IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Revue d'Intelligence Artificielle
Vol. 38, No. 2, April, 2024, pp. 575-585
Journal homepage: http://iieta.org/journals/ria

The Deep Learning Methods for Fusion Infrared and Visible Images: A Survey

Rusul Mohammed Neamah*, Tawfiq A. Al-Asadi

Department of Software, Information Technology College, University of Babylon, Babylon 51002, Iraq

Corresponding Author Email: rusulneamah@gmail.com

https://doi.org/10.18280/ria.380221

Received: 15 December 2023
Revised: 2 February 2024
Accepted: 4 March 2024
Available online: 24 April 2024

Keywords:
CNN, auto-encoder, GAN, fused image, infrared and visible image

ABSTRACT

The use of deep learning techniques in the infrared and visible picture fusion domain has dramatically enhanced the effectiveness of image fusion approaches. Deep learning (DL) has significantly boosted the fusion process, improving efficiency and efficacy. This advancement has produced fused images that exhibit a broad spectrum of possible applications. Nevertheless, additional investigation and innovation are required to tackle the difficulties and debates related to the utilization of DL in picture fusion, guaranteeing the ongoing progress of this domain. Various imaging techniques are at one's disposal to capture and present information within the infrared and visible segments of the electromagnetic spectrum. These imaging modalities can encompass a diverse array of intricate details and features. The selection of an imaging approach carries distinct advantages and disadvantages contingent upon the specific application and disparities between infrared and visible depictions in object representation. In addition to their ability to convey finer elements like color, texture, shape, and contrast, visual images also conform to the perceptual traits inherent to human observation. Nonetheless, infrared images may exhibit a different level of intricacy evident in color, texture, shape, and contrast than their visible counterparts. Leveraging advanced deep learning technologies, the amalgamation of visual and infrared photographs synergizes the textural insights of visual images with the thermal data offered by infrared images, thereby affording a composite set of advantages. This article explores deep learning techniques for combining infrared and visible images, focusing on their application in image fusion. It reviews various fusion approaches, including CNNs, GANs, auto-encoders, and transformers, and evaluates fused images using subjective and objective methods. The survey provides a comprehensive overview of current research and suggests future directions in deep learning-based fusion methods.

1. INTRODUCTION

Image fusion improves images by combining those captured through different types of sensors. A reliable and instructive image is created to aid further processing or decision-making. It is essential to extract image information efficiently and apply fusion principles appropriately to formulate a successful fusion approach It is possible to extract useful information from input photos using feature extraction techniques, which can then be smoothly combined with the fused image to prevent artifacts from being introduced into the final image. This process may be carried out in several different ways. Various image pairs can be fused, including visible and SAR images, infrared and SAR images, infrared and visible images, and medical images such as CT and MRI scans [1, 2]. Figure 1 summarizes the general method of fusion images.

In addition to remote sensing and medical imaging, image fusion technology also finds extensive applications in security, surveillance, human visual assistance systems, and the military. Therefore, its study holds significant importance in these application [3].

Due to its wide range of applications, image fusion has gained significant attention in the research community in recent decades. With the advancement of this field, many image transforms and spatial filters have been developed to accommodate both general and specific types of images. A fundamental goal of image fusion is to produce visually appealing results while also achieving high levels of objective results [4]. The ability to describe details and accurately represent hot targets can be greatly enhanced by combining infrared and visible light pictures in a fusion process. It is the infrared sensor's ability to detect infrared radiation that results in the infrared image being created. Usually, the target area of the image is well-lit, so distractions like darkness and smoke can be ignored. Despite this, the infrared sensor cannot detect finer details in a scene due to its insensitivity to brightness variations. Because visible sensor images contain a great deal of texture information and are highly spatially resolved, they are ideal for human eyesight. Visual sensors, however, can be influenced by smoke, darkness, etc., causing them to be hard to see. In light of the aforementioned complementary aspects, combing infrared and visible images is extremely beneficial in retaining the target area [5].

In order to integrate thermal images and visual images, the researchers explored various methods. A deep learning technique and a traditional method are the two types of techniques used in this field. It is common for traditional methods to follow a three-step process when merging pixel-level or predefined features. The first step is to extract features from the original photographs, called feature extraction. Once the features have been extracted, combination schemes are applied to combine them during feature fusion. As a final step, the fused image is recreated by utilizing its matching inverse transformation [6, 7]. The mathematical transformations used in traditional infrared and visible image fusion techniques can be further divided into six groups: multi-scale transform-based (MST) [8, 9], low-rank representation (LRR) [10, 11], sparse representation-based [12, 13], saliency-based [14, 15], subspace-based [16], and hybrid-based methods [17]. The following Figure 2 shows some examples used in the traditional method.

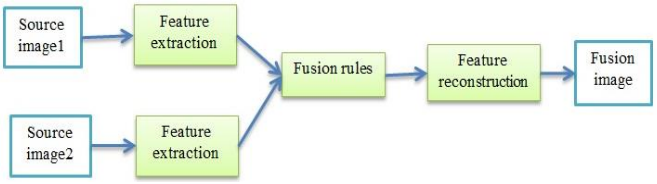
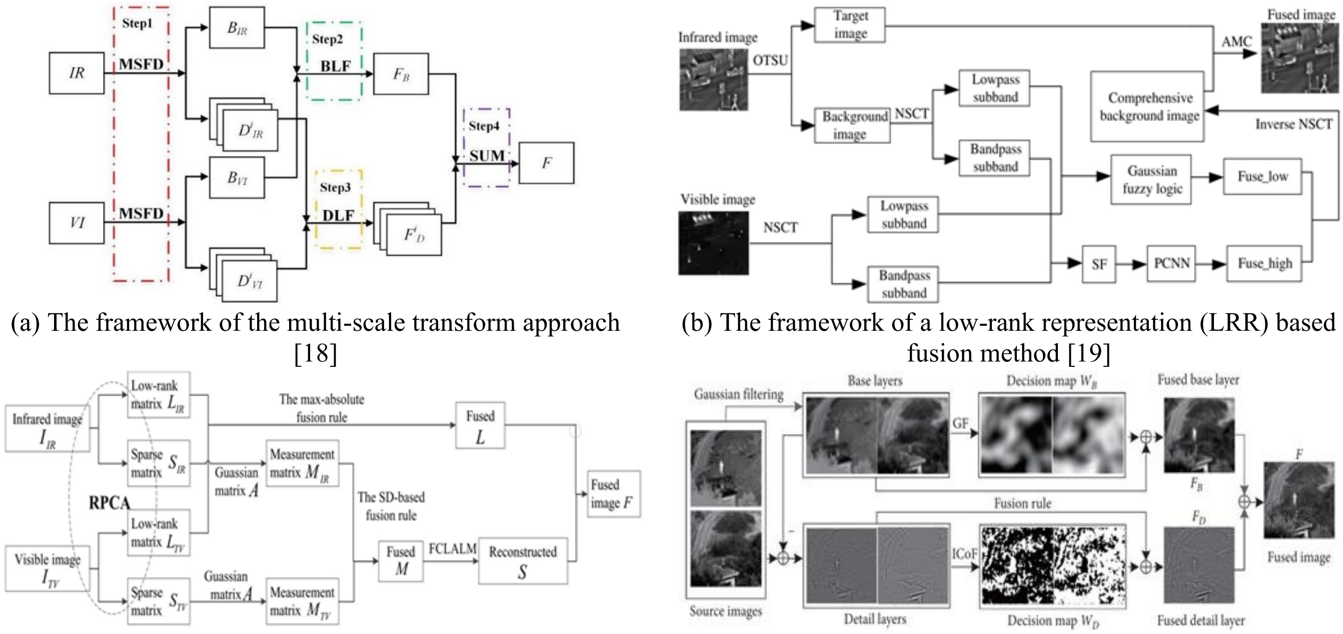


Figure 1. General image fusion method



(a) The framework of the multi-scale transform approach [18]

(b) The framework of a low-rank representation (LRR) based fusion method [19]

(c) the structure of the subspace-based technique [20]

(d)The structure of the hybrid-based technique [21]

Figure 2. Examples of traditional infrared and visible images

By using transform operators, MST methods decompose an image into sub-layers, design strategies of fusion to combine the sub-layers, and use the inverse transformation to produce the final picture. The transform operator like the Laplacian pyramid and produced the weight map which was utilized to combine the relevant layers by taking into account brightness local entropy, contrast, and; therefore, even in low-light situations, excellent results can be achieved [22]. However, the MST method is heavily influenced by the transformation used, and if the fusion rules are incorrect, the results can show artifacts [23].

SR (sparse representations) are an alternative to multiscale transforms (MST). In SR, the goal was to build a very comprehensive dictionary that could be used sparsely to indicate the input images. From the merged sparse representation coefficients, the output (fused) image can be recreated [6].

In order to display visible and infrared pictures, the discrete cosine transform was applied using a fixed over-complete dictionary. It is possible to enhance the visual impact (quality) of combined pictures in target-oriented fusion techniques by using salience methods because salience approaches help to preserve the stability of the key goal area in a fused image [24]. A visual saliency map and guidance Gaussian filter and rolling were used to separate the pictures into different layers and fuse the layers to create a fused image to increase the amount of visual information in the fusion output [25]. A low-rank representation (LRR) decomposes pictures into sparse and low-rank components in a method that is efficient. The low-rank components represent the image's global structure, while the sparse components represent its local details. In order to produce the final fused image, the sub-layers are fused according to appropriate rules. Although traditional image fusion methods have produced satisfactory results, they still have three limitations. The final result is determined by the goodness of the handcrafted features used in the combining process. In addition, conventional approaches, such as sparse representation (SR), can be computationally expensive. In addition, fixed fusion methods need to be tailored to different datasets of images [6]. Conventional image processing methods have limitations due to their human-made design, limited generalization capacity, and computational complexity. These methods depend on input photos and output attributes and may not capture crucial information or handle dynamic scenarios. They also need help with high-resolution or multi-modal images and may be unable to take advantage of the growing amount and diversity of image data and generate a fused image with fewer imperfections, reducing computational expenses.

Deep learning (DL) is a widely used method for picture fusion because of its adaptability, resilience to errors, and ability to reduce noise. Conventional image fusion techniques typically use a sequential procedure consisting of feature extraction, combining schemes, and inverse transformation.

However, this approach can lead to unwanted artifacts in the fused image and can be both intricate and time-consuming to develop. On the other hand, deep learning (DL) techniques can modify the weights of the fusion model using an adaptive mechanism. This enables the model to understand the many characteristics of the photos and generate a fused image with fewer imperfections. Deep learning approaches also exhibit much reduced computational expenses compared to traditional fusion rules, a critical factor in numerous fusion scenarios. In addition, deep learning algorithms can automatically extract features from photos, which makes them highly suitable for tasks such as fusion. They can also handle high-resolution or multi-modal images, resulting in a fused image with fewer flaws. Hence, deep learning algorithms have surpassed traditional methods and are widely employed in image fusion.

2. FUSION METHODS OF INFRARED AND VISIBLE IMAGE FUSION BASED ON DEEP LEARNING

By fusing multiple pictures with various characteristics into a single, high-quality picture, deep learning is a technique that uses deep neural networks. Recently, deep learning techniques for image fusion have grown in popularity because they can automatically extract features from images, which makes them well-suited for jobs like fusion. To overcome the drawbacks of traditional fusion approaches, deep learning techniques are used for feature extraction in several applications, including image classification, image processing and object recognition. In deep learning image fusion, there are several types of neural networks: convolutional neural networks, generative and adversarial networks, auto-encoders and transformers. Despite the high quality of image fusion results produced by deep learning techniques, there are still some areas that need improved. To provide a complete picture of each method, we now discuss its different aspects separately.

2.1 Convolution neural network-based fusion approaches

The fusion of infrared and visible images utilizing a deep learning architecture is a straightforward and efficient technique [26]. By splitting low-frequency data and texture data into two components, the authors are able to extract deep features from meticulous content by utilizing the multilayer fusion strategy of the VGG-19 network [27]. Some loss functions are significantly impacted on CNN's capacity for learning. Method proposed for transferring the style of one image to another utilizing CNN [28]. The process extracts deep features from the produced picture, the style image, and the content image at different layers of the VGG-19 network [27], then minimizes the difference between the created and original images' deep characteristics.

The ResNet50 pre-trained network, as recommended in reference [29], was employed to extract deep features from the source images. This network comprises of 50 weight layers and 5 2D convolutional blocks. The technique of zero-phase component analysis was employed to standardize the deep features and acquire the initial weight map. Ultimately, the soft-max procedure was employed to get the ultimate weights for the source images, and the merged image was reconstructed using the weight-averaging strategy. Using a multi-channel convolutional network, three channels were employed for obtaining features: visible features, infrared features, and features that are common to both infrared and visible images. With the addition and averaging of the featured pictures, the decoding module produces the fused image, and in order to deal with the lack of labeled data, a variety of loss function techniques were utilized. By reworking the loss function, the visible and thermal infrared images were combined adaptively, and noise interference was reduced. The technique is computationally efficient and may preserve important texture details and characteristics without showing any obvious artifacts [30, 31].

Convolutional neural network (CNN) fusion methods are highly efficient at merging infrared and visible images, extracting profound characteristics, maintaining data integrity, and improving contrast and visibility. These technologies are versatile and can be applied to various circumstances and applications, including medical imaging, night vision, and remote sensing. Nevertheless, these models necessitate substantial quantities of training data and computational resources, and they may encounter problems such as overfitting, generalization, or transferability challenges, as well as potentially introducing artifacts or distortions. The enhancements encompass the utilization of sophisticated network structures, integration of pre-existing knowledge, and the creation of resilient assessment criteria. Artifacts can be managed via pre-processing techniques, skip connections, residual blocks, or loss functions. Potential areas for future research involve investigating the integration of several modalities, examining dynamic or temporal images, and implementing fusion techniques in many domains, such as biomedical imaging, security, surveillance, and cultural heritage.

2.2 Generative and adversarial network-based fusion approaches

Deep learning technology is typically used as a foundation for CNN model-based image fusion; however, in this case, the model requires ground truth, so establishing fusion picture standards for combining visible and infrared images is not practical. The ground truth is not taken into consideration when building a deep model that assesses the blurriness in each area of the source image, and then determines the weight. By using a network that generates countermeasures, it is possible to avoid the aforementioned problems by fusing infrared and visible images [32]. Through the use of a target edge-enhancement loss function, target textures were optimized, and the target is now more clearly visible in the fusion output [33]. They also created a detail loss function for more semantic information, as the FusionGAN may lose pixel information from the infrared images. To balance the information between the infrared and visible images, GAN with multi-classification restrictions was suggested [34]. In contrast, these methods emphasize improving visual quality, while ignoring the importance of facilitating the fusion of outcomes following high-level vision challenges by utilizing a prefused picture as the generator's label [35]. As a result, the generator has been trained to produce images that are as similar as possible to the prefused picture. This method ensures that the fused picture retains both the infrared image's thermal information and the rich texture of the viewable picture. Scientists argue that this approach outweighed its disadvantages even though it was computationally expensive due to the pre-fused picture for each training cycle. Figure 3 shows an image fusion technique based on GAN [36].

GAN-based fusion techniques have demonstrated potential in image fusion by producing high-quality fused images that incorporate significant characteristics from both infrared and visible images. Nevertheless, they can incur substantial computing costs and fail to address complex visual tasks. GAN-based fusion algorithms address potential artifacts by creating countermeasures, but their efficacy relies on the quality of the pre-fused image utilized throughout the training process. Future investigations may center on advancing more effective training techniques, novel loss functions, and the customization of the fusion process for diverse datasets.

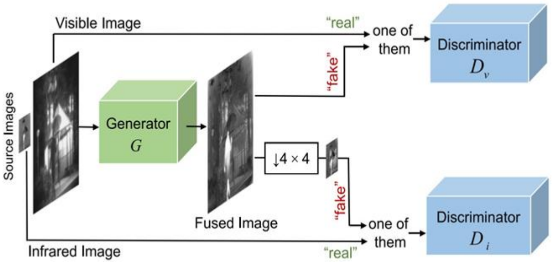


Figure 3. DDcGAN for image fusion

2.3 Auto-encoder-based fusion approaches

Presented an unsupervised auto-encoder network [37], the network elicits the feature from the original pictures using CNN and dense blocks. Using the appropriate fusion technique, the fused feature is then decoded by the decoding module, which incorporates the dense block into the encoding module, preserving as much data as possible. Figure 4 illustrates the auto-encoder's fusion approach. In Nestfuse, the nest connection architecture is utilized as the decoding network, while the encoder network is converted into a multi-scale network [38]. In order to fuse the prominent parts of the picture with the background information, spatial/channel attention fusion techniques are implemented, but multi-modal features cannot be successfully utilized with this handcrafted approach. Utilized the RGB-thermal fusion network (RTFNet) [39], a three-part system: an RGB encoder, an infrared encoder for extracting features from RGB and thermal images, and a decoder for restoring feature picture quality. The accuracy of the estimated feature map may be recovered with a new encoder when using RTFNet for feature extraction if the encoder and decoder are geographically symmetrical. As the method's primary application is scene segmentation, the edges are not crisp.

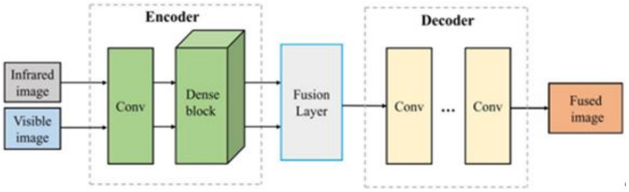


Figure 4. Auto-encoder-based infrared and visible image fusion [39]

Auto-encoder-based fusion methods, such as the unsupervised auto-encoder network, Nestfuse, and RTFNet, have shown promise in image fusion by efficiently extracting features and maintaining data integrity. Nevertheless, they encounter difficulties dealing with multi-modal signals and achieving precise edge recognition. Furthermore, these solutions need to specifically tackle the artifact management issue, highlighting the need for additional research and advancement in this domain. Potential areas for further study in auto-encoder image fusion methods include:

•Enhancing the utilization of multi-modal features and the sharpness of edges;

•Investigating techniques for handling artifacts;

•Optimizing the symmetry of the encoder and decoder to improve the estimate of feature maps.

2.4 Transformer-based fusion approaches

Transformer has experienced significant success in its initial application to natural language processing [40], and while CNN focuses on local aspects, its attention mechanism can assist in developing long-range reliance, allowing it to utilize global data in both deep and shallow layers better. According to the vision transformer concept [41], the vision transformer has a lot of potential for computer vision (CV). Recently, CV researchers have been using more transforms, such as object identification, multiple object tracking, segmentation, and others, to do so. Transformers are based mainly on attention mechanisms [42]. An integrated model based on the transformer was suggested, and it performed well on several low-level visual tasks [43]. The global spatial dependency of transformers has been applied to several areas of computer vision. We focus on the overall correlation of picture space and channels throughout the fusion process, motivated by the properties of the trans-former, as proposed TGFuse, which involves using a lightweight transformer module and adversarial learning for visible and infrared image fusion [44]. Through the use of the transformer technique to build efficient global fusion interactions, shallow features extracted by a CNN in the transformer fusion module can interact with each other. This interaction simultaneously improves the fusion connection across channels and within the spatial range. By enforcing competitive consistency from the inputs during the training process, adversarial learning can enhance outcome discrimination. An improved fusion model for focal Transformers, based on the multi-modal feature self-adaptive fusion technique, is proposed to provide a fused image that is both visually appealing and more informative by fusing infrared and visible information [6]. A spatio-transformer (ST) fusion method was used to fuse images obtained from different sensors in the proposed technique [45]. There are three parts to the image fusion transformer: an encoder network, an ST fusion network, and a nested decoder network. The ST fusion network, which consists of spatial and transformer branches, then fuses features at multiple scales.

Transformer-based fusion methods have demonstrated potential in the field of picture fusion. They efficiently leverage worldwide data and enhance integration linkages between channels within the spatial scope. Nevertheless, individuals could encounter difficulties when dealing with the intricacy of the transformer strategy and the computational expense of adversarial learning.

These methods do not directly deal with the management of artifacts. Further research and development in this area would be advantageous.

Potential enhancements can be achieved in the efficacy of worldwide fusion interactions and adversarial learning. Possible future research directions include investigating artifact handling approaches and optimizing the encoder's and decoder's symmetry to enhance feature map estimation.

Furthermore, prospective studies might focus on applying these methods to datasets with greater diversity and including more sophisticated attention mechanisms, which would be highly beneficial.

3. ASSESSMENT OF FUSED IMAGE

The optimal algorithm, approach, or measure for improved picture evaluation is often chosen by comparing different image processing approaches. For a variety of image-enhancing tasks, including the fine-tuning of image resolutions for alignment, the overlaying of two picture products, and the mixing of images for feature extraction and target recognition, image fusion is a common option. Since image fusion is used in many geospatial and night vision applications as well as objectively evaluating image fusion algorithms [46], it is crucial to understand these methods. Different point-specific assessment indicators can be used by researchers to make quantitative references and precise image fusion comparisons. Subjective evaluation and objective evaluation can be used to categorize the available integration indicators [47].

3.1 Subjective evaluation approaches

The subjective assessment is evaluated in absolute and relative terms using well-known five-level quality scales and obstacle scales, respectively [48]. An effective subjective assessment method involves visually inspecting the picture without any aids and carefully analyzing its characteristics, distortion, contrast, and image integrity to evaluate different fusion processes. Subjective assessors can use the assessment criteria to assign a quality grade to the merged picture. However, various people have different standards for evaluating the same image, and these standards can be easily influenced by context, environment, and other variables, resulting in inaccurate answers to the merged image. Given its poor goodness and delayed timeliness, it is not easy to assess fusion images using this approach in several dimensions. To accurately assess fusion outcomes, objective assessment markers must be combined [49].

The subjective evaluations of fused pictures are constrained by the divergent criteria employed by different assessors, which might be swayed by factors such as context, surroundings, and personal biases. Consequently, this can result in consistent and correct assessments. For instance, the interpretation of an image can vary among individuals due to their subjective perceptions, environmental influences, and personal biases, leading to inconsistent and incorrect assessments.

3.2 Objective evaluation approaches

Objective evaluation measures are created and utilized to overcome the constraints of subjective evaluations. These metrics use accurate formulas to produce relevant index data of the fused image. Image fusion benefits from their inclusion by offering a more standardized and consistent evaluation method. The fundamental principles of these metrics, including those derived from information theory, structural similarity, feature similarity, and source and output images, strive to offer a more quantitative and impartial evaluation of the quality and effectiveness of fused images. Using reference and non-reference standards, these metrics provide a more systematic and dependable approach to assessing the efficacy of picture fusion techniques [49, 50].

3.2.1 Metrics Based on Information Theory

(1) Entropy (EN)

$$H = -\sum_{i=0}^{L-1} P_i \log_2 P_i \qquad (1)$$

where, L is the gray level of the image from 0 to 255, P_i is the probability of the gray level I in the image. EN might indicate the texture richness and average info in the merged picture. The quantity of info in the fused picture is more plentiful the greater the EN is. And one of the most often used indicators for evaluating image quality is EN. If the fused picture had noise and artifacts, however, the value of EN would significantly rise and cannot accurately reflect the goodness of the final picture. Particularly, IR images will have a lot of noise. Therefore, we believe EN should only be employed as a secondary assessment metric on IR and VI image fusion [48, 51, 52].

(2) Mutual Information (MI)

MI is used to calculate the amount of info that was transmitted from the source image to the fusion image. According to information theory, MI denotes the statistical interdependence of two random variables [25, 53] and has the following mathematical definition:

$$MI_F^{AB} = MI_{FA} + MI_{FB} \qquad (2)$$

A high MI metric signifies that a lot of info is transmitted from the input pictures to the fused image, which signals good fusion performance, whereas MIFA and MIFB indicate the amount of info that went into creating the fusion pictures from the infrared and visible photographs, respectively.

(3) Peak Signal-To-Noise Ratio (PSNR)

SNR is used to compute the peak power to noise value of ower [54, 55]. These are the criteria for this metric:

$$PSNR = 10\log_{10}\left\{\frac{r^2}{MSE}\right\} \qquad (3)$$

In Eq. (3), r denotes the fused image's peak value. A high PSNR value indicates that the fusion procedure is less damaged and that the fused picture is identical to the input image.

3.2.2 Metrics based on structural similarity

(1) Structural Similarity Index Measure (SSIM)

Mathematically, SSIM between two components U and V is expressed as

$$SSM(U,V) = \frac{\sigma_{uv}}{\sigma_u \sigma_v} \frac{2\mu_u \mu_v}{\mu_u^2 + \mu_v^2} \frac{2\sigma_u \sigma_v}{\sigma_u^2 + \sigma_v^2} \qquad (4)$$

where, σ_U, σ_V,σ_{UV} are the variances and covariance and μ_U, μ_V are mean intensities. The structure, contrast, and luminance distortion between the fused picture and the original pictures are combined in the design of SSIM by modeling any image distortion as a contrast distortion, mix of loss correlation, and radiometric [56, 57].

(2) Mean Squared Error (MSE)

The fault and the actual distinction between the perfect or estimated outcomes are computed using MSE [58, 59]. According to its definition:

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}\left(A_{ij} - B_{ij}\right)^2 \tag{5}$$

where, m and n are the height and width of the picture, indicating the pixel rows and columns, A and B are the ideal and evaluate able compound pictures, respectively, and i and j are the pixel row and column indexes.

(3) Correlation Coefficient (CC)

CC has the following mathematical definition and assesses the degree of linear correlation between a fused picture and visible and infrared pictures [60, 61]:

$$CC = \frac{\left(r_{IF} + r_{VF}\right)}{2} \tag{6}$$

$$r_{XF} = \frac{\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X(i,j) - X\right)\left(F(i,j) - F\right)}{\sqrt{\sum_{i=1}^{H}\sum_{j=1}^{W}\left(X(i,j) - X\right)^2\left(\sum_{i=1}^{H}\sum_{j=1}^{W}\left(F(i,j) - F\right)^2\right)}} \tag{7}$$

where, X denotes the original image. X and F represent fused images, and H and W stand for the length and width of the original picture.

3.2.3 Metrics based on feature similarity

(1) Average Gradient (AG)

The fused image's gradient information is quantified by the average gradient (AG) metric, which also exemplifies its detail and texture [49, 62, 63], following defines:

$$AG = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{\frac{VF_i^2(i,j) + VF_y^2(i,j)}{2}} \tag{8}$$

where, $F_i = F_{k,l} - F_{k+1,l}$, $F_j = F_{k,l} - F_{k,l+1}$, M and N represent the dimension of fused picture F at pixel level. The greater the average gradient value, the greater the data in the picture, resulting in a superior fused outcome.

(2) Standard Deviation (SD)

The idea is the distribution and contrast of the merged picture serve as the foundation for the standard deviation (SD) measure [2, 64]. SD is described mathematically as follows:

$$SD = \sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(F(i,j) - \mu\right)^2} \tag{9}$$

where, μ stands for the fused image's mean value. Our eyes are naturally drawn to areas with strong contrast as we are highly sensitive to visual differences.

(3) Spatial Frequency (SF)

The concept can be split into two parts: spatial column frequency (CF) and spatial row frequency (RF). The formulas for both are displayed below. Spatial frequency, which indicates the total activity of a picture in the spatial domain.

$$SF = \sqrt{RF^2 + CF^2} \tag{10}$$

$$CF = \sqrt{\frac{1}{M \times N}\sum_{i=1}^{N}\sum_{j=1}^{M}\left[F(i,j) - F(i,j-1)\right]^2} \tag{11}$$

$$RF = \sqrt{\frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N}\left[F(i,j) - F(i,j-1)\right]^2} \tag{12}$$

SF stands for both the image's spatial change and the precision of the details. The textures and edges get richer as the SF gets bigger. Additionally, it operates apart from the reference image. The value of SF will increase due to the undesired artefacts in the combined IR and VI pictures. The quality of the merged image cannot be accurately reflected by the SF in this situation [14, 48, 65].

(4) Gradient-Based Fusion Performance ($Q^{AB/F}$)

Based on the presumption that the edge information in the original pictures is preserved in the fused picture, $Q^{AB/F}$ assesses the quantity of edge info that is transmitted from the original photos to the fused picture [66, 67]. Following is a definition of $Q^{AB/F}$:

$$Q^{AB/F} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{M} Q^{AF}(i,j)\,w^A(i,j) + Q^{BF}(i,j)\,w^B(i,j)}{\sum_{i=1}^{N}\sum_{j=1}^{M}\left(w^A(i,j) + w^B(i,j)\right)} \tag{13}$$

3.2.4 Metrics based on source and produced images

(1) Visual Information Fidelity (VIF)

VIF was created based on visual information fidelity (VIF) and is solely utilized in image fusion [68]. The visual data from the original image was extracted using the VIF model by Han et al. [69]. After additional processing to eliminate the distortion of information, they were able to successfully fuse the visual data. The VIF, which is specifically utilized for fusion assessment, is generated after incorporating all of the visual info. summarizes the VIF calculating procedure into four stages. It is necessary to first filter and partition the fusion image into numerous blocks from the source image. Check to see whether any of the blocks have distorted visual data next. Third, check the accuracy of the visual data in each block. In the fourth stage, the overall index based on VIF is determined [69].

(2) Other measures

The metrics Q_{CB} and Q_{CV}, which gauge how well-fused pictures work visually, are based on what humans see. An important metric for assessing an algorithm's performance is running time. The computational effectiveness of the model is assessed using the time-consuming nature of an image fusion technique [47, 49, 70].

4. EXPERIMENT

The number of studies and methods in the field of image merging is growing every day. The primary goal is to explore the present issues and potential directions for image fusion as they relate to diverse fields, including surveillance, photography, medical diagnosis, and remote sensing. The following data sets were used in the tests for the visual and infrared image fusion in this field: TNO dataset [71] is a collection of multispectral nighttime images captured by several multiband camera systems in various military-relevant settings. The FLIR dataset offers comparable RGB pictures and annotated thermography datasets. 14,452 infrared pictures altogether are included in the collection. The majority of the 15,488 pairs of photos in the LLVIP collection [72] were captured in extremely dark environments, and each pair is perfectly matched in time and location. The KAIST [73] data collection contains different broad sceneries of a campus, a street, and a rural area. Each image has an associated visual picture and thermal picture. With a spatial resolution of 480 × 640, the infrared and visible picture pairings in the MSRS [74] collection include both daylight and nighttime settings. Table 1 shows performance of some deep learning-based image fusion techniques. Experiments were conducted on 10 pair of images collected from KAIST data set.

Table 2 shows the results of some fusion methods based on deep learning conducted on 21 pair of images collected from TNO dataset.

Table 1. The performance of some methods using 10 pair pictures from KAIST data set and the best first two values are indicated in bold and red Italic font

Technique	SF	EN	$Q^{AB/F}$	SSIM	MI	SD	VIF
DenseFuse [37]	9.3238	6.8526	0.4735	0.8692	13.7053	81.7283	0.6875
NestFuse [38]	9.7807	6.8745	*0.5011*	0.8817	13.7491	*83.0530*	0.7195
TGFuse [44]	11.3149	**6.9838**	**0.5863**	**0.9160**	**13.9676**	**94.7203**	**0.7746**
DeepFuse [75]	8.3500	6.6102	0.3847	0.9138	13.2205	66.8872	0.5752
IFCNN [76]	**11.8590**	6.6454	0.4962	0.9129	13.2909	73.7053	0.6090
U2Fusion [77]	11.0368	6.7227	0.3934	*0.9147*	13.4453	66.5035	*0.7680*
RFN-Nest [78]	5.8457	6.7274	0.3292	0.8959	13.4547	67.8765	0.5404
FusionGAN [79]	8.0476	6.5409	0.2682	0.6135	13.0817	61.6339	0.4928

Table 2. The performance of some techniques based on deep learning using 21 pair of pictures from TNO dataset and best first two values are specified in bold and red Italic font

Technique	EN	SD	AG	SF	SSIM	VIF	MI
DDcGAN [36]	**7.5306**	*50.5463*	*6.3313*	11.6881	0.5098	0.6387	**15.0611**
DenseFuse [37]	6.9307	35.3016	2.9871	5.9371	0.6861	0.5013	13.8614
DeepFuse [75]	6.8825	34.1770	4.0241	8.0985	*0.7122*	0.5529	13.6546
IFCNN [76]	6.7259	32.0164	5.9337	11.5053	**0.7193**	0.3754	13.4489
U2Fusion [77]	4.0501	37.3202	6.3172	*11.7124*	0.6371	*0.6672*	14.0979
fusionGAN [79]	6.6094	30.5280	3.1951	6.2315	0.6826	0.2614	13.2159
MTNO [80]	*7.3101*	**51.6398**	**10.9079**	**21.2000**	0.5598	**1.0053**	*14.6201*

According to objective experimental findings, each fusion technique has benefits and downsides, and diverse techniques exhibit distinct benefits in various contexts. The integration of infrared and visible pictures utilizing deep learning techniques has led to the continual advance of enhanced new technologies. Evaluation criteria shown in Table 1 quantify the excellence and efficiency of several image fusion techniques that rely on deep learning. The metrics encompass measurements related to the differentiation in brightness, the level of detail, the distinctness, and the accuracy of the merged pictures. According to the values, the TGFuse approach demonstrates the top scores in most metrics, with NestFuse, and U2Fusion closely following. These findings indicate that TGFuse is the most efficient and resilient technique for image fusion, mainly when applied to the KAIST dataset. According to the information presented in Table 2, the MTNO technique exhibits highest scores in most criteria, with DDcGAN, U2Fusion, IFCNN, and DeepFuse closely trailing behind. The findings suggest that MTNO is the optimal and robust method for image fusion, mainly when used with the TNO dataset. FusionGAN and DenseFuse exhibit inferior performance across all criteria, indicating their diminished efficacy in picture fusion compared to the other approaches. Nevertheless, diverse datasets and settings may necessitate distinct evaluation metrics and criteria contingent upon the aim and application of picture fusion.

The objective experimental findings indicate that the TGFuse technique achieves the highest scores in most criteria, especially when applied to the KAIST dataset. Similarly, the MTNO approach demonstrates superior performance in most criteria, particularly when used in the TNO dataset. Hence, the authors should concentrate on extensively investigating and enhancing the TGFuse and MTNO approaches for image fusion. These methods have demonstrated superior efficiency and durability in their specific contexts, and their ongoing progress through deep learning techniques has resulted in the creation of improved new technologies. By conducting a more thorough investigation of the methods and algorithms utilized in TGFuse and MTNO, the authors have the potential to reveal valuable insights that can further advance the field of image fusion. Furthermore, it is advantageous for the authors to contemplate the possible versatility of these strategies about other datasets and environments, as suggested by the results, to guarantee their strength and effectiveness in a wide range of picture fusion applications.

While the image fusion technique is making some progress, there are still several issues for which there is no ideal answer. In the future, it will be necessary to enhance and explore the issues with picture fusion. Although several convolutional neural network-based image fusion models perform well, most of them fall short of perfection. Finally, to completely maintain the feature information acquired from each layer of convolution, the fusion approach utilizing the convolution neural network must give focus on improving the fluidity of the intermediate layer network's features.

5. CONCLUSIONS

The "Deep Learning Methods for Fusion of Infrared and Visible Images: A Survey" concludes the progress made in image fusion, explicitly focusing on deep learning techniques. The survey systematically assesses the fusion methods by employing various picture metrics, distinguishing their respective contributions, advantages, and constraints. Furthermore, it clearly defines the current state of research on the fusion of infrared and visible images and provides a framework for possible future study directions.

The survey highlights the notable progress of employing deep learning approaches in infrared and visible image fusion. It emphasizes the enhanced efficacy of image fusion methods, creating fused images with various possible uses. Nevertheless, the survey recognizes the current difficulties and discussions surrounding the application of deep learning in picture fusion. It highlights the necessity for further investigation and creativity to tackle these intricacies and guarantee the continuous advancement of this domain.

From an objective and subjective standpoint, the survey's comprehensive assessment of the fusion algorithms offers significant insights into the effectiveness of different fusion procedures. This analysis discerns the advantages and constraints of various methodologies, providing a comprehensive comprehension of their efficacy in varied circumstances. The survey's emphasis on objective and subjective judgments highlights the need to use a complete evaluation methodology to appropriately measure the quality and efficiency of image fusion techniques.

The "Deep Learning Methods for Fusion of Infrared and Visible Images: A Survey" is a helpful resource for scholars and practitioners in image fusion. This presentation demonstrates the progress made possible by deep learning and highlights the unresolved obstacles and the prospective directions for future research. The survey enhances the improvement and innovation in the field of image fusion by offering a comprehensive and unbiased evaluation of the present state of the art.

ACKNOWLEDGMENT

We would like to express our special thanks to all the individuals who contributed to the completion of this research. I thank the supervisors who provided guidance and academic scrubs.

REFERENCES

[1] Wang, J., Peng, J., Feng, X., He, G., Fan, J. (2014). Fusion method for infrared and visible images by using non-negative sparse representation. Infrared Physics & Technology, 67: 477-489. https://doi.org/10.1016/j.infrared.2014.09.019

[2] Piella, G. (2003). A general framework for multiresolution image fusion: From pixels to regions. Information Fusion, 4(4): 259-280. https://doi.org/10.1016/S1566-2535(03)00046-0

[3] Chen, Y., Xiong, J., Liu, H.L., Fan, Q. (2014). Fusion method of infrared and visible images based on neighborhood characteristic and regionalization in NSCT domain. Optik, 125(17): 4980-4984. https://doi.org/10.1016/j.ijleo.2014.04.006

[4] Sharma, A.M., Dogra, A., Goyal, B., Vig, R., Agrawal, S. (2020). From pyramids to state-of-the-art: A study and comprehensive comparison of visible-infrared image fusion techniques. IET Image Processing, 14(9): 1671-1689. https://doi.org/10.1049/iet-ipr.2019.0322

[5] Zhang, T.Y., Zhou, Q., Feng, H.J., Xu, Z.H., Li, Q., Chen, Y.T. (2013). Fusion of infrared and visible light images based on nonsubsampled shearlet transform. In International Symposium on Photoelectronic Detection and Imaging 2013: Infrared Imaging and Applications, 8907: 366-373. https://doi.org/10.1117/12.2032470

[6] Liu, X., Gao, H., Miao, Q., Xi, Y., Ai, Y., Gao, D. (2022). MFST: Multi-modal feature self-adaptive transformer for infrared and visible image fusion. Remote Sensing, 14(13): 3233. https://doi.org/10.3390/rs14133233

[7] Tang, L., Xiang, X., Zhang, H., Gong, M., Ma, J. (2023). DIVFusion: Darkness-free infrared and visible image fusion. Information Fusion, 91: 477-493. https://doi.org/10.1016/j.inffus.2022.10.034

[8] Chen, J., Li, X., Luo, L., Mei, X., Ma, J. (2020). Infrared and visible image fusion based on target-enhanced multiscale transform decomposition. Information Sciences, 508: 64-78. https://doi.org/10.1016/j.ins.2019.08.066

[9] Chen, J., Li, X., Luo, L., Ma, J. (2021). Multi-focus image fusion based on multi-scale gradients and image matting. IEEE Transactions on Multimedia, 24: 655-667. https://doi.org/10.1109/TMM.2021.3057493

[10] Gao, C., Song, C., Zhang, Y., Qi, D., Yu, Y. (2021). Improving the performance of infrared and visible image fusion based on latent low-rank representation nested with rolling guided image filtering. IEEE Access, 9: 91462-91475. https://doi.org/10.1109/ACCESS.2021.3090436

[11] Li, H., Wu, X.J. (2018). Infrared and visible image fusion using latent low-rank representation. arXiv preprint arXiv:1804.08992. https://doi.org/10.48550/arXiv.1804.08992

[12] Wei, Q., Bioucas-Dias, J., Dobigeon, N., Tourneret, J. Y. (2015). Hyperspectral and multispectral image fusion based on a sparse representation. IEEE Transactions on Geoscience and Remote Sensing, 53(7): 3658-3668. https://doi.org/10.1109/TGRS.2014.2381272

[13] Zhang, Q., Liu, Y., Blum, R. S., Han, J., Tao, D. (2018). Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. Information Fusion, 40: 57-75. https://doi.org/10.1016/j.inffus.2017.05.006

[14] Zhang, X., Ma, Y., Fan, F., Zhang, Y., Huang, J. (2017). Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition. JOSA A, 34(8): 1400-1410. https://doi.org/10.1364/josaa.34.001400

[15] Liu, C.H., Qi, Y., Ding, W.R. (2017). Infrared and visible image fusion method based on saliency detection in sparse domain. Infrared Physics & Technology, 83: 94-102. https://doi.org/10.1016/j.infrared.2017.04.018

[16] Zhang, J., Wei, L., Miao, Q., Wang, Y. (2004). Image fusion based on nonnegative matrix factorization. In 2004 International Conference on Image Processing,

2004. ICIP '04., Singapore, pp. 973-976. https://doi.org/10.1109/ICIP.2004.1419463

[17] Liu, Y., Wu, Z., Han, X., Sun, Q., Zhao, J., Liu, J. (2022). Infrared and visible image fusion based on visual saliency map and image contrast enhancement. Sensors, 22(17): 6390. https://doi.org/10.3390/s22176390

[18] Yan, H., Li, Z. (2020). Infrared and visual image fusion based on multi-scale feature decomposition. Optik, 203: 163900. https://doi.org/10.1016/j.ijleo.2019.163900

[19] He, K., Zhou, D., Zhang, X., Nie, R., Wang, Q., Jin, X. (2017). Infrared and visible image fusion based on target extraction in the nonsubsampled contourlet transform domain. Journal of Applied Remote Sensing, 11(1): 015011. https://doi.org/10.1117/1.jrs.11.015011

[20] Li, J., Song, M., Peng, Y. (2018). Infrared and visible image fusion based on robust principal component analysis and compressed sensing. Infrared Physics & Technology, 89: 129-139. https://doi.org/10.1016/j.infrared.2018.01.003

[21] Zhang, Y., Li, D., Zhu, W. (2020). Infrared and visible image fusion with hybrid image filtering. Mathematical Problems in Engineering, 2020: 1757214. https://doi.org/10.1155/2020/1757214

[22] Vanmali, A.V., Gadre, V.M. (2017). Visible and NIR image fusion using weight-map-guided Laplacian–Gaussian pyramid for improving scene visibility. Sādhanā, 42: 1063-1082. https://doi.org/10.1007/s12046-017-0673-1

[23] Zhou, X., Wang, W. (2016). Infrared and visible image fusion based on tetrolet transform. In Proceedings of the 2015 International Conference on Communications, Signal Processing, and Systems, pp. 701-708. https://doi.org/10.1007/978-3-662-49831-6_72

[24] Bin, Y., Chao, Y., Guoyu, H. (2016). Efficient image fusion with approximate sparse representation. International Journal of Wavelets, Multiresolution and Information Processing, 14(4): 1650024. https://doi.org/10.1142/S0219691316500247

[25] Ma, J., Zhou, Z., Wang, B., Zong, H. (2017). Infrared and visible image fusion based on visual saliency map and weighted least square optimization. Infrared Physics & Technology, 82: 8-17. https://doi.org/10.1016/j.infrared.2017.02.005

[26] Li, H., Wu, X.J., Kittler, J. (2018). Infrared and visible image fusion using a deep learning framework. In 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, pp. 2705-2710. https://doi.org/10.1109/ICPR.2018.8546006

[27] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556

[28] Handa, A., Garg, P., Khare, V. (2018). Masked neural style transfer using convolutional neural networks. In 2018 International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), Bhubaneswar, India, pp. 2099-2104. https://doi.org/10.1109/ICRIEECE44171.2018.9008937

[29] Li, H., Wu, X.J., Durrani, T.S. (2019). Infrared and visible image fusion with ResNet and zero-phase component analysis. Infrared Physics & Technology, 102: 103039. https://doi.org/10.1016/j.infrared.2019.103039

[30] Wang, H., An, W., Li, L., Li, C., Zhou, D. (2022). Infrared and visible image fusion based on multi-channel convolutional neural network. IET Image Processing, 16(6): 1575-1584. https://doi.org/10.1049/ipr2.12431

[31] Hou, R., Zhou, D., Nie, R., Liu, D., Xiong, L., Guo, Y., Yu, C. (2020). VIF-Net: An unsupervised framework for infrared and visible image fusion. IEEE Transactions on Computational Imaging, 6: 640-651. https://doi.org/10.1109/tci.2020.2965304

[32] Liu, Y., Chen, X., Cheng, J., Peng, H., Wang, Z. (2018). Infrared and visible image fusion with convolutional neural networks. International Journal of Wavelets, Multiresolution and Information Processing, 16(3): 1850018. https://doi.org/10.1142/S0219691318500182

[33] Ma, J., Liang, P., Yu, W., Chen, C., Guo, X., Wu, J., Jiang, J. (2020). Infrared and visible image fusion via detail preserving adversarial learning. Information Fusion, 54: 85-98. https://doi.org/10.1016/j.inffus.2019.07.005

[34] Ma, J., Zhang, H., Shao, Z., Liang, P., Xu, H. (2020). GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion. IEEE Transactions on Instrumentation and Measurement, 70: 1-14. https://doi.org/10.1109/TIM.2020.3038013

[35] Li, Q., Lu, L., Li, Z., Wu, W., Liu, Z., Jeon, G., Yang, X. (2019). Coupled GAN with relativistic discriminators for infrared and visible images fusion. IEEE Sensors Journal, 21(6): 7458-7467. https://doi.org/10.1109/JSEN.2019.2921803

[36] Ma, J., Xu, H., Jiang, J., Mei, X., Zhang, X. P. (2020). DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Transactions on Image Processing, 29: 4980-4995. https://doi.org/10.1109/TIP.2020.2977573

[37] Li, H., Wu, X.J. (2018). DenseFuse: A fusion approach to infrared and visible images. IEEE Transactions on Image Processing, 28(5): 2614-2623. https://doi.org/10.1109/TIP.2018.2887342

[38] Li, H., Wu, X.J., Durrani, T. (2020). NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. IEEE Transactions on Instrumentation and Measurement, 69(12): 9645-9656. https://doi.org/10.1109/TIM.2020.3005230

[39] Sun, Y., Zuo, W., Liu, M. (2019). Rtfnet: Rgb-thermal fusion network for semantic segmentation of urban scenes. IEEE Robotics and Automation Letters, 4(3): 2576-2583. https://doi.org/10.1109/LRA.2019.2904733

[40] Veltman, A., Pulle, D.W., De Doncker, R.W., Veltman, A., Pulle, D. W., De Doncker, R.W. (2016). The transformer. Fundamentals of Electrical Drives, pp. 47-82. https://doi.org/10.1007/978-3-319-29409-4_3

[41] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X. (2021). Thomas unterthiner mostafa dehghani matthias minderer georg heigold sylvain gelly jakob uszkoreit and neil houlsby. An image isworth 16× 16 words: Transformers for image recognition atscale. In International Conference on Learning Representations.

[42] Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. In 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

[43] Chen, H., Wang, Y., Guo, T., et al. (2021). Pre-trained image processing transformer. In 2021 IEEE/CVF

Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, pp. 12294-12305. https://doi.org/10.1109/CVPR46437.2021.01212

[44] Rao, D., Xu, T., Wu, X.J. (2023). TGFuse: An infrared and visible image fusion approach based on transformer and generative adversarial network. IEEE Transactions on Image Processing. https://doi.org/10.1109/TIP.2023.3273451

[45] Vs, V., Valanarasu, J.M.J., Oza, P., Patel, V.M. (2022). Image fusion transformer. In 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, pp. 3566-3570. IEEE. https://doi.org/10.1109/ICIP46576.2022.9897280

[46] Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R., Wu, W. (2011). Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(1): 94-109. https://doi.org/10.1109/TPAMI.2011.109

[47] Chen, Y., Blum, R.S. (2009). A new automated quality assessment algorithm for image fusion. Image and vision computing, 27(10): 1421-1432. https://doi.org/10.1016/j.imavis.2007.12.002

[48] Jin, X., Jiang, Q., Yao, S., Zhou, D., Nie, R., Hai, J., He, K. (2017). A survey of infrared and visual image fusion methods. Infrared Physics & Technology, 85: 478-501. https://doi.org/10.1016/j.infrared.2017.07.010

[49] Sun, C., Zhang, C., Xiong, N. (2020). Infrared and visible image fusion techniques based on deep learning: A review. Electronics, 9(12): 2162. https://doi.org/10.3390/electronics9122162

[50] Ma, W., Wang, K., Li, J., Yang, S. X., Li, J., Song, L., Li, Q. (2023). Infrared and visible image fusion technology and application: A review. Sensors, 23(2): 599. https://doi.org/10.3390/s23020599

[51] Cvejic, N., Canagarajah, C.N., Bull, D.R. (2006). Image fusion metric based on mutual information and Tsallis entropy. Electronics Letters, 42(11): 1.

[52] Chen, Y., Shin, H. (2020). Multispectral image fusion based pedestrian detection using a multilayer fused deconvolutional single-shot detector. JOSA A, 37(5): 768-779. https://doi.org/10.1364/JOSAA.386410

[53] Jin, H., Jiao, L., Liu, F., Qi, Y. (2008). Fusion of infrared and visual images based on contrast pyramid directional filter banks using clonal selection optimizing. Optical Engineering, 47(2): 027002. https://doi.org/10.1117/1.2857417

[54] Ma, J., Ma, Y., Li, C. (2019). Infrared and visible image fusion methods and applications: A survey. Information Fusion, 45: 153-178. https://doi.org/10.1016/j.inffus.2018.02.004

[55] Gao, S., Cheng, Y., Zhao, Y. (2013). Method of visual and infrared fusion for moving object detection. Optics letters, 38(11): 1981-1983. https://doi.org/10.1364/OL.38.001981

[56] Li, S., Yin, H., Fang, L. (2012). Group-sparse representation with dictionary learning for medical image denoising and fusion. IEEE Transactions on Biomedical Engineering, 59(12): 3450-3459. https://doi.org/10.1109/TBME.2012.2217493

[57] Han, L., Shi, L., Yang, Y., Song, D. (2014). Thermal physical property-based fusion of geostationary meteorological satellite visible and infrared channel images. Sensors, 14(6): 10187-10202. https://doi.org/10.3390/s140610187

[58] Jiang, Y., Wang, M. (2014). Image fusion using multiscale edge-preserving decomposition based on weighted least squares filter. IET image Processing, 8(3): 183-190. https://doi.org/10.1049/iet-ipr.2013.0429

[59] Zhang, X. (2021). Deep learning-based multi-focus image fusion: A survey and a comparative study. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(9): 4819-4838. https://doi.org/10.1109/TPAMI.2021.3078906

[60] Xiang, T., Yan, L., Gao, R. (2015). A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain. Infrared Physics & Technology, 69: 53-61. https://doi.org/10.1016/j.infrared.2015.01.002

[61] Zhan, L., Zhuang, Y. (2016). Infrared and visible image fusion method based on three stages of discrete wavelet transform. International Journal of Hybrid Information Technology, 9: 407-418. http://dx.doi.org/10.14257/ijhit.2016.9.5.35

[62] Madheswari, K., Venkateswaran, N. (2017). Swarm intelligence based optimisation in thermal image fusion using dual tree discrete wavelet transform. Quantitative Infrared Thermography Journal, 14(1): 24-43. https://doi.org/10.1080/17686733.2016.1229328

[63] Liu, Z., Yin, H., Fang, B., Chai, Y. (2015). A novel fusion scheme for visible and infrared images based on compressive sensing. Optics Communications, 335: 168-177. https://doi.org/10.1016/j.optcom.2014.07.093

[64] Zhao, C., Guo, Y., Wang, Y. (2015). A fast fusion scheme for infrared and visible light images in NSCT domain. Infrared Physics & Technology, 72: 266-275. https://doi.org/10.1016/j.infrared.2015.07.026.

[65] Zhao, J., Cui, G., Gong, X., Zang, Y., Tao, S., Wang, D. (2017). Fusion of visible and infrared images using global entropy and gradient constrained regularization. Infrared Physics & Technology, 81: 201-209. https://doi.org/10.1016/j.infrared.2017.01.012

[66] Ma, J., Chen, C., Li, C., Huang, J. (2016). Infrared and visible image fusion via gradient transfer and total variation minimization. Information Fusion, 31: 100-109. https://doi.org/10.1016/j.inffus.2016.02.001

[67] Li, H., Ding, W., Cao, X., Liu, C. (2017). Image registration and fusion of visible and infrared integrated camera for medium-altitude unmanned aerial vehicle remote sensing. Remote Sensing, 9(5): 441. https://doi.org/10.3390/rs9050441.

[68] Sheikh, H.R., Bovik, A.C. (2006). Image information and visual quality. IEEE Transactions on image processing, 15(2): 430-444. https://doi.org/10.1109/TIP.2005.859378.

[69] Han, Y., Cai, Y., Cao, Y., Xu, X. (2013). A new image fusion performance metric based on visual information fidelity. Information Fusion, 14(2): 127-135. https://doi.org/10.1016/j.inffus.2011.08.002.

[70] Chen, H., Varshney, P. K. (2007). A human perception inspired quality metric for image fusion based on regional information. Information Fusion, 8(2): 193-207. https://doi.org/10.1016/j.inffus.2005.10.001

[71] Toet, A. (2014). TNO Image Fusion Dataset. https://figshare.com/articles. TN_Image_Fusion_Dataset/1008029.

[72] Jia, X., Zhu, C., Li, M., Tang, W., Zhou, W. (2021). LLVIP: A visible-infrared paired dataset for low-light

vision. In 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, pp. 3489-3497. https://doi.org/10.1109/ICCVW54120.2021.00389

[73] Hwang, S., Park, J., Kim, N., Choi, Y., So Kweon, I. (2015). Multispectral pedestrian detection: Benchmark dataset and baseline. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, pp. 1037-1045. https://doi.org/10.1109/CVPR.2015.7298706

[74] Tang, L., Yuan, J., Zhang, H., Jiang, X., Ma, J. (2022). PIAFusion: A progressive infrared and visible image fusion network based on illumination aware. Information Fusion, 83: 79-92. https://doi.org/10.1016/j.inffus.2022.03.007

[75] Ram Prabhakar, K., Sai Srikar, V., Venkatesh Babu, R. (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, pp. 4724-4732. https://doi.org/10.1109/ICCV.2017.505

[76] Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., Zhang, L. (2020). IFCNN: A general image fusion framework based on convolutional neural network. Information Fusion, 54: 99-118. https://doi.org/10.1016/j.inffus.2019.07.011

[77] Xu, H., Ma, J., Jiang, J., Guo, X., Ling, H. (2020). U2Fusion: A unified unsupervised image fusion network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(1): 502-518. https://doi.org/10.1109/TPAMI.2020.3012548

[78] Li, H., Wu, X.J., Kittler, J. (2021). RFN-Nest: An end-to-end residual fusion network for infrared and visible images. Information Fusion, 73: 72-86. https://doi.org/10.1016/j.inffus.2021.02.023

[79] Ma, J., Yu, W., Liang, P., Li, C., Jiang, J. (2019). FusionGAN: A generative adversarial network for infrared and visible image fusion. Information Fusion, 48: 11-26. https://doi.org/10.1016/j.inffus.2018.09.004

[80] Li, G., Lin, Y., Qu, X. (2021). An infrared and visible image fusion method based on multi-scale transformation and norm optimization. Information Fusion, 71: 109-129. https://doi.org/10.1016/j.inffus.2021.02.008