frontiers | Frontiers in Energy Research

TYPE Methods
PUBLISHED 30 August 2022
DOI 10.3389/fenrg.2022.1001450

OPEN ACCESS

EDITED BY
Lianbo Ma,
Northeastern University, China

REVIEWED BY
Cheng Xie,
Harbin Institute of Technology, China
Liu Jie,
Ministry of Industry and Information Technology, China
Hongjiang Wang,
Shenyang Institute of Engineering, China

*CORRESPONDENCE
Jing Zhang,
jzhangnuc@163.com

SPECIALTY SECTION
This article was submitted to Smart Grids,
a section of the journal
Frontiers in Energy Research

RECEIVED 23 July 2022
ACCEPTED 08 August 2022
PUBLISHED 30 August 2022

CITATION
Zhang J, Tang B and Hu S (2022), Infrared and visible image fusion based on particle swarm optimization and dense block.
Front. Energy Res. 10:1001450.
doi: 10.3389/fenrg.2022.1001450

Infrared and visible image fusion based on particle swarm optimization and dense block

Jing Zhang[1,2]*, Bingjin Tang[1] and Shuai Hu[1]

[1]School of Software, North University of China, Taiyuan, China, [2]Shanxi Military and Civilian Integration Software Engineering Technology Research Center, North University of China, Taiyuan, China

Infrared and visible image fusion aims to preserve essential thermal information and crucial visible details from two types of input images to generate an informative fusion image for better visual perception. In recent years, several hybrid methods have been applied in the field of infrared and visible image fusion. In this paper, we proposed a novel image fusion method based on particle swarm optimization and dense block for visible and infrared images. Particle swarm optimization is utilized to optimize the weighting factors of the coefficients obtained by discrete wavelet transform, then the coefficients are fused with the optimum weight to obtain the initial fusion image. The final fusion image is created by integrating the first fused image with the input visible image using a deep learning model, in which dense block is utilized for better feature extraction ability. The results of comparison experiments demonstrate that our method produces fusion images with richer details and texture features, and the fused image reduces the artifacts and noise.

KEYWORDS

image fusion, infrared and visible images, particle swarm optimization, dense block, deep learning

Introduction

The accurate operation and prompt detection of abnormal conditions are crucial for the safety of the power grid. In this way, the fusion of power grid images has aroused increasing attention in recent years for the smart grids (Ma et al., 2021), and the majority of images are utilized to spot anomalies in the personnel and the devices. The aim of the image fusion technique is to combine meaningful information from several sensors into a single informative fused image. By extracting all complementary features from the input images and avoiding any irregularities in the final fused image, effective image fusion preserves critical information. Fusion techniques for infrared and visible image have gained increasing attentions in the recent research on multi-sensor fusion field. Both visible and infrared sensors have advantages and disadvantages of their own. Infrared images are more sensitive to the thermal and radiative information, thus heat source targets have higher pixel values in infrared images and are displayed as high brightness, which is distinct to observe. However, due to the infrared sensor imaging limitations, infrared images contain less texture information. Visible images contain rich details since visible image sensors capture reflected light from objects. Thus, the fusion of images

obtained from both sensors can overcome the restrictions of a single sensor and produce many complementary characteristics. Fusion image aim to combine the strengths of source images for better visual effects results and provide rich information to enhance decision-making. Therefore, fusion algorithms have been adopted in various scenarios for different applications such as object tracking (Zhang et al., 2020), surveillance (Paramanandham and Rajendiran, 2018), face verification (Raghavendra et al., 2011).

To implement the fusion of visible and infrared images, a number of methods have been developed over the last few decades. These algorithms can be categorized into conventional methods and deep learning-based methods.

For conventional methods, several multi-scale decomposition methods have been attempted, the representative examples are the pyramid and wavelet transform based methods. For instance, Burt and Adelson (1985) firstly used Laplacian pyramid to encode images. Du et al. (2016) proposed a union Laplacian pyramid based image fusion method which used Laplacian pyramid to transform input images into multi-scale representations and inversed pyramid to obtain the fused image. Shen et al. (2014) proposed a boosting Laplacian pyramid method to fuse multiple exposure images. Li et al. (2018) proposed a feature fusion strategy using Gaussian pyramid. In contrast to the multi-scale pyramid transform, the wavelet transform coefficients are mutually independent (Ma et al., 2019a). Li et al. (2002) introduced a wavelet transform based image fusion method for multi-sensor images, which showed certain strengths over Laplacian pyramid based fusion methods. As a method of fusion task, the dual-tree discrete wavelet transform is applied for thermal image fusion (Madheswari and Venkateswaran, 2016). In the work of (Liu et al., 2015), this issue is addressed by using a general image fusion framework in which multi-scale transform is used for decomposition and reconstruction. However, the accuracy of conventional approaches is inadequate, which results in subpar fusion outcomes.

With respect to deep learning-based image fusion tasks, convolutional neural networks (CNNs) have performed a number of impressive results. Liu et al. (2017) proposed a fusion method for multi-focus images in which CNN is applied for the first time. Li et al. (2018) introduced a deep learning network using a fixed VGG-19 network to extract detail content features of infrared and visible images. After that, Li and Wu, 2019 operated dense block (Li and Wu, 2019) to preserve useful information of input images. Moreover, by implicitly executing feature extraction and reconstruction process, the generative adversarial network (GAN) methods are also well-performed in visible and infrared image fusion tasks (Zhang et al., 2021). For instance, Ma et al. (2019) firstly introduced GANs by considering infrared and visible image fusion task as an adversarial problem. After that, they utilized GAN to preserve rich spectral information in remote sensing images (Ma et al., 2020). Li et al. (2020) proposed a dual discriminator generative adversarial network to keep more details and textures in the fused image. However, GAN methods have weaknesses on the balance of generator and discriminator. Additionally, autoencoder (AE)-based methods also exhibits good performances, where encoder extracts features from input images and decoder reconstruct the features to obtain fusion results, such as DenseFuse (Li and Wu, 2019) and VIF-Net (Hou et al., 2020).

Furthermore, evolutionary computation based optimization is a new trend in recent research. Many related algorithms are utilized in different systems to implement multiple scales and objectives optimization, such as particle swarm optimization (PSO) (Madheswari and Venkateswaran, 2016; Paramanandham and Rajendiran, 2018), grey wolf (Daniel et al., 2017) artificial bee colony (Chatterjee et al., 2017), differential evolution (Kaur and Singh, 2020), and they all perform good results. Appropriate weight parameters generated by optimization algorithms can be used to produce fusion image with better quality, and the fusion images assist the implementation of subsequent works in a smooth way.

However, the limitation of conventional methods is apparent in the treat of image details. Most deep learning based methods generally have weaknesses on training consumption and are prone to overfitting problems. Thus, some hybrid image fusion methods consequently exist to avoid artifact and blockiness. For instance, PSO is used to optimize weights for fusing discrete wavelet transform coefficients (Madheswari and Venkateswaran, 2016) and discrete cosine transform coefficients (Paramanandham and Rajendiran, 2018) in the fusion process. Differential evolution is used to enhance feature selection (Kaur and Singh, 2020). Fu et al. (2020) proposed a multimodal medical image fusion method using Laplacian pyramid combined with CNN. Wang et al. (2021) proposed a visible and infrared image fusion method based on Laplacian pyramid and GAN. These hybrid methods have improved fusion results in their respective applications.

Therefore, in this paper, we proposed an innovative hybrid image fusion method by combining conventional image fusion method with swarm intelligence technique and deep learning model. Firstly, the input images are fused using discrete wavelet transform (DWT) with a weight factor optimized by PSO. Due to power grid images necessitate greater observational detail, the final fused image is then produced by synthesizing the initial fused image with the input visible image using a dense block. With these fusion architectures, fused images can preserve more useful details and textures that support the further supervisory and decisional demands.

The contributions of this paper are summarized as follows:

1) A visible and infrared image fusion method based on the swarm intelligence technique is proposed. Specifically, PSO is used to optimize the weighting factors of coefficients obtained

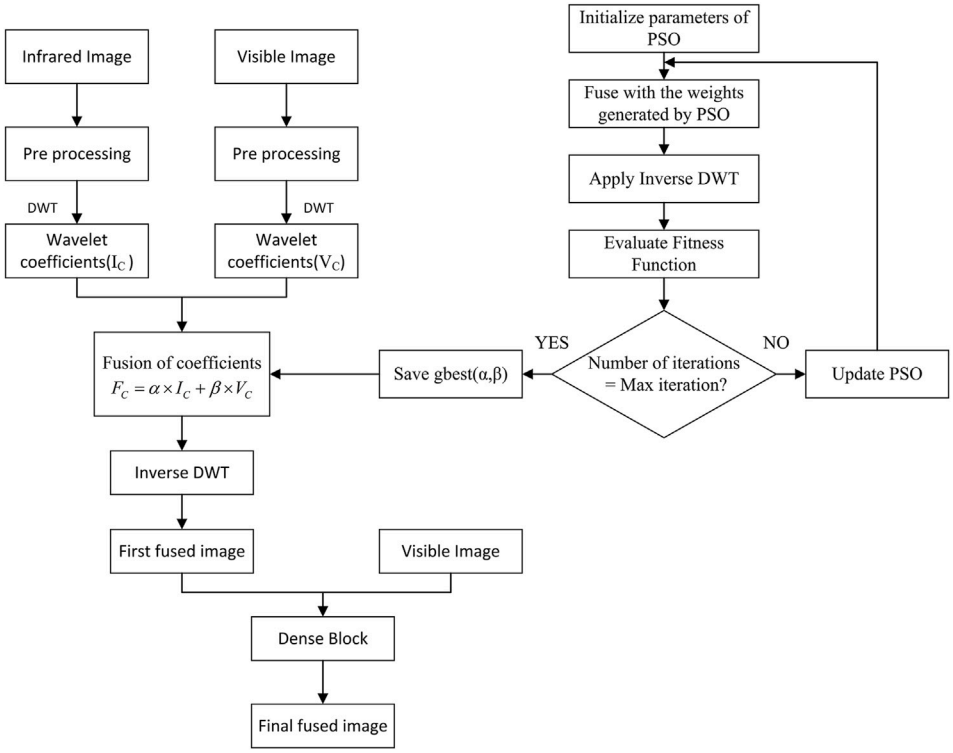

FIGURE 1
The framework of proposed image fusion method.

by DWT, then the coefficients are fused with the optimum weight to obtain the fusion image.

2) An image fusion method using deep learning model is proposed for preserving richer details and textures. To better maintain features from the visible image, dense block is used in feature extraction process to obtain the fusion image with high quality for better visual effect.
3) Based on the two mentioned methods, a novel image fusion method is proposed in this paper, which contains two parts. In the first part, the input visible and infrared images are decomposed to coefficients by DWT, and then fused with the optimal weight generated by PSO. Inverse DWT is utilized to obtain the initial fusion image. In the second part, the first fused image is integrated with the input visible image through a deep learning model in which dense block is utilized for feature extraction. In the end, the final fused image is obtained. By comparing with representative methods, the evaluation results verified the effectiveness of our method.

The remainder of the paper is structured as follows: *Methods Section* introduces the specific techniques that we use in our method. In *Proposed Fusion Method Section*, details on the proposed fusion method are provided. Experimental results are given in *Experiments and Results Section*, followed by the conclusions in *Conclusion Section*.

Methods

This section introduces the techniques we use, including DWT, PSO, and dense block, in the proposed image fusion framework.

Discrete wavelet transform

DWT is a commonly used wavelet transform method for image fusion. The DWT has more advantages compared to pyramid methods and discrete cosine transform (DCT). For instance, DWT provides increased directional information, higher signal-to-noise ratios and no blocking artifacts than pyramid-based fusion (Lewis et al., 2007); DWT provides good localization and higher flexibility than DCT (Wu et al., 2016). DWT can divide the source image into several sub-bands including low-low, low-high, high-low and high-high bands, which contains the approximate coefficients, vertical details,

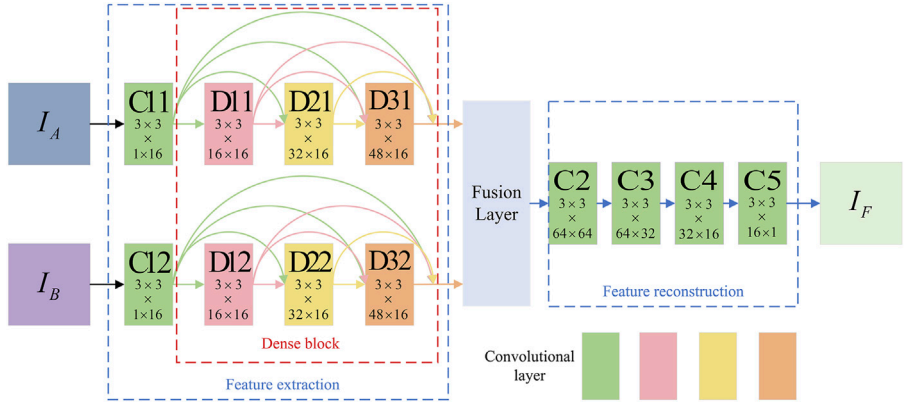

FIGURE 2
The architecture of the proposed network.

horizontal details and diagonal details coefficients, respectively (Shehanaz et al., 2021).

In this paper, DWT is used to decompose the input infrared and visible images into wavelet coefficients. The decomposed coefficients are fused with optimized weights generated by PSO, and by applying the inverse DWT to the fused coefficients, the fused image is obtained.

Particle swarm optimization

The PSO is a population-based optimization algorithm which was first introduced by Kennedy and Eberhart (1995) in 1995, its fundamental idea came from research on the flock feeding behavior of birds. The main purpose of PSO is to solve a problem by utilizing a population of randomly generated particles in the search space. Swarm is the term for the entire population, and particles are used to describe each individual. The swarm is denoted as $S = \{x_1, x_2, ..., x_N\}$, where x_i represents the i^{th} particle, and the population contains particle position as well as random velocity. Each particle position is updated by two best values called *pbest* and *gbest*. The *pbest* is the best position achieved by each particle so far, and the *gbest* is the global best position among particles in the swarm. The position and velocity of each particle are updated according to Eqs 1, 2,

$$V_{n+1}^{i} = \omega V^{i} + k_1 r_1 \left(pbest_n^i - X_n^i\right) + k_2 r_2 \left(gbest_n - X_n^i\right) \quad (1)$$

$$X_{n+1}^{i} = X_n^i + V_{n+1}^i \quad (2)$$

where V_{n+1}^{i} and V^{i} are the new and current velocity of the i^{th} particle, X_{n+1}^{i} and X_n^i are the new and current position of the i^{th} particle, ω is the inertia factor, large inertia weight tends to promote global search while small inertia weight tends to promote local search, k_1 and k_2 are two positive acceleration factors (both k_1, k_2= 2), r_1 and r_2 are numbers randomly generated between 0 and 1, $pbest_n^i$ is the *pbest* position of the i^{th} particle at time set n, $gbest_n$ is the *gbest* position of the optimizer at time set n.

In this paper, PSO generates the optimum fusion weight by searching the best solution, and the optimum weight is used to fuse the decomposed coefficients to enhance the fusion results.

Dense block

CNNs have achieved substantial progress in the field of image processing during the past few years. With the strong ability in feature extraction, convolutional neural networks have provided several novel ways for image fusion. Dense block is a key component of the DenseNet (Huang et al., 2017), the key concept is that for each layer, the feature mappings of all preceding layers are utilized as the input of the current layer while their feature mappings are used as the input for the subsequent layers, which form a full connection. The feature mappings extracted from each layer are available for the following layers. Thus, the features of input images can be effectively extracted and preserved, which establish the foundation for subsequent fusion. Advantages of dense block architecture are as follows: 1) this architecture can alleviate vanishing gradient and model degradation, which makes the network easily trained; 2) this architecture can enhance feature preservation; 3) this architecture reduces the number of parameters.

Proposed fusion method

The proposed fusion method is thoroughly discussed in this section. The overall framework of the proposed method is shown in Figure 1. As shown in Figure 1, the proposed framework has two main steps. Firstly, the input images are fused using DWT

TABLE 1 Architecture of the network.

Parts	Layers	Size	Stride	Input	Output	Activation
Feature extraction	Conv(C11,C12)	3 × 3	1	1	16	ReLU
	Conv(D11,D12)	3 × 3	1	16	16	ReLU
	Conv(D21,D22)	3 × 3	1	32	16	ReLU
	Conv(D31,D32)	3 × 3	1	48	16	ReLU
Feature reconstruction	Conv(C2)	3 × 3	1	64	64	ReLU
	Conv(C3)	3 × 3	1	64	32	ReLU
	Conv(C4)	3 × 3	1	32	16	ReLU
	Conv(C5)	3 × 3	1	16	1	ReLU

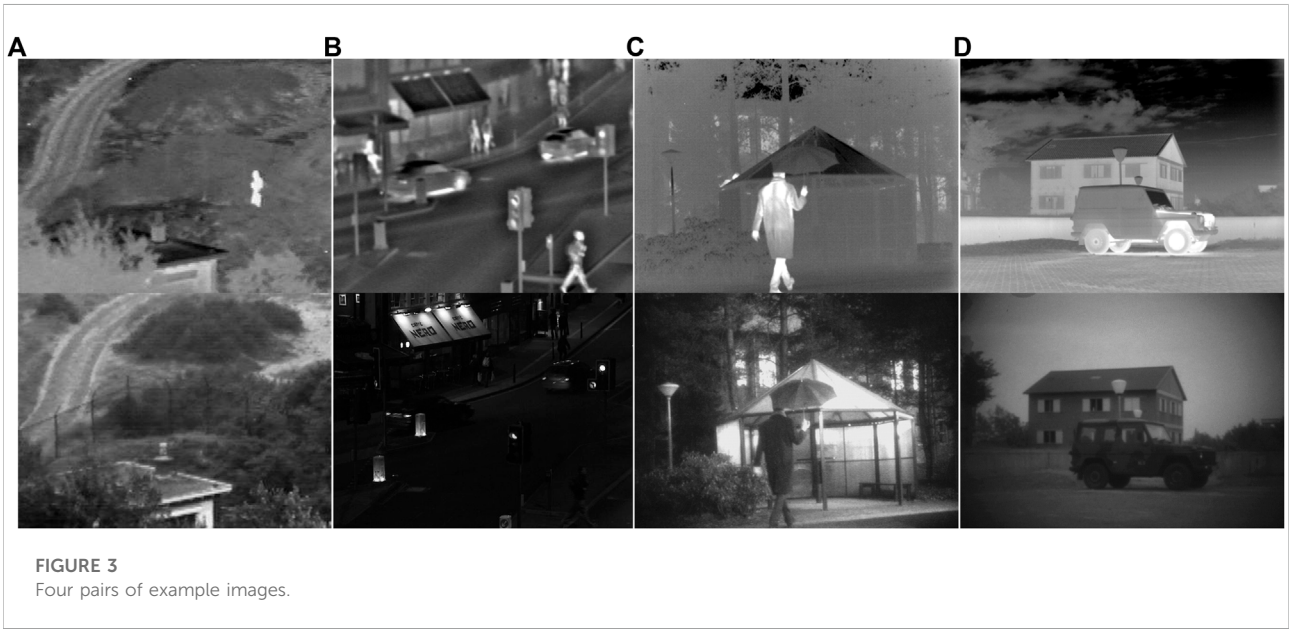

FIGURE 3
Four pairs of example images.

with a weight factor optimized by PSO to obtain the first fused image. Then, the final fused image is obtained by synthesizing the input visible image with the initial fused image using the dense block.

PSO based image fusion rule

Discrete wavelet transform (DWT) is used for decomposing the input infrared and visible images into approximate and detailed coefficients, which is used for fusion process to generate better results. The approximate and detailed coefficients of input infrared image are calculated by Eqs 3, 4, respectively.

$$C_{I\varphi}(m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y)\varphi(m,n) \quad (3)$$

$$C_{I\varphi}^{Hi}(m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y)\varphi^{Hi}(m,n) \quad (4)$$

The decomposed coefficients of input visible image are calculated by Eqs 5, 6, respectively.

$$C_{V\varphi}(m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} V(x,y)\varphi(m,n) \quad (5)$$

$$C_{V\varphi}^{Hi}(m,n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} V(x,y)\varphi^{Hi}(m,n) \quad (6)$$

where $M \times N$ represents the size of input images, $i = \{1, 2, 3\}$, C^{H_1}, C^{H_2} and C^{H_3} are the horizontal orientation, vertical orientation and diagonal orientation, respectively.

Since the approximate and detailed coefficients of input infrared and visible images decomposed by wavelet transform contain complementary information and features, the conventional fusion strategy such as average strategy and maximum select strategy may not provide

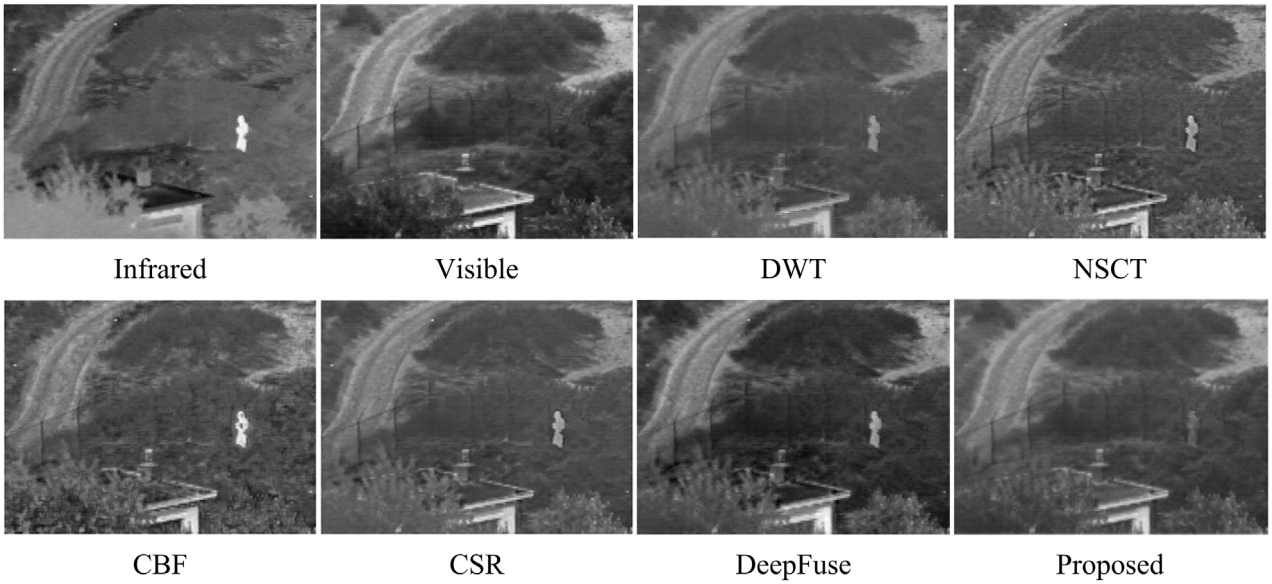

FIGURE 4
Fusion results with different comparison methods on image (a).

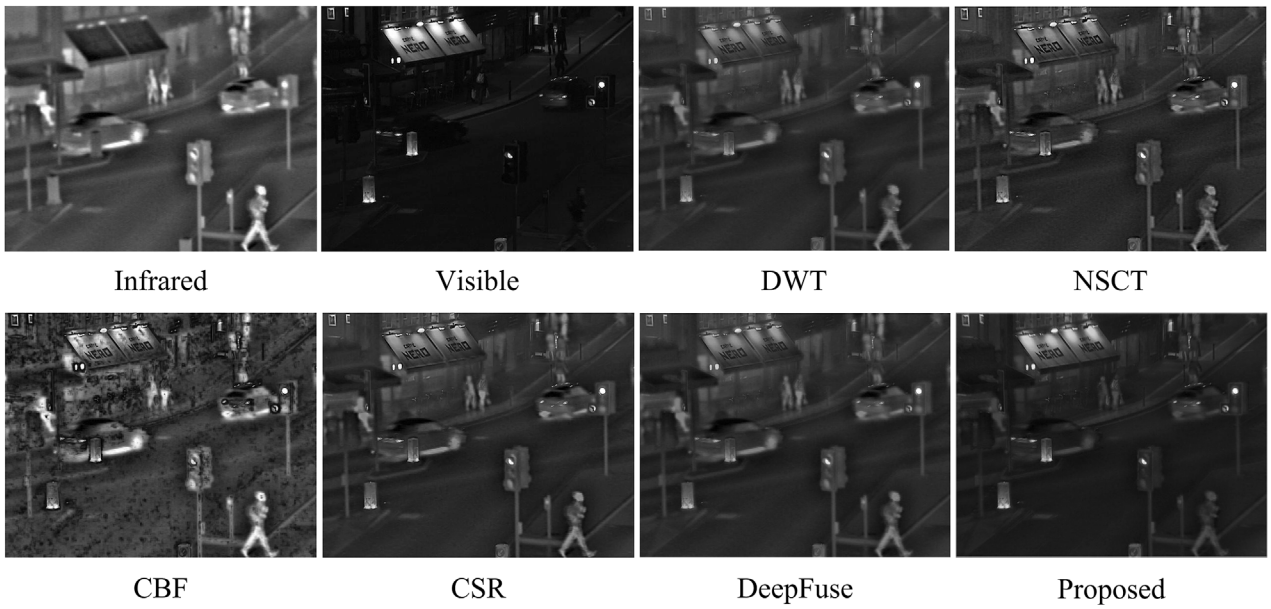

FIGURE 5
Fusion results with different comparison methods on image (b).

meaningful fusion of complementary saliency features. Thus, particle swarm optimization (PSO) is used to improve the fusion strategy with optimum weights to enhance entire fusion performance. The PSO algorithm computes with a population of random particles and updates the generations for the search of ideal option.

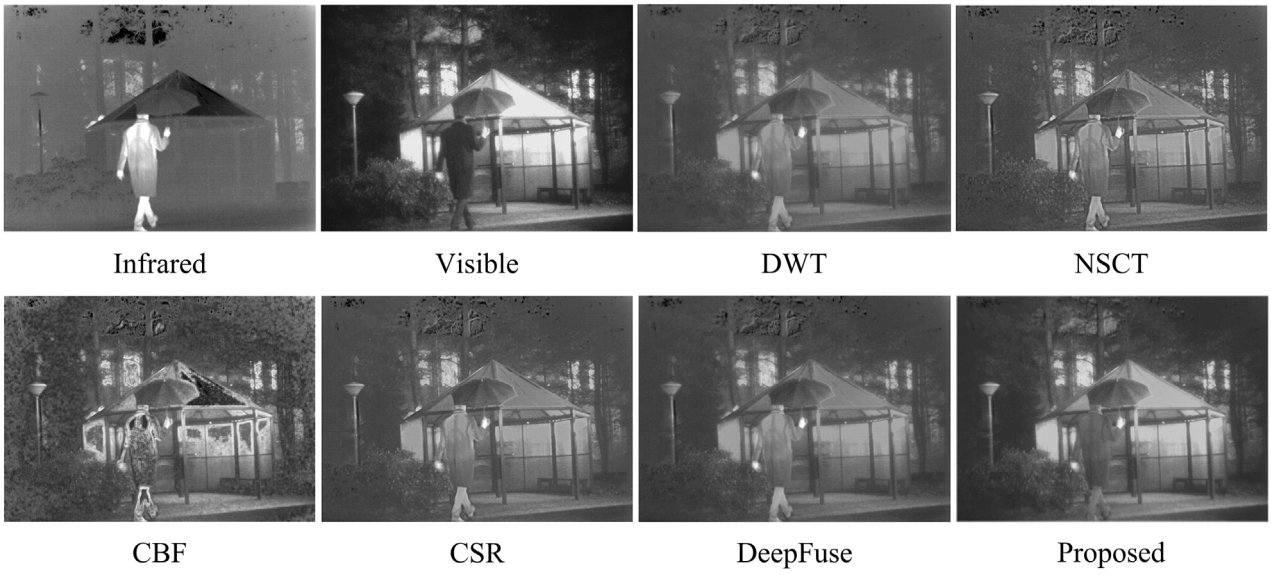

FIGURE 6
Fusion results with different comparison methods on image (c).

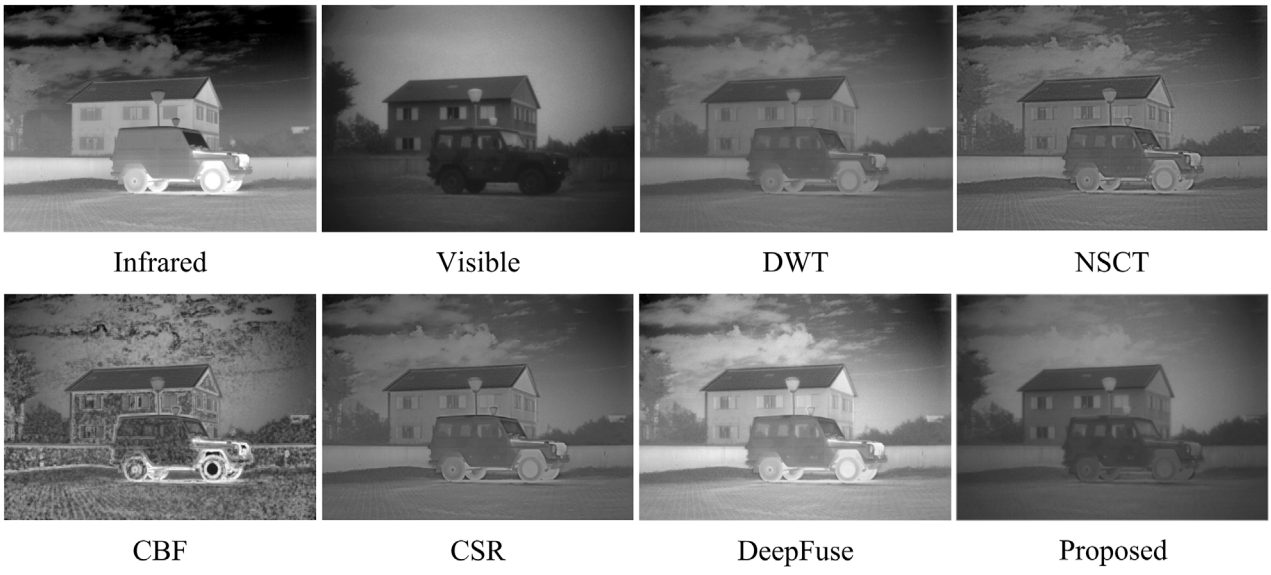

FIGURE 7
Fusion results with different comparison methods on image (d).

The optimum fusion weights are saved when the iteration ends and are used to fuse the decomposed coefficients based on the fusion rule. The fusion rule is defined as following Eq. 7.

$$F_C = \alpha \times I_C + \beta \times V_C \tag{7}$$

where F_c is the fused coefficients, I_c and V_c are the coefficients of infrared and visible images, α and β are the best weights optimized by PSO.

The first fused image is then generated by applying the inverse discrete wavelet transform to the fused coefficients.

Dense block based image fusion rule

Network architecture

The architecture of the proposed network is shown in Figure 2, and the network contains three main parts including feature extraction, fusion layer and feature reconstruction. The input images are denoted as I_A and I_B, respectively. The feature extraction part contains two channels to extract deep features. Channel A is made up of C11 and a dense block which is consists of D11, D21 and D31. Channel B contains C12 and a dense block that includes D12, D22 and D32. The first layer (C11, C12) contains 3 × 3 filters to extract rough features, and each dense block also have three convolutional layers which contain 3 × 3 filters. The input channel number of feature map is 16 in each convolutional layer in feature extraction part.

With this strategy, the input two channels have same architecture and share same weights, thus the feature extraction part has two advantages. First, the computational complexity and consumption are reduced. Second, the input images can be any size. Architecture of the network is outlined in Table 1. In fusion layer, we choose addition fusion strategy to directly concatenate features. The result of the fusion layer will be the input of feature reconstruction part. The feature reconstruction part contains another four convolutional layers (C2, C3, C4, C5), which also contain 3 × 3 filters, to obtain the fused result I_F.

Loss function

Loss function is used for computing the difference between the prediction and ground-truth to find appropriate parameters to reconstruct the input image more accurately and sufficiently. The structural similarity index (Wang et al., 2004) is a useful metric for comparing the structural similarity of two images. SSIM is sensitive to the perception of local structural changes which resembles the human visual system (HVS), and it contains three components: luminance comparison, structure comparison and contrast comparison. The three comparisons can be combined, and result in a specific form as $SSIM(A,F)$ and $SSIM(B,F)$, which are respectively defined as Eqs 8, 9.

$$SSIM(A,F) = \frac{(2\mu_A\mu_F + C_1)(2\sigma_{AF} + C_2)}{(\mu_A^2 + \mu_F^2 + C_1)(\sigma_A^2 + \sigma_F^2 + C_2)} \tag{8}$$

$$SSIM(B,F) = \frac{(2\mu_B\mu_F + C_1)(2\sigma_{BF} + C_2)}{(\mu_B^2 + \mu_F^2 + C_1)(\sigma_B^2 + \sigma_F^2 + C_2)} \tag{9}$$

In our network, the loss function SSIM are calculated by Eq. 10,

$$SSIM(A,B,F) = \frac{1}{2}(SSIM(A,F) + SSIM(B,F)) \tag{10}$$

where F is the fused image, and A, B are input images; μ_A, μ_B and μ_F represent the mean of A, B and F; μ_A^2, μ_B^2 and μ_F^2 represent the variance of A, B and F; σ_{AF} and σ_{BF} represent the joint variances of A, B and F. Both C_1 and C_2 are stable coefficients which are set as 0 in this work.

SSIM can characterize the differences between two images more precisely, and the larger SSIM value indicates the smaller difference between the images.

TABLE 2 Comparison results of evaluation metrics for image (a).

Method	EN	MI	SD	MSE	PSNR
DWT	6.2986	1.4527	7.8290	0.0167	41.8264
NSCT	6.4782	1.4758	8.0003	0.0133	42.8285
CBF	6.6181	1.4934	8.0733	0.0158	42.0793
CSR	6.4007	1.5533	7.9221	0.0134	42.7958
DeepFuse	**6.7653**	1.7104	8.3172	0.0179	41.5324
Proposed	6.5913	**2.5036**	**8.3314**	**0.0131**	**42.8870**

TABLE 3 Comparison results of evaluation metrics for image(b).

Method	EN	MI	SD	MSE	PSNR
DWT	5.9789	1.8314	6.9421	0.0307	39.1955
NSCT	6.0585	1.8502	6.9375	0.0248	40.1214
CBF	5.8384	1.2152	6.8541	0.0346	38.6801
CSR	5.9121	2.0877	6.9599	0.0246	40.1481
DeepFuse	5.9399	2.1950	7.0347	**0.0206**	40.9211
Proposed	**6.0659**	**2.5194**	**7.1588**	0.0205	**40.9459**

Training

In order to train the network to reconstruct the input image more accurately, we discard fusion layer and only consider feature extraction and reconstruction parts. The purpose of this training phase is to enhance the ability of the encoder and decoder in the autoencoder network to extract and reconstruct features. In training phrase, in order to improve the feature extraction ability of visible images to satisfy the requirements for better visual perception, we train the weights of encoder and decoder using the MS-COCO (Lin et al., 2014) dataset. The images in MS-COCO are all resized to 256 × 256 and transformed to gray scale in the pre-processing phrase, and the learning rate is set to 10^{-4}.

Experiments and results

In order to verify the performance of the proposed algorithm, we conduct extensive evaluation and comparison experiments. TNO database is used in this work, and four example images numbered as (a, b, c, d) are shown in Figure 3. The top row

TABLE 4 Comparison results of evaluation metrics for image(c).

Method	EN	MI	SD	MSE	PSNR
DWT	6.3672	1.6763	8.0665	0.0306	39.2135
NSCT	6.4668	1.6467	8.0793	0.0262	39.8776
CBF	6.3062	1.5914	8.2067	0.0324	38.9566
CSR	6.3993	1.7999	8.0709	0.0260	39.9096
DeepFuse	6.4134	2.1705	8.2431	**0.0227**	**40.5093**
Proposed	**6.7803**	**3.0469**	**8.7027**	0.0232	40.4037

TABLE 5 Comparison results of evaluation metrics for image(d).

Method	EN	MI	SD	MSE	PSNR
DWT	6.5126	1.4833	8.8842	0.0449	37.5427
NSCT	6.6107	1.4644	9.2933	0.0405	37.6900
CBF	6.5452	0.9850	9.4885	0.0512	36.9743
CSR	6.5897	1.6243	9.3471	**0.0404**	**38.0011**
DeepFuse	7.0314	2.0268	9.4639	0.0813	34.9668
Proposed	**7.3205**	**2.1473**	**10.5195**	0.0467	37.3682

represents infrared images, while the second row represents visible images.

In our experiments, we compare the proposed method with several typical fusion methods, including discrete wavelet transform (DWT) (Li et al., 2002), non-subsampled contourlet (NSCT) (Da Cunha et al., 2006), cross bilateral filter (CBF) (Shreyamsha Kumar, 2015), convolutional sparse representation (CSR) (Liu et al., 2016), and the DeepFuse method (DeepFuse) (Prabhakar et al., 2017). The fifilter size is also set as 3 × 3 for DeepFuse methods in our experiment. Figures 4–7 represents the contrast experiments results of the four example images, respectively. The evaluation results of comparison experiments are all conducted by MATLAB-2021a.

Evaluation metrics

To quantitatively evaluate the proposed method with the comparison methods, five commonly used quality metrics are utilized for evaluation. They are: entropy (EN), mutual information (MI), standard deviation (SD), mean square error (MSE), peak signal to noise ratio (PSNR).

Information entropy is a significant metric to assess the depth of visual information, which reflects the richness of information contained in the fused image, it can be calculated by Eq. 11,

$$EN = \sum_{i=1}^{L} p_i \log_2 p_i \tag{11}$$

where L represents the overall number of image pixels, p_i denotes the probability of distribution at each gray level. The more information that is present in the fused image, the larger the EN value, and the higher the quality of the fusion.

MI measures the information fused image obtained from the input images, it can be calculated as follows,

$$MI = MI_{AF} + MI_{BF} \tag{12}$$

$$MI_{AF}(a,f) = \sum_{a,f} P_{AF}(a,f) \log_2 \frac{P_{AF}(a,f)}{P_A(a)P_F(f)} \tag{13}$$

$$MI_{BF}(b,f) = \sum_{b,f} P_{BF}(b,f) \log_2 \frac{P_{BF}(b,f)}{P_B(b)P_F(f)} \tag{14}$$

where $P_A(a)$, $P_B(b)$, and $P_F(f)$ represents the normalized histogram of the image A, image B and the fused image, respectively. $P_{AF}(a,f)$ denotes the joint normalized histograms between the fused image and the source images A, $P_{BF}(b,f)$ represents the joint normalized histograms between the fused image and the source images B. The larger the MI value, the more information fused image preserves from input source images.

SD reflects the degree of dispersion in the image between each pixel value and the average value, it can be calculated by Eq. 15,

$$SD = \sqrt{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(F(m,n) - \bar{F}\right)^2} \tag{15}$$

where $F(m,n)$ is the pixel at relative coordinates, $\bar{F}$ is the average pixel value. The more dispersed the pixel level distribution, the higher the image contrast, and the better the fusion effect, the larger the SD value.

MSE measures the mean square error of the images, which is a reverse indicator evaluating the accuracy in integrating information from input images. PSNR indicates the distortion degree between the source images and the fused image. They can be calculated as follows.

$$MSE_{FA} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(F(m,n) - A(m,n)\right)^2 \tag{16}$$

$$MSE_{FB} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(F(m,n) - B(m,n)\right)^2 \tag{17}$$

$$MSE = \frac{1}{2}\left(MSE_{FA} + MSE_{FB}\right) \tag{18}$$

$$PSNR = 10\log_{10}\left(\frac{(M \times N)^2}{MSE}\right) \tag{19}$$

The fused image is less distorted and more similar to the source images with a higher PSNR value.

Subjective analysis

Figures 4–7 demonstrates the comparison between our method and five other methods. We can observe that images generated by DWT lost too much details, which leads to poor perception. For the

fused images of NSCT, some region details are blurred and lost some texture. More artificial noise and unclear saliency features are present in images fused by CBF, such as the person and sky shown in Figure 6 and Figure 7. In CSR, fused images have some obvious salient features loss and are darker than other images. As for DeepFuse method, more infrared image features are obtained and the fused images have too much brightness, which is unnatural for observation, such as the image shown in Figure 7. Compared with these five methods, our fusion method preserves abundant detail and texture information, the images are clearer, which are qualified for human visual perception.

Objective analysis

In this section, our method is compared with the five methods using evaluation metrics mentioned above. The objective evaluation results of five metrics achieved by the comparison methods and proposed method for the four groups of source images are shown in Table 2-5, in which the best results of the five indicators are marked in bold. The results demonstrate that our method outperforms the comparison methods in almost five assessment metrics. This indicates that the fused images obtained by our method have lower artifact and noise levels but richer details and texture features. The fused images are clearer and have a better visual effect.

In brief, the proposed method can more effectively extract and preserve detail information and texture features from source images and fuse with the best scale, which improves human visual perception. Additionally, the fusion results are validated using evaluation metrics, verifying the qualification for subsequent observation and detection.

Conclusion

In this paper, we introduce a novel method based on particle swarm optimization and dense block. This method combines conventional image fusion method with swarm intelligence technique and deep learning model. Firstly, DWT is utilized to decompose the input images into coefficients, then PSO is used to optimize the weighting factors and the coefficients are fused with the optimum weight to obtain the initial fusion image. To satisfy the observational requirements, the initial fusion image is then fused with the input visible image using a deep learning model in which dense block extracts and preserves rich image information. The experimental results demonstrate that the output fused images have abundant details and texture features, which are suitable for visual perception. The fused results are qualified for subsequent observation and detection.

Although the proposed method achieves good results, additional works in the fields of medical image and remote sensing image is required to broaden its application to multi-modality image fusion.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

JZ: Conceptualization, methodology, writing, and editing. BT: conceptualization, experiments and writing. SH: Experiments and data pre-processing.

Funding

This study is supported by the Technology Field Fund (2021-JCJQ-JJ-0726).

Conflict of interest

The 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.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Burt, P. J., and Adelson, E. H. (1985). "Merging images through pattern decomposition," in SPIE proceedings, San Diego, August 20, 1985. doi:10.1117/12.966501

Chatterjee, A., Biswas, M., Maji, D., Jana, D., Brojabasi, S., Sarkar, G., et al. (2017). "Discrete wavelet transform based V-I image fusion with artificial bee colony optimization," in 2017 IEEE 7th annual computing and communication workshop and conference (CCWC), Las Vegas, NV, USA, January 09–11, 2017. doi:10.1109/ccwc.2017.7868491

Da Cunha, A., Zhou, J., and Do, M. (2006). The nonsubsampled contourlet transform: theory, design, and applications. *IEEE Trans. Image Process.* 15 (10), 3089–3101. doi:10.1109/TIP.2006.877507

Daniel, E., Anitha, J., Kamaleshwaran, K., and Rani, I. (2017). Optimum spectrum mask based medical image fusion using Gray Wolf Optimization. *Biomed. Signal Process. Control* 34, 36–43. doi:10.1016/j.bspc.2017.01.003

Du, J., Li, W., Xiao, B., and Nawaz, Q. (2016). Union Laplacian pyramid with multiple features for medical image fusion. *Neurocomputing* 194, 326–339. doi:10.1016/j.neucom.2016.02.047

Fu, J., Li, W., Du, J., and Xiao, B. (2020). Multimodal medical image fusion *via* laplacian pyramid and convolutional neural network reconstruction with local gradient energy strategy. *Comput. Biol. Med.* 126, 104048. doi:10.1016/j.compbiomed.2020.104048

Hou, R., Zhou, D., Nie, R., Liu, D., Xiong, L., Guo, Y., et al. (2020). VIF-Net: an unsupervised framework for infrared and visible image fusion. *IEEE Trans. Comput. Imaging* 6, 640–651. doi:10.1109/TCI.2020.2965304

Huang, G., LiuVan Der Maaten, Z. L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in 2017 IEEE conference on computer vision and pattern recognition (CVPR), 2261–2269. doi:10.1109/CVPR.2017.243

Kaur, M., and Singh, D. (2020). Multi-modality medical image fusion technique using multi-objective differential evolution based deep neural networks. *J. Ambient. Intell. Humaniz. Comput.* 12 (2), 2483–2493. doi:10.1007/s12652-020-02386-0

Kennedy, J., and Eberhart, R. C. (1995). "Particle swarm optimization," in IEEE international conference on neural network, 1942–1948.

Lewis, J., O'Callaghan, R., Nikolov, S., Bull, D., and Canagarajah, N. (2007). Pixel- and region-based image fusion with complex wavelets. *Inf. Fusion* 8 (2), 119–130. doi:10.1016/j.inffus.2005.09.006

Li, H., Manjunath, B. S., and Mitra, S. K. (2002). Multi-sensor image fusion using the wavelet transform. *Graph. Models Image Process.* 57 (3), 235–245. doi:10.1006/gmip.1995.1022

Li, H., and Wu, X.-J. (2019). DenseFuse: a fusion approach to infrared and visible images. *IEEE Trans. Image Process.* 28 (5), 2614–2623. doi:10.1109/TIP.2018.2887342

Li, H., Wu, X.-J., and Kittler, J. (2018). "Infrared and visible image fusion using a deep learning framework," in 2018 24th international conference on pattern recognition (ICPR), Beijing, China, August 20–24, 2018. doi:10.1109/ICPR.2018.8546006

Li, J., Huo, H., Liu, K., and Li, C. (2020). Infrared and visible image fusion using dual discriminators generative adversarial networks with Wasserstein distance. *Inf. Sci.* 529, 28–41. doi:10.1016/j.ins.2020.04.035

Li, S., Hao, Q., Kang, X., and Benediktsson, J. (2018). Gaussian pyramid based multiscale feature fusion for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11 (9), 3312–3324. doi:10.1109/JSTARS.2018.2856741

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., et al. (2014). "Microsoft coco: common objects in context," in Computer Vision – ECCV 2014, Zurich, Switzerland, September 6–12, 2014, 740–755. doi:10.1007/978-3-319-10602-1_48

Liu, Y., Chen, X., Peng, H., and Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* 36, 191–207. doi:10.1016/j.inffus.2016.12.001

Liu, Y., Chen, X., Ward, R., and Jane Wang, Z. (2016). Image fusion with convolutional sparse representation. *IEEE Signal Process. Lett.* 23 (12), 1882–1886. doi:10.1109/lsp.2016.2618776

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

Ma, J., Ma, Y., and Li, C. (2019a). Infrared and visible image fusion methods and applications: a survey. *Inf. Fusion* 45, 153–178. doi:10.1016/j.inffus.2018.02.004

Ma, J., Yu, W., Chen, C., Liang, P., Guo, X., and Jiang, J. (2020). Pan-GAN: an unsupervised pan-sharpening method for remote sensing image fusion. *Inf. Fusion* 62, 110–120. doi:10.1016/j.inffus.2020.04.006

Ma, J., Yu, W., Liang, P., Li, C., and Jiang, J. (2019b). FusionGAN: a generative adversarial network for infrared and visible image fusion. *Inf. Fusion* 48, 11–26. doi:10.1016/j.inffus.2018.09.004

Ma, L., Wang, X., Wang, X., Wang, L., Shi, Y., and Huang, M. (2021). TCDA: Truthful combinatorial double auctions for mobile edge computing in industrial internet of things. *IEEE Trans. Mob. Comput.* 99, 1. doi:10.1109/TMC.2021.3064314

Madheswari, K., and Venkateswaran, N. (2016). Swarm intelligence based optimisation in thermal image fusion using dual tree discrete wavelet transform. *Quantitative InfraRed Thermogr. J.* 14 (1), 24–43. doi:10.1080/17686733.2016.1229328

Paramanandham, N., and Rajendiran, K. (2018). Infrared and visible image fusion using discrete cosine transform and swarm intelligence for surveillance applications. *Infrared Phys. Technol.* 88, 13–22. doi:10.1016/j.infrared.2017.11.006

Prabhakar, K. R., Srikar, V. S., and Babu, R. V. (2017). "DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs," in 2017 IEEE international conference on computer vision (ICCV), Venice, Italy, October 22–29, 2017.

Raghavendra, R., Dorizzi, B., Rao, A., and Hemantha Kumar, G. (2011). Particle swarm optimization based fusion of near infrared and visible images for improved face verification. *Pattern Recognit.* 44 (2), 401–411. doi:10.1016/j.patcog.2010.08.006

Shehanaz, S., Daniel, E., Guntur, S., and Satrasupalli, S. (2021). Optimum weighted multimodal medical image fusion using particle swarm optimization. *Optik* 231, 166413. doi:10.1016/j.ijleo.2021.166413

Shen, J., Zhao, Y., Yan, S., and Li, X. (2014). Exposure fusion using boosting laplacian pyramid. *IEEE Trans. Cybern.* 44 (9), 1579–1590. doi:10.1109/TCYB.2013.2290435

Shreyamsha Kumar, B. (2013). Image fusion based on pixel significance using cross bilateral filter. *Signal Image Video Process.* 9 (5), 1193–1204. doi:10.1007/s11760-013-0556-9

Wang, J., Ke, C., Wu, M., Liu, M., and Zeng, C. (2021). Infrared and visible image fusion based on Laplacian pyramid and generative Adversarial Network. *KSII Trans. Internet Inf. Syst.* 15 (5), 1761–1777. doi:10.3837/tiis.2021.05.010

Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* 13 (4), 600–612. doi:10.1109/TIP.2003.819861

Wu, X., Wang, D., Kurths, J., and Kan, H. (2016). A novel lossless color image encryption scheme using 2D DWT and 6D hyperchaotic system. *Inf. Sci.* 349-350, 137–153. doi:10.1016/j.ins.2016.02.041

Zhang, H., Xu, H., Tian, X., Jiang, J., and Ma, J. (2021). Image fusion meets deep learning: a survey and perspective. *Inf. Fusion* 76, 323–336. doi:10.1016/j.inffus.2021.06.008

Zhang, X., Ye, P., Leung, H., Gong, K., and Xiao, G. (2020). Object fusion tracking based on visible and infrared images: a comprehensive review. *Inf. Fusion* 63, 166–187. doi:10.1016/j.inffus.2020.05.002