ \ subject \ \ to \ \ \mathbf{V} = \mathbf{D}\boldsymbol{a} + \boldsymbol{E}, \tag{2}$$

where $\boldsymbol{E}$ is the noise item and is expected to be kept as small as possible, λ is a trade-off parameter, and $\|\cdot\|_{2,1}$ is L21 norm and measures the level of noise. Some research has shown that low-rank constraints can further capture the complex structural information [27]. If the rank is not too large, the rank can be measured with nuclear norm [28]. We aim to combine sparse and low-rank constraints to enhance the performance of infrared and visible image fusion, so Equation 2 is changed as

$$\min_{\boldsymbol{a}} \ \|\boldsymbol{a}\|_1 + \alpha\|\boldsymbol{a}\|_* + \lambda\|\boldsymbol{E}\|_{2,1} \ \ subject \ \ to \ \ \mathbf{V} = \mathbf{D}\boldsymbol{a} + \boldsymbol{E}, \tag{3}$$

where $\|\cdot\|_*$ is the nuclear norm and measures the rank of a matrix and α is the balance parameter. The problem of Equation 3 is an optimization problem with constraint; we introduce the alternating direction method of multipliers (ADMM) [29] to solve Equation 3. We first introduce two auxiliary variables, $\boldsymbol{a}_1, \boldsymbol{a}_2$, and Equation 3 can be converted into

$$\min_{\boldsymbol{a}} \ \|\boldsymbol{a}_1\|_1 + \alpha\|\boldsymbol{a}_2\|_* + \lambda\|\boldsymbol{E}\|_{2,1} \ \ subject \ \ to \ \ \mathbf{V} = \mathbf{D}\boldsymbol{a} + \boldsymbol{E}, \boldsymbol{a} = \boldsymbol{a}_1, \boldsymbol{a} = \boldsymbol{a}_2. \tag{4}$$

Optimization of Equation 4 is equal to minimize the following augmented Lagrange multiplier (ALM) function L

$$\begin{aligned} \mathrm{L} = & \|\boldsymbol{a}_1\|_1 + \alpha\|\boldsymbol{a}_2\|_* + \lambda\|\boldsymbol{E}\|_{2,1} \\ & + \langle \boldsymbol{Y}_1, \mathbf{V} - \mathbf{D}\boldsymbol{a} - \boldsymbol{E}\rangle + \langle \boldsymbol{Y}_2, \boldsymbol{a} - \boldsymbol{a}_1\rangle + \langle \boldsymbol{Y}_3, \boldsymbol{a} - \boldsymbol{a}_2\rangle \\ & + \frac{\mu}{2}\left(\|\mathbf{V} - \mathbf{D}\boldsymbol{a} - \boldsymbol{E}\|_F^2 + \|\boldsymbol{a} - \boldsymbol{a}_1\|_F^2 + \|\boldsymbol{a} - \boldsymbol{a}_2\|_F^2\right), \end{aligned} \tag{5}$$

where μ represents positive penalty parameters, $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, and $\boldsymbol{Y}_3$ are Lagrange multipliers, and $\|\cdot\|_F$ denotes the Frobenius norm. There are four variables $\boldsymbol{a}, \boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{E}$, and three Lagrange multipliers $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, and $\boldsymbol{Y}_3$ in Equation 5. ADMM iteratively optimizes each variable while the others are fixed. Equation 5 is divided into five subproblems to separately optimize each variable.

(I) Update $\boldsymbol{a}$:

$$\boldsymbol{a} = \arg\ \min_{\boldsymbol{a}} \ \left\|\mathbf{V} - \mathbf{D}\boldsymbol{a} - \boldsymbol{E} + \frac{\boldsymbol{Y}_1}{\mu}\right\|_F^2 + \left\|\boldsymbol{a} - \boldsymbol{a}_1 + \frac{\boldsymbol{Y}_2}{\mu}\right\|_F^2 + \left\|\boldsymbol{a} - \boldsymbol{a}_2 + \frac{\boldsymbol{Y}_3}{\mu}\right\|_F^2. \tag{6}$$

Equation 6 is a convex optimization problem, and the solution can be calculated by taking the first partial derivatives to variable $\boldsymbol{a}$ and setting them to zero. Let $\mathbf{G}_1 = \mathbf{V} - \boldsymbol{E} + \frac{\boldsymbol{Y}_1}{\mu}, \mathbf{G}_2 = \boldsymbol{a}_1 - \frac{\boldsymbol{Y}_2}{\mu}, \mathbf{G}_3 = \boldsymbol{a}_2 - \frac{\boldsymbol{Y}_3}{\mu}$, it is clear that the close form solution of Equation 6 is

$$\boldsymbol{a} = \left(\mathbf{D}^{\mathrm{T}}\mathbf{D} + 2\mathbf{I}\right)^{-1}\left(\mathbf{D}^{\mathrm{T}}\mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_3\right). \tag{7}$$

(II) Update $\boldsymbol{a}_1$:

$$\boldsymbol{a}_1 = \arg\ \min_{\boldsymbol{a}_1} \ \|\boldsymbol{a}_1\|_1 + \frac{\mu}{2}\left\|\boldsymbol{a} - \boldsymbol{a}_1 + \frac{\boldsymbol{Y}_2}{\mu}\right\|_F^2. \tag{8}$$

The solution of $\boldsymbol{a}_1$ can be efficiently calculated with a soft-shrinkage operator [30]. Let $\boldsymbol{G}_4 = \boldsymbol{a} + \frac{\boldsymbol{Y}_2}{\mu}$, the solution of Equation 8 can be obtained with the soft-shrinkage operator and has the following form:

$$\boldsymbol{a}_1 = sign(\boldsymbol{G}_4) \odot \max\left(|\boldsymbol{G}_4| - \frac{1}{\mu}, 0\right), \tag{9}$$

where sign$(\cdot)$ is sign function and $\odot$ is Hadamard product.

(III) Update $\boldsymbol{a}_2$:

$$\boldsymbol{a}_2 = \arg\ \min_{\boldsymbol{a}_2} \ \alpha\|\boldsymbol{a}_2\|_* + \frac{\mu}{2}\left\|\boldsymbol{a} - \boldsymbol{a}_2 + \frac{\boldsymbol{Y}_3}{\mu}\right\|_F^2. \tag{10}$$

The problem in Equation 10 can be solved by the singular value thresholding scheme [31]. Let $\boldsymbol{G}_5 = \boldsymbol{a} + \frac{\boldsymbol{Y}_3}{\mu}$, the singular value decomposition of $\boldsymbol{G}_5$ is $\boldsymbol{G}_5 = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^{\mathrm{T}}$. Then, the solution of Equation 10 has the following form according to the singular value thresholding scheme:

$$\boldsymbol{a}_2 = \boldsymbol{U} \cdot sign(\boldsymbol{S}) \odot \max\left(|\boldsymbol{S}| - \frac{\alpha}{\mu}, 0\right) \cdot \boldsymbol{V}^{\mathrm{T}}. \tag{11}$$

(IV) Update $\boldsymbol{E}$:

$$\boldsymbol{E} = \arg\ \min_{\boldsymbol{E}} \ \lambda\|\boldsymbol{E}\|_{2,1} + \frac{\mu}{2}\left\|\mathbf{V} - \mathbf{D}\boldsymbol{a} - \boldsymbol{E} + \frac{\boldsymbol{Y}_1}{\mu}\right\|_F^2. \tag{12}$$

This sub-problem of Equation 12 can be efficiently solved by the half-quadratic optimization strategy [32]. Let $\boldsymbol{G}_6 = \mathbf{V} - \mathbf{D}\boldsymbol{a} + \frac{\boldsymbol{Y}_1}{\mu}$, this sub-problem of Equation 12 can be efficiently solved by the half-quadratic optimization strategy as

$$\boldsymbol{E}(:,\mathrm{i}) = \begin{cases} \dfrac{\|\boldsymbol{G}_6(:,\mathrm{i})\|_2 - \frac{\lambda}{\mu}}{\|\boldsymbol{G}_6(:,\mathrm{i})\|_2}\boldsymbol{G}_6(:,\mathrm{i}), & \|\boldsymbol{G}_6(:,\mathrm{i})\|_2 > \dfrac{\lambda}{\mu} \\ 0, & \|\boldsymbol{G}_6(:,\mathrm{i})\|_2 \le \dfrac{\lambda}{\mu} \end{cases}, \mathrm{i} = 1, \ldots, \mathrm{k}, \tag{13}$$

where $\boldsymbol{E}(:,\mathrm{i}), \boldsymbol{G}_6(:,\mathrm{i})$ denote the i-th column of $\boldsymbol{E}$ and $\boldsymbol{G}_6$.

(V) Update Lagrange multipliers $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$, and $\boldsymbol{Y}_3$:

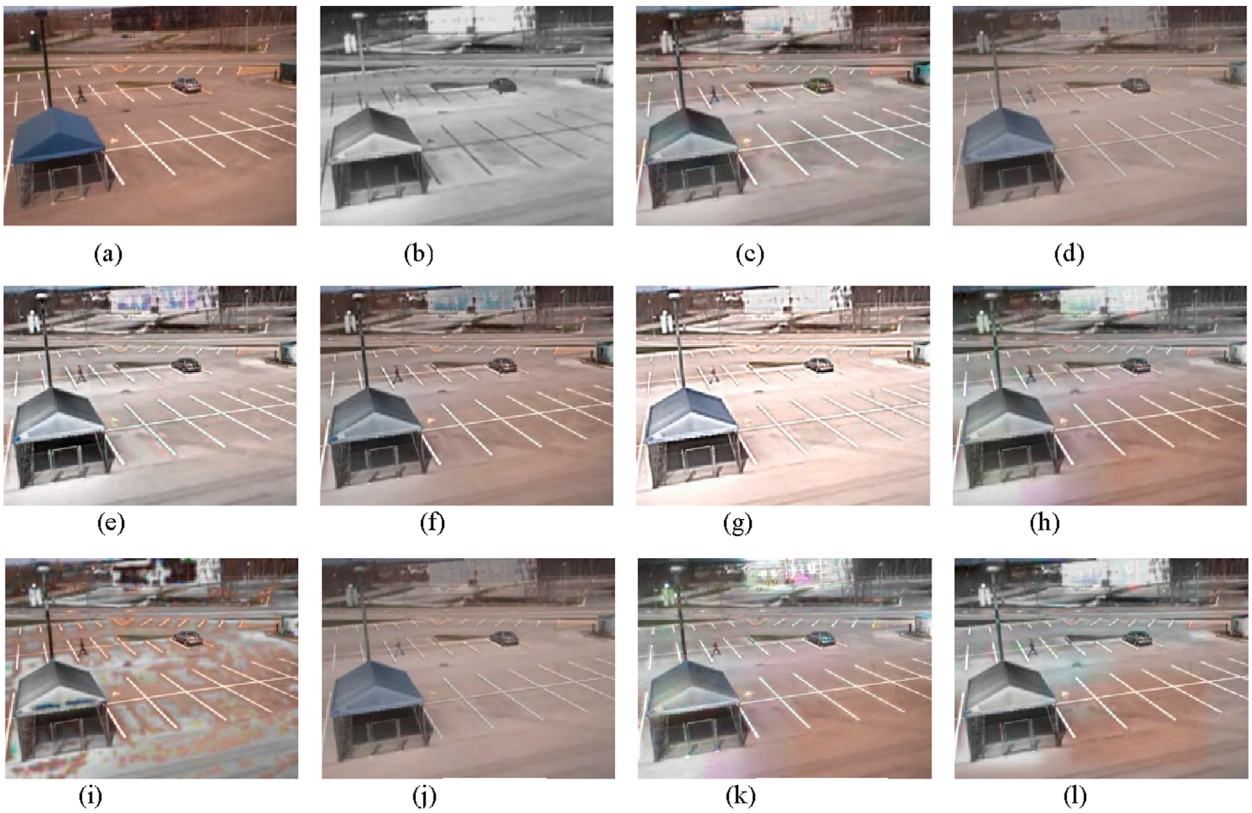

FIGURE 5
The visible and infrared image fusion result on the manCar image pair. The results of the first 9 methods are from the workbench in reference [22]. **(a)** Visible **(b)** Infrared **(c)** CNN **(d)** DLF **(e)** GFCE **(f)** MSD **(g)** LatLRR **(h)** MST_SR **(i)** NSCT_SR **(j)** ResNet **(k)** RP_SR **(l)** Our.

$$\begin{cases} \boldsymbol{Y}_1 = \boldsymbol{Y}_1 + \mu(\mathbf{V} - \mathbf{D}\boldsymbol{a} - \boldsymbol{E}) \\ \boldsymbol{Y}_2 = \boldsymbol{Y}_2 + \mu(\boldsymbol{a} - \boldsymbol{a}_1) \\ \boldsymbol{Y}_3 = \boldsymbol{Y}_3 + \mu(\boldsymbol{a} - \boldsymbol{a}_2) \end{cases}, \tag{14}$$

where μ increases dynamically from small values by $\mu = \rho\mu$, and ρ is a positive value.

The sparse low-rank coefficient $\boldsymbol{a}$ in Equation 5 can be solved by iteratively calculating the above five subproblems and obtaining their solutions as shown in Equations 7, 9, 11, 13, 14.

3 Fusion method based on sparse low-rank representation

In this section, we introduce the proposed infrared and visible image fusion method in detail. The proposed method consists of two steps: image decomposition and feature fusion. LP is used to decompose images into low-pass and high-pass parts. Low-pass parts are a smooth version of source images; they are fused based on sparse low-rank representation. High-pass parts contain more details and the fusion is based on the max-absolute rule. The schematic diagram of the proposed fusion method is shown in Figure 1.

3.1 Image decomposition

MST is a popular transformation method and can transform the source images into a multi-scale domain. The source images are firstly transformed into a multi-scale domain by LP to obtain different information. Figure 2 shows the diagram of LP. In the first layer of decomposition, source image $\mathbf{I}$ is first convolved twice and down sampled once to obtain an interlayer result $\mathbf{M}_\mathbf{d}^\mathbf{1}$. The interlayer result $\mathbf{M}_\mathbf{d}^\mathbf{1}$ is used to compute the second interlayer result $\mathbf{M}_\mathbf{d}^\mathbf{2}$ and the first high-pass band $\mathbf{H}_\mathbf{1}$. The second interlayer result $\mathbf{M}_\mathbf{d}^\mathbf{2}$ for the next decomposition level is obtained by two convolutions and a down sampling. Meanwhile, after the interlayer result $\mathbf{M}_\mathbf{d}^\mathbf{1}$ goes through an up sampling and two convolutions, a high-pass interlayer result $\mathbf{HP}_\mathbf{1}$ is obtained. The first high-pass band $\mathbf{H}_\mathbf{1}$ can be calculated by subtracting the high-pass interlayer result $\mathbf{HP}_\mathbf{1}$ from source image $\mathbf{I}$. In following i-th layer decomposition, similar conductions to those in the first layer are performed on the interlayer result $\mathbf{M}_\mathbf{d}^\boldsymbol{i}(i = 1, 2, ..., n)$ from the previous decomposition. The single difference is that, in the subtraction, the high-pass interlayer result $\mathbf{HP}_\boldsymbol{i}$ is subtracted from the interlayer result $\mathbf{M}_\mathbf{d}^{\boldsymbol{i-1}}$ from the previous level subtracts to calculate the i-th high-pass band $\mathbf{H}_\boldsymbol{i}$. After n layers of decompositions, one low-pass band $\mathbf{L}$ and n high-pass bands $\mathbf{H}_\mathbf{1}$, $\mathbf{H}_\mathbf{2}$, …, $\mathbf{H}_\mathbf{n}$ can be obtained; the number of high-pass bands is equal to the layers of decomposition.

TABLE 3 Different methods' average results of metrics on the TNO dataset. Red, green, and blue mean the best, second best, and third best results, numbers in the second column brackets are the statistical result of the best, second best, and third best results correspondingly. Some red, green, and blue results are ranked according to the third decimal place. Considering the size of table, all results are only reserved for two decimal places.

Method	statistic	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	Q^{CB}	Q^{CV}	RMSE	SF	SSIM	SD
CNN	(1,2,1)	4.89 ± 3.02	**0.97** ± 0.66	48.15 ± 30.50	7.11 ± 0.49	1.78 ± 0.50	58.94 ± 1.53	**0.58** ± 0.04	0.55 ± 0.09	**335** ± 261	0.09 ± 0.03	12.81 ± 6.86	1.40 ± 0.19	**47.40** ± 16.11
DLF	(2,1,0)	2.80 ± 1.57	1.47 ± 0.89	27.75 ± 15.94	6.45 ± 0.59	1.64 ± 0.55	**59.61** ± 1.60	0.37 ± 0.05	0.48 ± 0.07	498 ± 367	**0.08** ± 0.03	7.18 ± 3.54	**1.52** ± 0.22	27.27 ± 11.09
GFCE	(0,4,1)	**7.01** ± 3.00	1.95 ± 0.84	**67.34** ± 29.71	**7.21** ± 0.27	1.26 ± 0.40	57.46 ± 1.46	0.42 ± 0.10	**0.55** ± 0.08	534 ± 425	0.12 ± 0.05	**18.72** ± 7.46	1.14 ± 0.19	44.81 ± 10.97
MSD	(0,1,3)	5.15 ± 3.24	1.28 ± 1.00	50.29 ± 32.62	7.01 ± 0.49	1.65 ± 0.44	**59.26** ± 1.56	0.53 ± 0.05	0.54 ± 0.09	**355** ± 274	**0.08** ± 0.03	13.54 ± 7.38	**1.42** ± 0.18	41.68 ± 14.10
LatLRR	(3,0,1)	**7.69** ± 3.07	2.52 ± 0.69	**76.41** ± 31.09	6.57 ± 0.47	1.26 ± 0.67	56.97 ± 1.62	0.37 ± 0.04	0.48 ± 0.09	**403** ± 352	0.14 ± 0.04	**20.53** ± 6.86	1.18 ± 0.19	44.77 ± 13.87
NSCT-SR	(3,0,0)	3.83 ± 2.81	**0.67** ± 0.54	38.24 ± 28.56	5.34 ± 0.56	**2.22** ± 0.86	44.21 ± 1.64	0.44 ± 0.07	0.43 ± 0.10	825 ± 1,061	**0.07** ± 0.04	9.72 ± 6.35	0.97 ± 0.19	32.26 ± 14.42
ResNet	(0,3,0)	2.73 ± 5.80	1.49 ± 0.69	27.19 ± 55.38	6.45 ± 0.48	1.47 ± 0.51	**59.60** ± 1.66	0.36 ± 0.09	0.49 ± 0.07	527 ± 618	**0.08** ± 0.04	7.02 ± 14.78	**1.52** ± 0.30	27.06 ± 14.09
RP-SR	(0,0,5)	**6.15** ± 1.48	1.04 ± 0.98	**58.33** ± 15.05	**7.16** ± 0.59	1.41 ± 0.49	58.71 ± 1.60	0.45 ± 0.06	0.54 ± 0.07	709 ± 378	0.09 ± 0.03	**17.41** ± 3.26	1.29 ± 0.22	**46.20** ± 11.03
MST-SR	(1,1,1)	4.95 ± 4.90	0.97 ± 1.06	48.82 ± 50.04	7.16 ± 0.85	**1.84** ± 0.30	58.94 ± 1.23	**0.58** ± 0.03	**0.56** ± 0.07	485 ± 337	0.09 ± 0.04	12.84 ± 10.70	1.40 ± 0.16	46.08 ± 17.18
Our	(3,1,1)	4.95 ± 3.03	**0.84** ± 0.42	48.85 ± 30.05	**7.21** ± 0.24	**2.05** ± 0.70	58.78 ± 0.81	**0.58** ± 0.08	**0.56** ± 0.10	509 ± 426	0.09 ± 0.01	12.82 ± 8.92	1.39 ± 0.12	**47.62** ± 8.74

10.3389/fphy.2025.1514476

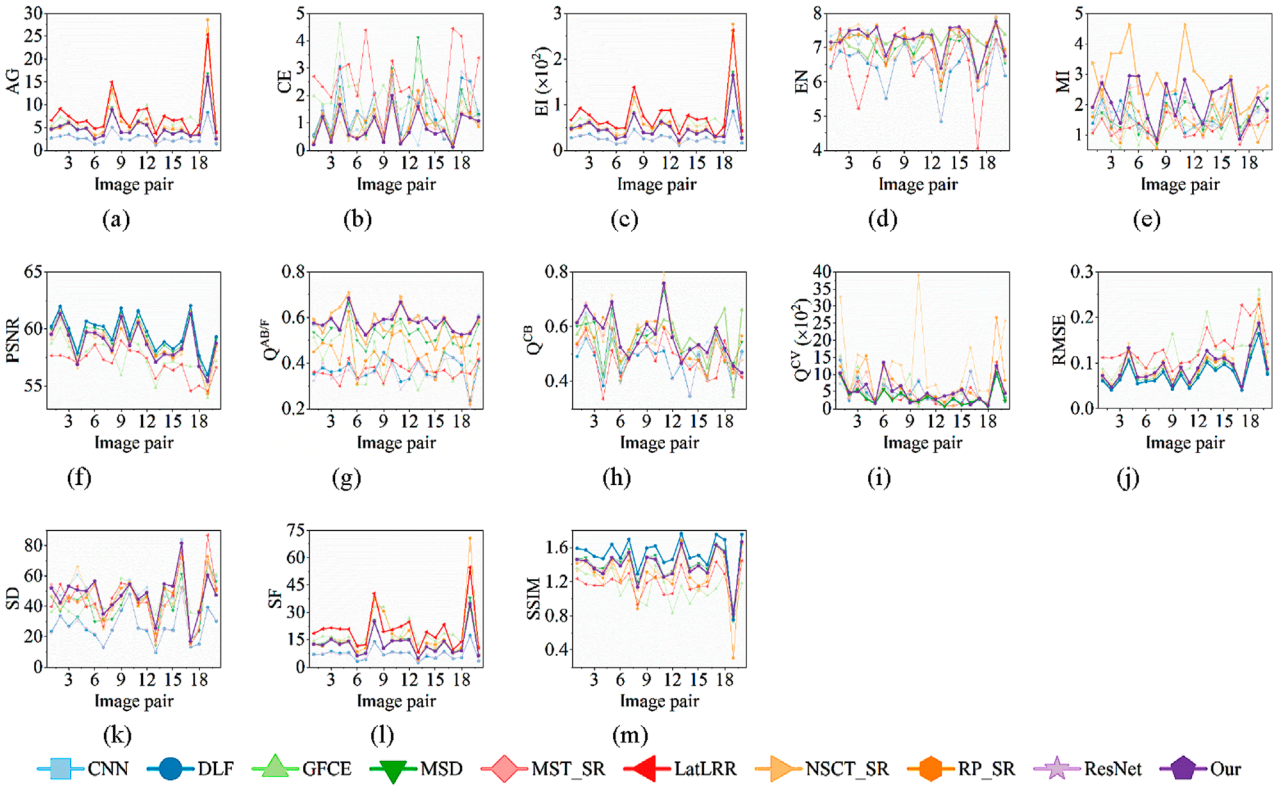

FIGURE 6
Metrics of 20 randomly selected image pairs. The best result and our result are bolded, and there is one bolded result if our result is the best. **(a)** AG **(b)** CE **(c)** EI **(d)** EN **(e)** MI **(f)** PSNR **(g)** $Q^{AB/F}$ **(h)** Q^{CB} **(i)** Q^{CV} **(j)** RMSE **(k)** SD **(l)** SF **(m)** SSIM.

3.2 Fusion strategy

After the decomposition is completed and the low-pass bands and high-pass bands are obtained, the fusion step is conducted. Different fusion strategies are used to deal with low-pass bands and high-pass bands.

3.2.1 Fusion strategy of low-pass bands

The low-pass bands of infrared and visible image are represented by $\mathbf{L}_A$ and $\mathbf{L}_B$; they are fused based on sparse low-rank representation. The sparse low-rank coefficients of $\mathbf{L}_A$ and $\mathbf{L}_B$ are firstly computed according to Equation 6. In order to improve the computing efficiency, the low-pass bands are divided into several image patches of size $\sqrt{n} \times \sqrt{n}$ by a sliding window. The sliding window slides from the top left to the bottom right and the step size of the sliding window is set as s pixels. In order to avoid block effect and keep enough overlap between different patches, the step size is less than the patch size. However, the step size cannot be too small or the overlapping region will be smoothed. Suppose there are k patches for $\mathbf{L}_A$ and $\mathbf{L}_B$, each patch is represented as $\mathbf{P}^i$ (i = 1, 2, …, k). In the following procedure, every patch is rearranged in column vector form, and the column vector of $\mathbf{P}^i$ is marked as $\mathbf{V}^i \in \mathbf{R}^{n\times 1}$.

In order to further improve the computing efficiency, the column vector $\mathbf{V}^i$ (i = 1, 2, …, k) is not directionally input into our sparse low-rank representation model. Firstly, the column vector $\mathbf{V}^i$ is zero-averaged as follows

$$\hat{\mathbf{V}}^i = \mathbf{V}^i - \overline{\mathbf{V}}^i \cdot \mathbf{1}, \tag{15}$$

where $\hat{\mathbf{V}}^i$ is zero-averaged patch, $\overline{\mathbf{V}}^i$ is the mean value of $\mathbf{V}^i$, and $\mathbf{1} \in \mathbf{R}^{n\times 1}$ is an all-one vector. Then, $\hat{\mathbf{V}}^i$ is used to solve the sparse low-rank coefficient. According to Equation 5, the optimization model can be expressed as

$$\min_{\boldsymbol{a}} \left\|\boldsymbol{a}_1^i\right\|_1 + \alpha\left\|\boldsymbol{a}_2^i\right\|_* + \lambda\|\boldsymbol{E}\|_{2,1} \quad subject\ \ to\ \ \hat{\mathbf{V}}^i = \mathbf{D}\boldsymbol{a}^i + \boldsymbol{E}, \boldsymbol{a}^i = \boldsymbol{a}_1^i, \boldsymbol{a} = \boldsymbol{a}_2^i, \tag{16}$$

It is worth noting that dictionary $\mathbf{D}$ is learnt based on $\hat{\mathbf{V}}^i$, and $\hat{\mathbf{V}}^i$ only contains the structural information. Therefore, the learnt $\mathbf{D}$ by Equation 1 is a universal dictionary and it can be used for any specific transform domain or parameter setting. Besides, because $\mathbf{D}$ is trained with the image patches $\hat{\mathbf{V}}^i$ rather than whole images, $\mathbf{D}$ can keep a small size, which is important to deduce the requirements of computers and improve the solving efficiency.

Let $\mathbf{V}_A^i$ and $\mathbf{V}_B^i$ be the i-th patch of $\mathbf{L}_A$ and $\mathbf{L}_B$ in column vector form, the sparse low-rank coefficients $\boldsymbol{a}_A^i, \boldsymbol{a}_B^i$ of $\mathbf{L}_A$ and $\mathbf{L}_B$ can be obtained by Equations 15 and 16. Then the max-L1 norm rule is used to fuse $\boldsymbol{a}_A^i$ and $\boldsymbol{a}_B^i$, and the fused sparse low-rank coefficient $\boldsymbol{a}_F^i$ can

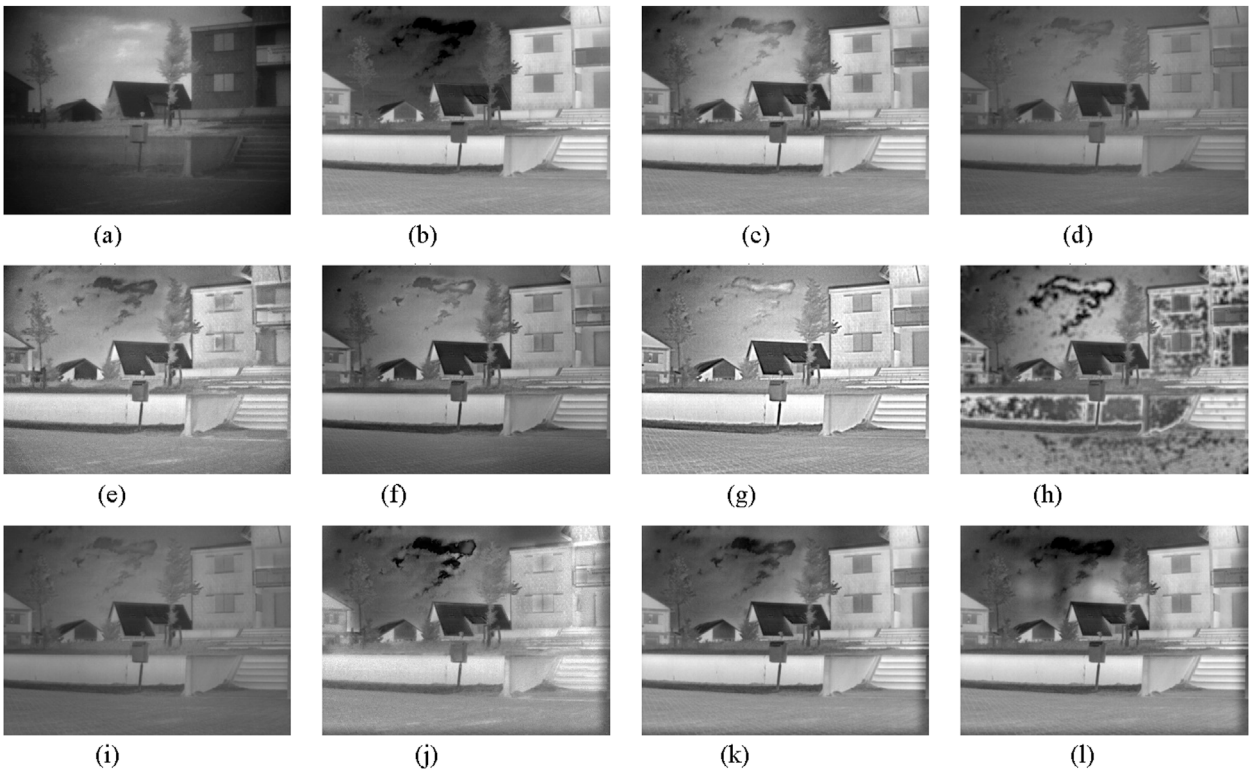

FIGURE 7
The visible and infrared image fusion result on the 29-th image pair. **(a)** Visible **(b)** Infrared **(c)** CNN **(d)** DLF **(e)** GFCE **(f)** Hybrid_MSD **(g)** LatLRR **(h)**NSCT_SR **(i)** ResNet **(j)** RP_SR **(k)** MST_SR **(l)** Our.

be obtained by

$$\boldsymbol{a}_F^i = \begin{cases} \boldsymbol{a}_A^i, & \left\|\boldsymbol{a}_A^i\right\|_1 > \left\|\boldsymbol{a}_B^i\right\|_1 \\ \boldsymbol{a}_B^i, & \text{else} \end{cases}, i = 1, ..., \mathrm{k}, \tag{17}$$

and the fused patch $\mathbf{V}_F^i$ can be obtained by

$$\mathbf{V}_F^i = \mathbf{D}\boldsymbol{a}_F^i + \overline{\mathrm{V}}_F^i \cdot \mathbf{1}, i = 1, ..., \mathrm{k}, \tag{18}$$

where $\overline{\mathrm{V}}_F^i$ is an average. If $\left\|\boldsymbol{a}_A^i\right\|_1 > \left\|\boldsymbol{a}_B^i\right\|_1$, $\overline{\mathrm{V}}_F^i$ takes as the average of $\mathbf{V}_A^i$, otherwise $\overline{\mathrm{V}}_F^i$ is the average of $\mathbf{V}_B^i$. After all k patches are fused with Equations 17, 18, each patch in the column vector form is converted to two-dimensional form. All the two-dimensional patches are spliced to obtain the fused low-pass bands $\mathbf{L}_\mathrm{F}$.

3.2.2 Fusion strategy of high-pass bands

The high-pass bands are directly fused based with the max-absolute rule. There are n high-pass bands and every high-pass band pair is fused separately. The absolute value of high-pass band $\mathbf{H_i}$ (i = 1, 2, …, n) is first computed. Then, the absolute value goes through 2-D order-statistic filtering. The filtering results of visible and infrared high-pass bands are compared directly. The bigger is reserved and the smaller is discarded. The fused high-pass band $\mathbf{H}_\mathbf{F}^i$ can be obtained by adding the comparation results. The detailed information can be seen in [33].

3.3 Reconstruction of fused image

After $\mathbf{L_F}$ and $\mathbf{H_F}$ are obtained, a reverse LP is conducted to reconstruct the fused image $\mathbf{I}_\mathrm{F}$. The diagram of reconstruction is shown in Figure 3. The reconstruction is the reverse operation of decomposition, there are also n layers of reconstruction. The low-pass band $\mathbf{L}_\mathrm{F}$ firstly goes through an upsampling and two convolutions, and we can obtain an interlayer result $\mathbf{M}_\mathbf{r}^\mathbf{1}$. The interlayer result $\mathbf{M}_\mathbf{r}^\mathbf{1}$ is added to the n-th fused high-pass band $\mathbf{H}_\mathbf{F}^\mathbf{n}$, and the sum result is transferred into the second layer. In every following reconstruction layer, a similar operation is conducted. The only difference is that the high-pass band of each layer is used in the additional operation. In the n-th layer of reconstruction, the fused image $\mathbf{I_F}$ can be reconstructed by adding the interlayer result $\mathbf{M}_\mathbf{r}^\mathbf{n}$ and the first fused high-pass band $\mathbf{H}_\mathbf{F}^\mathbf{1}$.

4 Experiments and evaluation

In this section, we verify and analyze our method through extensive experiments. All the experiments were performed on a desktop with Windows 10 operating system, 11th Gen Intel(R) Core (TM) i7-11700F @ 2.50 GHz CPU, and 16 GB RAM. In our method, the level of decomposition is 4, both of a, and λ is 1e-5. It is noteworthy that the convolution kernel is the same across all the convolution layers, because the impact of convolution kernels is much weaker than the number of decomposition layers [15]. The

TABLE 4 Different methods' average results of metrics on the LLVIP dataset. Red, green, and blue mean the best, second best, and third best results, numbers in the second column brackets are the statistical result of the best, second best, and third best results correspondingly. Considering the size of table, all results are only reserved for two decimal places.

Method	statistic	AG	CE	EI	EN	MI	PSNR	$Q^{AB/F}$	Q^{CB}	Q^{CV}	RMSE	SF	SSIM	SD
CNN	(0,2,1)	4.47 ± 3.03	0.77 ± 0.51	46.06 ± 29.96	7.09 ± 0.36	1.95 ± 0.53	62.75 ± 0.77	0.67 ± 0.10	0.48 ± 0.10	427 ± 219	0.04 ± 0.006	15.59 ± 8.85	1.24 ± 0.11	49.74 ± 9.25
DLF	(3,0,0)	2.66 ± 1.95	1.21 ± 0.63	27.38 ± 18.43	6.73 ± 0.35	1.94 ± 0.22	63.54 ± 0.91	0.48 ± 0.06	0.41 ± 0.05	648 ± 261	0.03 ± 0.006	9.39 ± 6.12	1.38 ± 0.12	31.67 ± 8.75
GFCE	(1,3,1)	7.89 ± 3.00	2.02 ± 0.55	81.60 ± 29.55	7.46 ± 0.23	1.18 ± 0.26	60.41 ± 0.60	0.43 ± 0.14	0.53 ± 0.02	651 ± 335	0.06 ± 0.008	23.35 ± 7.85	0.83 ± 0.14	48.51 ± 8.00
MSD	(1,2,3)	4.61 ± 3.09	0.81 ± 0.53	47.23 ± 30.59	6.83 ± 0.43	1.64 ± 0.64	63.08 ± 0.79	0.69 ± 0.09	0.54 ± 0.09	412 ± 254	0.03 ± 0.005	16.15 ± 8.94	1.33 ± 0.12	39.00 ± 12.73
LatLRR	(4,1,0)	8.19 ± 3.04	0.97 ± 0.52	84.92 ± 30.05	7.34 ± 0.23	1.60 ± 0.55	61.55 ± 0.77	0.42 ± 0.08	0.45 ± 0.10	556 ± 290	0.05 ± 0.006	28.47 ± 8.89	1.14 ± 0.11	58.14 ± 8.76
NSCT_SR	(2,1,0)	5.07 ± 1.73	0.57 ± 0.62	53.08 ± 16.56	7.06 ± 0.17	2.46 ± 0.19	62.45 ± 0.67	0.69 ± 0.04	0.52 ± 0.05	683 ± 239	0.04 ± 0.007	15.74 ± 5.21	1.22 ± 0.10	43.99 ± 8.55
ResNet	(2,1,0)	2.50 ± 2.86	1.19 ± 0.30	25.84 ± 28.15	6.74 ± 0.34	1.79 ± 0.81	63.53 ± 0.89	0.43 ± 0.10	0.41 ± 0.10	609 ± 404	0.03 ± 0.007	8.55 ± 8.68	1.38 ± 0.15	31.91 ± 10.05
RP_ SR	(1,0,6)	6.83 ± 2.95	0.73 ± 0.48	70.13 ± 29.25	7.33 ± 0.23	1.33 ± 0.46	62.26 ± 0.93	0.48 ± 0.18	0.55 ± 0.04	715 ± 411	0.04 ± 0.008	22.09 ± 7.96	1.10 ± 0.24	49.60 ± 7.97
MST_SR	(1,0,3)	4.65 ± 5.11	0.73 ± 0.49	47.88 ± 50.77	7.25 ± 0.36	2.08 ± 0.23	62.71 ± 0.91	0.70 ± 0.09	0.52 ± 0.06	493 ± 754	0.04 ± 0.006	15.93 ± 15.24	1.32 ± 0.13	47.52 ± 11.33
Our	(1,3,0)	4.11 ± 3.06	0.69 ± 0.50	42.26 ± 30.97	7.20 ± 0.45	2.24 ± 0.68	62.69 ± 1.58	0.70 ± 0.04	0.52 ± 0.09	570 ± 401	0.04 ± 0.004	14.16 ± 6.84	1.35 ± 0.18	45.38 ± 13.89

10.3389/fphy.2025.1514476

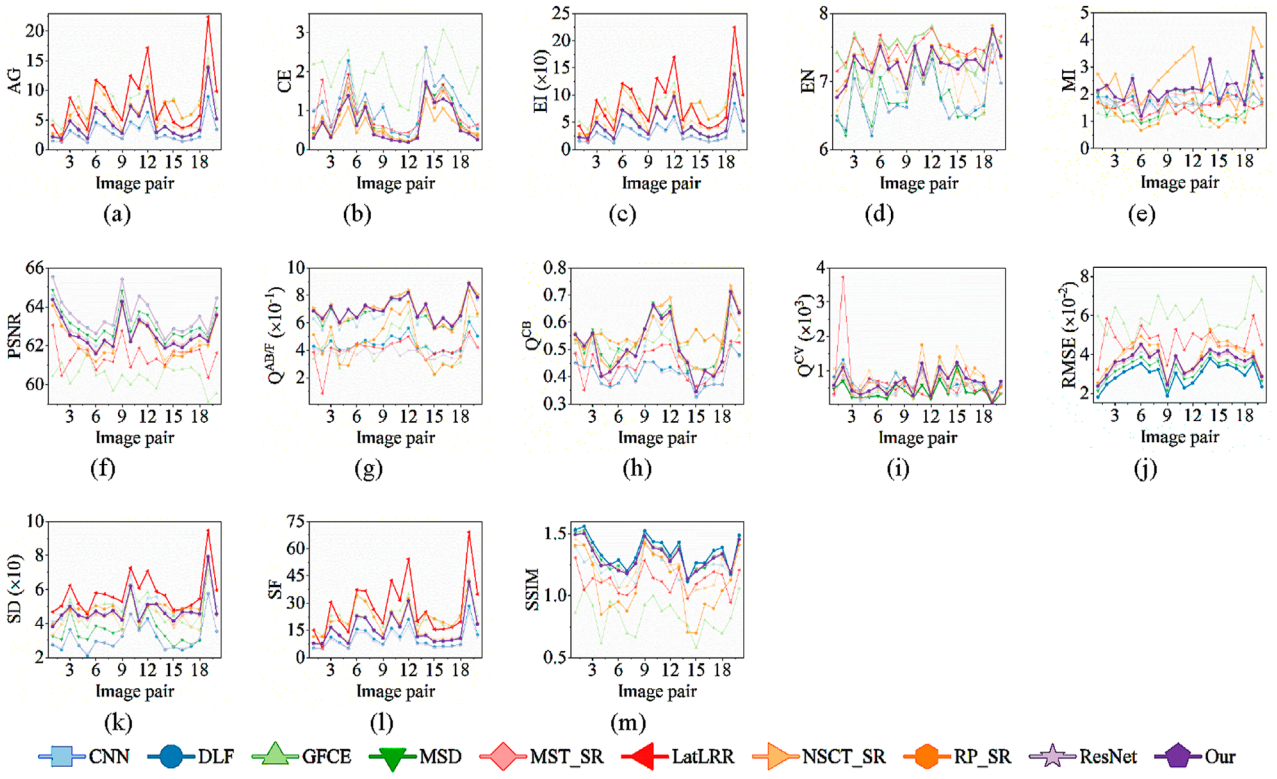

FIGURE 8
Metrics of 20 randomly selected image pairs. The best result and our result are bolded, and there is one bolded result if our result is the best. **(a)** AG **(b)** CE **(c)** EI **(d)** EN **(e)** MI **(f)** PSNR **(g)** $Q^{AB/F}$ **(h)** Q^{CB} **(i)** Q^{CV} **(j)** RMSE **(k)** SD **(l)** SF **(m)** SSIM.

convolution kernel in the following experiments is a Gauss filter [1 4 6 4 1]/16.

4.1 Experimental setups

4.1.1 Source images

In order to verify the effectiveness and analyze the characteristics of our method, three open access visible and infrared image datasets are adopted in the following experiments, namely, VIFB [22], TNO [34], LLVIP [35]. In VIFB, there are 21 pairs of visible and infrared images. In TNO, there are 37 image pairs; the resolution of images is 640 × 480. In LLVIP, there are 50 image pairs for validation and the images are mainly taken in dark environments.

4.1.2 Objective evaluation metrics

The metrics of image fusion can be divided into four types: information theory-based, image feature-based, image structural similarity-based, and human perception-based [22]. To evaluate and compare performance comprehensively and objectively, 13 evaluation metrics are used in the following experiments. The 13 metrics and their meaning are listed in Table 1. Considering that there are two source images and a single fused image, the 13 metrics can be categorized into three types based on their calculation methods: metrics independent of the source images, metrics computed separately on the fused and source images and then averaged, and metrics computed separately on the fused and source images and then summed. All the calculation methods are based on [15]. Metrics, including AG, EI, EN, SF, and SD, belong to the first category. The second type of metrics includes PSNR, $Q^{AB/F}$, Q^{CB}, Q^{CV}, and RMSE, and the rest of the metrics are categorized into the third type.

4.2 Experiment results on VIFB dataset

The proposed method is first tested on the VIFB dataset and compared with 20 kinds of methods, namely, anisotropic diffusion-based image fusion (ADF) [7], cross bilateral filter fusion method (CBF) [8], convolutional neural network (CNN) [19], Deep learning framework (DLF) [20], fourth order partial differential equations (FPDE) [3], guided filter context enhancement (GFCE) [9], guided filtering-based fusion method (GFF) [10], Gradient Transfer Fusion (GTF) [4], guided filter-based hybrid multi-scale decomposition [9], hybrid multi-scale decomposition [14], infrared feature extraction and visual information preservation (IFEVIP) [2], latent low-rank representation (LatLRR) [5], multi-scale guided filtered-based fusion (MGF) [11], MST and sparse representation (MST_SR) [15], multi-resolution singular value decomposition (MSVD) [12], nonsubsampled contourlet transform and sparse representation (NSCT_SR) [15], ResNet [21], Two-scale image fusion (TIF) [6], visual saliency map and weighted least square (VSMWLS) [16], and ratio of low-pass pyramid and sparse representation (RP_SR)

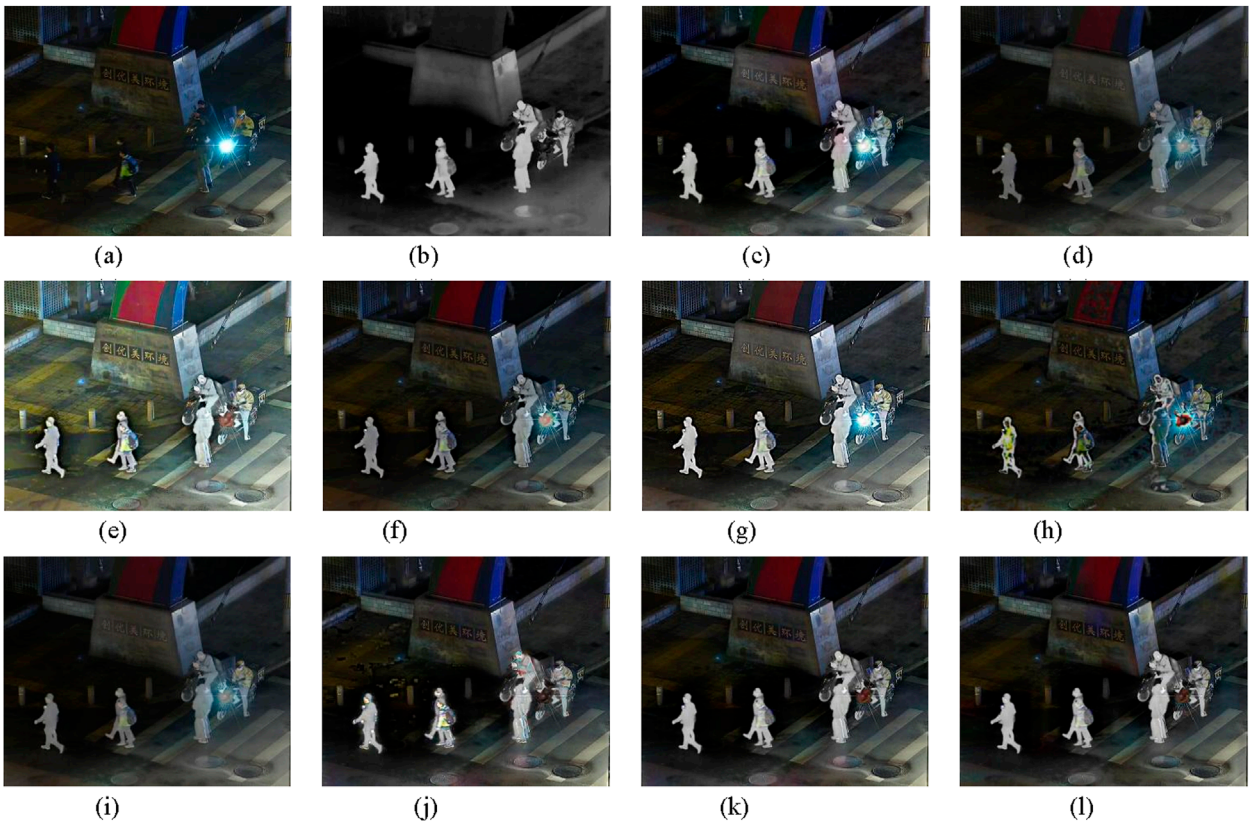

FIGURE 9
The visible and infrared image fusion results on the *third* image pair. **(a)** Visible **(b)** Infrared **(c)** CNN **(d)** DLF **(e)** GFCE **(f)** Hybrid_MSD **(g)** LatLRR **(h)** NSCT_SR **(i)** ResNet **(j)** RP_SR **(k)** MST_SR **(l)** Our.

TABLE 5 The runtime of different methods in three datasets (seconds per image pair). Red is the best result.

Dataset	CNN	DLF	GFCE	MSD	LatLRR	NSCT_SR	ResNet	RP_SR	MST_SR	Our
VIFB	36.019	10.94	2.151	8.66	269.823	2,119.319	**2.058**	11.952	3.993	12.5
TNO	38.866	4.106	**0.71**	3.148	97.474	1,205.000	3.064	5.31	5.33	17.932
LLVIP	168.878	58.231	10.422	39.383	1,252.133	11,108.11	**9.763**	79.72	99.611	79.781
Average	81.254	24.426	**4.428**	17.064	539.81	4812.810	4.962	32.327	36.311	36.738

[15]. The results of the 20 methods are from the workbench in reference [22].

We can see that our method obtains four best results and two second-best results. The four best results include three categories of metrics: information theory-based metrics, image feature-based metrics, and human visual perception. It is worth pointing out that the other metrics among the three categories also keep a relatively good level. From the information theory viewpoint, our method avoids losing information both in LP and image fusion. LP is a reversible transform, so there is no information to be lost. Sparse low-rank coefficients with maximum L1 norm contain most of the information, so we can believe that our method can keep as much information as possible. However, the performance in structural similarity-based metrics does not really stand out. This is an inherent drawback of sparse representation. All the information of fused low-pass is from the fixed dictionary. The dictionary is learnt from a specific image dataset. It is inevitable to lose some structural information. A larger dictionary helps to improve the performance, but a bigger size induces more computational costs. Besides, as previously mentioned, the overlap between different patches is smoothed and some structural information is also lost in this stage. It is helpful if the step size of the sliding window remains relatively big. As a contrast, LatLRR also receives a good result, but its performance in structural similarity-based metrices is not satisfactory. DLF, MSVD, and RestNet can achieve good results in structural similarity-based metrices, however the

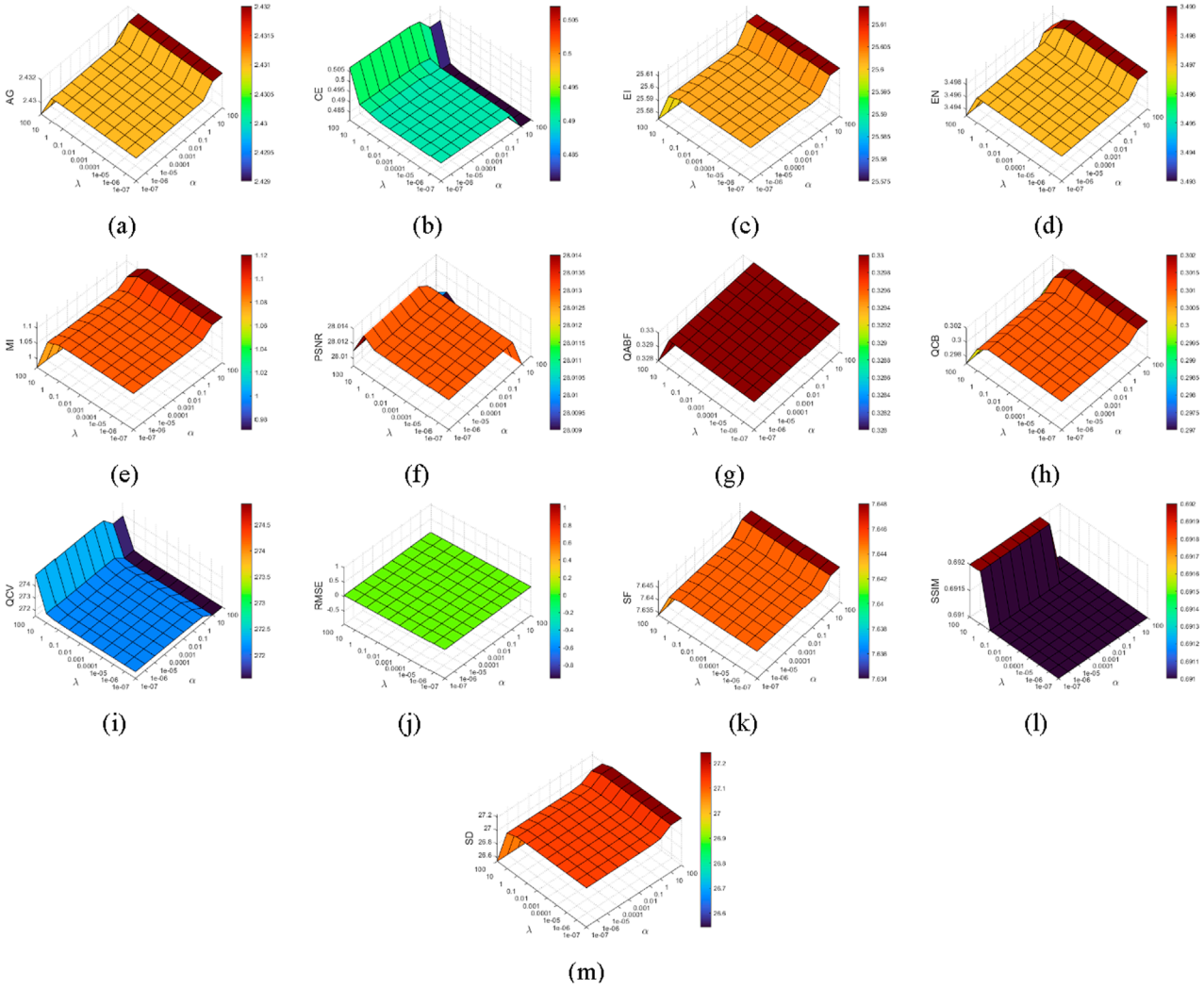

FIGURE 10
Analysis of parameter sensitivity. **(a)** AG **(b)** CE↓ **(c)** EI **(d)** EN **(e)** MI **(f)** PSNR **(g)** $Q^{AB/F}$ **(h)** Q^{CB} **(i)** Q^{CV}↓ **(j)** RMSE↓ **(k)** SF **(l)** SSIM **(m)** SD.

other metrics cannot be maintained as outstanding. The results of the proposed method also demonstrate favorable performance in terms of standard deviations (STD). Specifically, smaller standard deviations are observed for the AG, CE, and EN metrics, slightly poorer performance is seen for the MI and Q^{CB} metrics, and the remaining metrics exhibit performances close to the average level.

Furthermore, we draw the metrics of all 21 image pairs as shown in Figure 4. Only nine methods with better results in Table 2 are compared. Firstly, it is easy to notice that the results vary significantly from image pair to image pair. Our method shows better performance in Q^{CB} and Q^{CV} than other methods, as Figures 4h,i show. In CE, MI, and $Q^{AB/F}$, the performance of our method is very near the best result of NSCT-SR. In EN, the performance of our method is very near the best result of GFCE. In PSNR, RMSE, and SSIM, the performance of our method is very near the best result of DLF. In AG, EI, and SF, the performance of our method is average and LatLRR shows the best results. The results are consistent with Table 2. However, we can find that our method shows better stability among different image pairs, as Figures 4d,e,g show. Although our method does not obtain the best result, the variance among different image pairs is smaller than the methods with the best results.

Additionally, Figure 5 shows the fused image on the man Car image pair. Our method preserves as much information as possible: the majority of details from the visible and infrared image can be found in the fused image. In contrast, there are some notable discontinuities in NSCT_SR, which results in the introduction of certain new features and the loss of some crucial information. While DLF, MSD, and ResNet save more information about visible image, CNN, LatLRR, and GFCE store more information about the infrared image. The results of RP_SR and MST_SR are comparatively close to our method. It is possible to include some artificial features because the dictionary of sparse representation is fixed. Both colors of the car's rear door and the roof adjacent to it have altered, as we can see in Figure 5k. There are some pink zones on the ground in Figures 5h,k, they are not present in the source image pair. In contrast, our method effectively suppresses the artificial features, as Figure 5l shows. This demonstrates once more the advantages of our

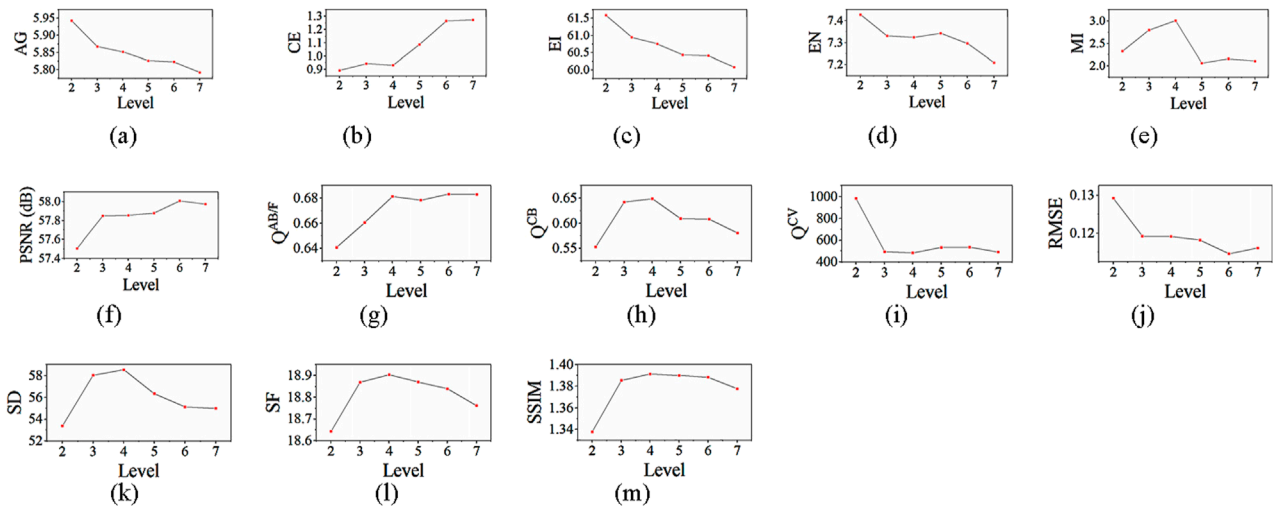

FIGURE 11
Influence of decomposition levels. **(a)** AG **(b)** CE↓ **(c)** EI **(d)** EN **(e)** MI **(f)** PSNR **(g)** $Q^{AB/F}$ **(h)** Q^{CB} **(i)** Q^{CV}↓ **(j)** RMSE↓ **(k)** SD **(l)** SF **(m)** SSIM.

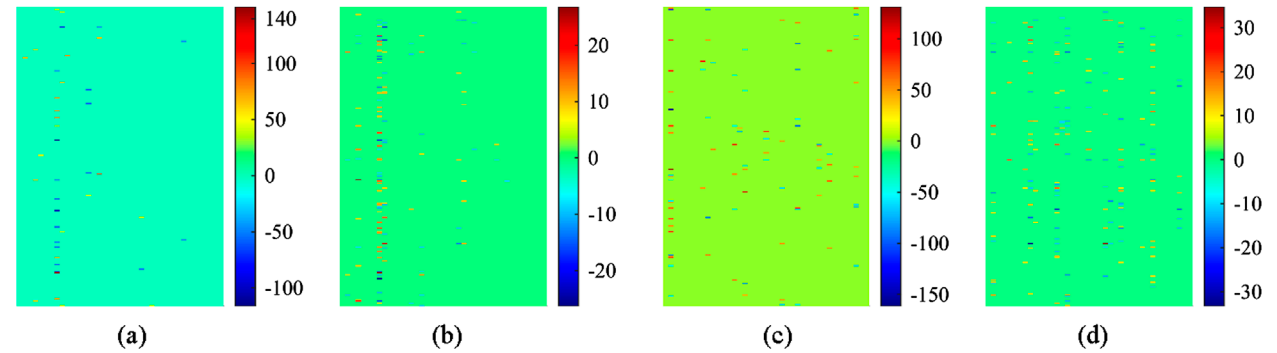

FIGURE 12
Coefficient visualization of the visible and infrared images. **(a)** SR coefficients of the visible image **(b)** SLR coefficients of the visible image **(c)** SR coefficients of the infrared image **(d)** SLR coefficients of the visible image.

method in terms of maintaining image information and suppressing artificial features. Figures 5c,d,j show that the overall images are smooth and the fused images are close to the source image pairs, but some details are not well preserved. The white zone beside the car is unobtrusive in Figures 5d,j, it is also hard to observe the trees in front of the farthest house. This indicates that DLF and RestNet lose more details while they capture more structural information; this is consistent with the metric results.

4.3 Experiment results on TNO dataset

We carried out another experiment on the TNO dataset to further examine the performance of our method; only nine methods that performed well on the VIFB dataset are compared in this experiment. The results of metrics are listed in Table 3. Our method receives three best, 2 seconds, one-third, and outperforms the others. It is easy to find that the advantages remain in information theory-based metrics, image feature-based metrics, and human visual perception, which is consistent with the results on the VIFB dataset. The performance of Q^{CV} declines compared with the results on the VIFB dataset. But our method continues to perform better in Q^{CB}; both Q^{CB} and Q^{CV} belong to human visual perception metrics, so we can still believe that our method has good ability to capture the major features in the human visual system. RMSE and SSIM are not prominent, this phenomenon can also be observed in other superior methods, including CNN, GFCE, LatLRR, RP_SR, and MST_SR. DLF and ResNet still obtain the best results in RMSE and SSIM. According to the STDs, smaller values are observed for the CE, EN, PSNR, RMSE, SSIM, and SD metrics, slightly poorer performance is seen for the MI and Q^{CB} metrics, and the remaining metrics exhibit performances close to the average level.

Furthermore, Figure 6 displays the metrics of 20 image pairs that were selected at random from the TNO dataset. The metrics of several image pairs also varied significantly from one another. Overall, our method performs better in CE, EN, Q^{CB}, and SD, as Figures 6b,d,h,k show. Even though our method's average CE in Table 3 is not the best, it outperforms NSCT_SR in more image

pairs, which proves the stability of our method. Additionally, our method's performance is extremely close to the best ones in PSNR, $Q^{AB/F}$, Q^{CV}, RMSE, and SSIM, as Figures 6f,g,i,j,m show, while our method's performance is mediocre in AG, EI, and SF.

Figure 7 displays the fused images of different methods on the 29-th image pair. There are still some discontinuities in NSCT_SR, as Figure 7b shows. In GFCE and RP_SR, there are several unexpected new features about the window, as Figures 7e,j show. DLF and RestNet contain more information than the infrared image and the bright message in the sky is lost, as Figures 7d,i show. DLF and RP_SR lose some details about the letter box, as Figures 7d,i show. MSD and LatLRR save more information about the cloud and the visible and infrared features are very clear. Our method only contains a little of the cloud features of the visible image, but the dark and bright feature of the sky is retained more, as Figure 7l shows.

4.4 Experiment results on the LLVIP dataset

In order to further investigate the performance of our method, we conducted another experiment on the LLVIP dataset; the same methods that were used on TNO dataset are compared in this experiment. The image pairs from the LLVIP dataset are captured in a very dark environment. Table 3 lists the results of the metrics, our method receives one best and 3 seconds. Although there are fewer top three metric results, it is worth noting that the metric results of our method are very close to the better ones. Overall, LatLRR obtains the better scores. DLF and ResNet still share good results with RMSE and SSIM while the other metric results are not outstanding. According to the STDs, smaller values are observed for the $Q^{AB/F}$, RMSE, and SF metrics, slightly poorer performance is seen for the EN, PSNR and SD metrics, while the remaining metrics exhibit performances close to the average level. The assessment scores for the fusion results of all methods exhibit significant fluctuations across different image data sets, due mainly to various degrees of statistical complexities amongst the data sets. However, it is observed that the performance of the proposed algorithm exhibits relatively smaller fluctuations across most evaluation metrics, as indicated by the smaller standard deviations in Tables 2–4, indicative of the better stability in our proposed method.

Additionally, Figure 8 shows the metrics of 20 image pairs selected at random from the LLVIP dataset. Our method yields the best outcome in $Q^{AB/F}$, as Figure 8g shows. As shown in Figures 8b,e,i,m, our method's performance is nearly as good as the best in CE, MI, Q^{CB}, Q^{CV}, and SSIM. In AG, EI, EN, PSNR, RMSE, SD, and SF, our method performs mediocrely. LatLRR performs better overall in AG, EI, SD, and SF but exhibits subpar results in PSNR, RMSE, and $Q^{AB/F}$, as Figures 8f,g,j show. Our approach presents better balanced results in all metrics than LatLRR, despite a slight performance drop when compared to the VIFB and TNO datasets. DLF also shows better performance in PSNR, RMSE, and SSIM, but its performance is not exceptional in CE, EI, EN, Q^{CB}, and SD.

Figure 9 shows the selected fusion results on the 3-th image pair, which was taken in a very dim setting. It is easy to observe in Figure 9e that GFCE lost the dark information. The wall and pedestrians still have some noticeable discontinuities in NSCT_SR, and RP_SR also exhibits this problem. Due to the loss of some information from the visible source image, the words on the wall are not discernible in DLF and ResNet. The fusion performance is close in Figures 9c,f,g,k,l, CNN and LatLRR achieve superior results in motorbike light out of the five results. Our approach and MST_SR preserve more dark information while losing the visible source image's bright light information.

4.5 Discussion

4.5.1 Runtime comparison

Table 5 lists the runtime of various algorithms on three datasets. It is evident that there are substantial differences in runtime across the various methods. ResNet performs better than other methods on VIFB and LLVIP datasets, while GFCE has a lower runtime cost on TNO dataset. NSCT_SR takes the longest on all three datasets, while LatLRR also needs to take a lengthy time to fuse a single image pair. Additionally, we can see that a particular method's runtime varies greatly depending on the dataset; this is particularly true for LatLRR and NSCT_SR. The size of source image pairs from LLVIP are larger than those from the other two datasets. It theoretically takes more time on the LLVIP dataset. On the other hand, even though the size of the source image pairs from TNO are larger than that from VIFB, some algorithms, including DLF, GFCE, MSD, LatLRR, NSCT_SR, and RP_SR, have shorter runtimes on the TNO dataset. This is because different datasets have varying levels of complexity. When complexity is higher, feature extraction needs more time. Overall, GFCE and ResNet share computational efficiency. Our method's runtime is comparable to that of DLF, MSD, RP_SR, and MST_SR. However, it can be observed that methods with notable advantages in runtime, such as ResNet and GFCE, perform far worse than the proposed method in image fusion. Conversely, methods that achieve comparable image fusion results to the proposed method, such as LatLRR and NSCT-SR, have a significantly higher time cost. As for MST-SR, although its time cost is slightly better than the proposed method on the VIFB and TNO datasets, its average time cost advantage is only 0.427 s, and its image fusion results across all datasets are far inferior to the proposed method. In conclusion, the proposed method, while maintaining an advantage in image fusion quality, also generally achieves improved runtime performance compared to methods with similar fusion quality.

Although LatLRR can obtain good fusion performance, the runtime cost is significantly expensive. NSCT_SR's fusion performance is not exceptional and its runtime is lengthy. The aforementioned results lead us to the conclusion that there needs to be a compromise between fusion performance and runtime.

4.5.2 Analysis of parameter sensitivity

In order to analyze the parameter sensitivity of our method, both α and λ in Equation 4 are set among a discrete set $\{1e^{-7}, 1e^{-6}, 1e^{-5}, 1e^{-4}, 1e^{-3}, 1e^{-2}, 1e^{-1}, 1, 10, 100\}$, and the results of 13 metrics are shown in Figure 10. The smaller the value, the better the result for three metrics: CE, Q^{CV}, and RMSE. When α and λ

are in the range of $1e^{-7}$ to 10, the three metrics remain small and stable. Meanwhile, other metrics hold a high level. This illustrates that our method has good robustness, and it is easy to find a parameter pair that results in good performance. In addition, there are several noteworthy results. Firstly, the metrics, including AG, EI, EN, MI, Q^{CB}, SF, SD, CE, and Q^{CV}, improve when α is set to 100. However, PSNR is lower. Therefore, α cannot exceed 10 for overall performance, as Figure 10f shows. Second, RMSE does not change with the variation of parameters and always maintains a modest value; this proves that our sparse low-rank representation can be well implemented, and the error of image reconstruction is very close to zero. Third, SSIM improves when λ surpasses 10, but other metrics worsen. Therefore, λ cannot exceed 10 for overall performance.

In order to analyze the influence of decomposition levels, we set it in the discrete set {2, 3, 4, 5, 6, 7} and the results of different metrics are shown in Figure 11. When the decomposition levels increase, there are generally three types of change trends: increasing, decreasing, and first increasing then decreasing. Metrics, including CE, PSNR and $Q^{AB/F}$, show the increasing trend and maintain good results when the decomposition level is 4. As for AG, EI, EN, Q^{CV}, and RMSE, they decrease as the decomposition level is increasing. AG, EI, and EN receive the best results when the level is 2, while the value of Q^{CV} and RMSE are near the best result when the level exceeds 3. The rest of the metrics, including MI, Q^{CB}, SD, SF, and SSIM, show the trend of first increasing then decreasing. They obtain the best results when the decomposition level is 4. In conclusion, when the decomposition level is 4, metrics such as CE, MI, PSNR, $Q^{AB/F}$, Q^{CB}, Q^{CV}, RMSE, SD, SF, and SSIM, can obtain the best results or keep very near the best results while the other metrics still maintain relatively good results. Therefore, the decomposition level is recommended to be 4.

4.5.3 Effectiveness analysis of low-rank constraint

To further analyze the effectiveness of low-rank constraints in extracting image information, Figure 12 presents the visualization results of sparse coefficients and sparse low-rank coefficients. First, it can be observed that the number of nonzero elements in the sparse coefficients is relatively small, with larger magnitude values. This indicates that, in the sparse representation of infrared and visible images, greater weight is assigned to key features, while some detailed features tend to be overlooked. In contrast, due to the presence of the low-rank constraint, the sparse low-rank coefficients contain more nonzero elements, with values an order of magnitude lower than those of the sparse coefficients. This suggests that the sparse low-rank representation can extract more image features compared to the sparse representation. Furthermore, since low-pass band fusion is performed based on the maximum L1-norm fusion rule, when an image contains numerous features, the fusion is influenced not only by a few key features but also by other features that play a decisive role. Therefore, in summary, compared to MST-SR, the proposed infrared and visible image fusion method based on sparse low-rank representation can retain key features while extracting more detailed features, thereby preserving as much information from the original images as possible in the fused image. This also explains why the proposed method performs well in both information theory-based metrics and image feature-based metrics, further validating that the introduced low-rank constraint effectively enhances the quality of infrared and visible image fusion.

5 Conclusion

In this paper, we combine sparse and low-rank constraints to present a novel visible and infrared image fusion method. Source images are firstly transformed into another domain to calculate their low-pass and high-pass bands by LP. Second, low-pass bands are represented with some sparse and low-rank coefficients. Then, specific fusion rules are adopted to fuse the low-pass and high-pass bands. Low-pass parts are a smooth version of source images; they are fused based on sparse low-rank representation. High-pass parts contain more details and the fusion is conducted based on the max-absolute rule. Finally, an inverse LP is conducted to obtain the fused image. Our method is validated on three public datasets. The results show that our method performs better in the three kinds of metrics: information theory-based, image feature-based, and human perception-based metrics. This means that low-rank constraints can effectively improve the performance of capturing details. Furthermore, our method obtains average performance in runtime cost and achieves relatively better balance between fusion performance and runtime cost. Our method also shows good parameter robustness; a good result can be obtained in a wide range of parameters.

Data availability statement

The VIFB dataset can be download from https://github.com/xingchenzhang/VIFB/tree/master (accessed on 17 June 2025). The TNO dataset can be download from https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029 (accessed on 17 June 2025). The LLVIP dataset can be download from https://bupt-ai-cz.github.io/LLVIP/ (accessed on 17 June 2025).

Author contributions

YZ: Writing – original draft, Data curation, Formal Analysis, Investigation, Methodology, Software, Validation. JW: Project administration, Resources, Supervision, Writing – review and editing. BY: Funding acquisition, Supervision, Writing – review and editing. HC: Methodology, Validation, Writing – review and editing. JF: Methodology, Writing – review and editing. ZW: Writing – review and editing. SY: Writing – review and editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. This research was funded by the Yunnan Provincial and Municipal Integration Special Project under Grant 202302AH360002 by the Young and Middle-Aged Academic and Technical

Leaders Reserve Talents Project of Yunnan Province under Grant 202305AC160062 and by the Scientific Research Fund Project of Yunnan Provincial Department of Education under Grant 2024J0077.

Conflict of interest

Author JF was employed by Guangxi Huasheng new material Co., Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

References

1. Zhang X. Benchmarking and comparing multi-exposure image fusion algorithms. *Inf Fusion* (2021) 74:111–31. doi:10.1016/j.inffus.2021.02.005

2. Zhang Y, Zhang L, Bai X, Zhang L. Infrared and visual image fusion through infrared feature extraction and visual information preservation. *Infrared Phys Technol* (2017) 83:227–37. doi:10.1016/j.infrared.2017.05.007

3. Bavirisetti D, Xiao G, Liu G. Multi-sensor image fusion based on fourth order partial differential equations. In: *Proceedings of the 20th international conference information fusion*. China: Xi'an (2017). p. 10–3.

4. Ma J, Chen C, Li C, Huang J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Inf Fusion* (2016) 31:100–9. doi:10.1016/j.inffus.2016.02.001

5. Li H, Wu X. Infrared and visible image fusion using latent low-rank representation. *arXiv preprint* (2018). Available online at: https://arxiv.org/abs/1804.08992.

6. Bavirisetti D, Dhuli R. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Phys Technol* (2016) 76:52–64. doi:10.1016/j.infrared.2016.01.009

7. Bavirisetti D, Dhuli R. Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform. *IEEE Sens J* (2016) 16:203–9. doi:10.1109/jsen.2015.2478655

8. Shreyamsha Kumar BK. Image fusion based on pixel significance using cross bilateral filter. *Signal Image Video P* (2013) 9:1193–204. doi:10.1007/s11760-013-0556-9

9. Zhou Z, Dong M, Xie X, Gao Z. Fusion of infrared and visible images for night-vision context enhancement. *Appl Opt* (2016) 55:6480. doi:10.1364/ao.55.006480

10. Li S, Kang X, Hu J. Image fusion with guided filtering. *IEEE Trans Image Process* (2013) 22:2864–75. doi:10.1109/TIP.2013.2244222

11. Bavirisetti D, Xiao G, Zhao J, Dhuli R, Liu G. Multi-scale guided image and video fusion: a fast and efficient approach. *Circuits Syst Signal Process* (2019) 38:5576–605. doi:10.1007/s00034-019-01131-z

12. Naidu V. Image fusion technique using multi-resolution singular value decomposition. *Def Sci J* (2011) 61:479. doi:10.14429/dsj.61.705

13. Qi B, Bai X, Wu W, Zhang Y, Lv H, Li G. A novel saliency-based decomposition strategy for infrared and visible image fusion. *Remote Sens.* (2023) 15:2624. doi:10.3390/rs15102624

14. 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* (2016) 30:15–26. doi:10.1016/j.inffus.2015.11.003

15. Liu Y, Liu S, Wang Z. A general framework for image fusion based on multi-scale transform and sparse representation. *Inf Fusion* (2015) 24:147–64. doi:10.1016/j.inffus.2014.09.004

16. Ma J, Zhou Z, Wang B, Zong H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Phys Technol* (2017) 82:8–17. doi:10.1016/j.infrared.2017.02.005

17. Chen W, Miao L, Wang Y, Zhou Z, Qiao Y. Infrared–visible image fusion through feature-based decomposition and domain normalization. *Remote Sens.* (2024) 16:969. doi:10.3390/rs16060969

18. Li L, Lv M, Jia Z, Jin Q, Liu M, Chen L, An effective infrared and visible image fusion approach via rolling guidance filtering and gradient saliency map. *Remote Sens* (2023) 15:2486. doi:10.3390/rs15102486

19. Liu Y, Chen X, Cheng J, Peng H, Wang Z. Infrared and visible image fusion with convolutional neural networks. *Int J Wavelets Multiresolut Inf Process* (2018) 16:1850018. doi:10.1142/s0219691318500182

20. Li H, Wu X, Kittler J. Infrared and visible image fusion using a deep learning framework. In: *Proceedings of the 24th international conference on pattern recognition (ICPR)*. Beijing, China (2018). p. 20–4.

21. Li H, Wu X, Durrani T. Infrared and visible image fusion with ResNet and zero-phase component analysis. *Infrared Phys Technol* (2019) 102. doi:10.1016/j.infrared.2019.103039

22. Zhang X, Ye P, Xiao G. VIFB: a visible and infrared image fusion benchmark. In: *Proceedings of the 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)*. Seattle, WA, USA (2020).

23. Wei Q, Liu Y, Jiang X, Su B, Yu M. DDFNet-A: attention-based dual-branch feature decomposition fusion network for infrared and visible image fusion. *Remote Sens.* (2024) 16:1795. doi:10.3390/rs16101795

24. Xu Y, Fang X, Wu J, Li X, Zhang D. Discriminative transfer subspace learning via low-rank and sparse representation. *IEEE Trans Image Process* (2016) 25:850–63. doi:10.1109/tip.2015.2510498

25. Xue J, Zhao Y, Bu Y, Liao W, Chan J, Philips W. Spatial-spectral structured sparse low-rank representation for hyperspectral image super-resolution. *IEEE Trans Image Process* (2021) 30:3084–97. doi:10.1109/tip.2021.3058590

26. Donoho D. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. Commun. *Pure Appl Math* (2006) LIX:0797–829. doi:10.1002/cpa.20131

27. Ahmed J, Gao B, Woo W, Zhu Y. Ensemble joint sparse low-rank matrix decomposition for thermography diagnosis system. *IEEE Trans Ind Electron* (2021) 68:2648–58. doi:10.1109/tie.2020.2975484

28. Lin Z, Chen M, Ma Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv*:2011. arXiv: 1109.5055. doi:10.48550/arXiv.1009.5055

29. Boyd S, Parikh N, Chu E, Peleato B, Eckstein J. Distributed optimization and statistical learning via the alternating direction method of multipliers. In: *Now foundations and trends*. Netherlands: Delft (2010). p. 1–122.

30. Donoho DL. De-noising by soft-thresholding. *IEEE Trans Inf Theor* (1995) 41:613–27. doi:10.1109/18.382009

31. Cai JF, Candès EJ, Shen ZW. A singular value thresholding algorithm for matrix completion. *SIAM J Optimiz* (2010) 20:1956–82. doi:10.1137/080738970

32. Liu G, Lin Z, Yan S, Sun J, Yu Y, Ma Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans Pattern Anal Mach Intell* (2013) 35:171–84. doi:10.1109/tpami.2012.88

33. Li H, Manjunath B, Mitra S. Multisensor image fusion using the wavelet transform. *Graph Models Image Process* (1995) 57:235–45. doi:10.1006/gmip.1995.1022

34. Toet A. The TNO multiband image data collection. *Data Brief* (2017) 15:249–51. doi:10.1016/j.dib.2017.09.038

35. Jia X, Zhu C, Li M, Tang W, Zhou W. LLVIP: a visible-infrared paired dataset for low-light vision. In: *Proceedings of the 2021 IEEE/CVF international conference on computer vision workshops (ICCVW)*. Montreal: BC, Canada (2021).