International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(7s) 2026 pp. 885-893
https://doi.org/10.70917/ijcisim-2026-3159
Article

Current Trends in Infrared-Visible Image Fusion (IVIF) Techniques, Evaluation Metrices and Open Issues

Hemali Lalitbhai Vithalani[1], Prof. Rajeev Mathur[2]

[1] Computer Engineering Dept., Research Scholar, Indus University, Ahmedabad, Gujarat, India, 382115.
[2] Dept of Computer Science, IICT – Indus University, Ahmedabad, Gujarat, India, 382115.
[1] vithalanihemali.24.rs@indusuni.ac.in;
[2] director@iict.indusuni.ac.in, mathur.rajeev96@gmail.com ;

Abstract— Infrared-visible image fusion (IVIF) is a major effort in computer vision meant to enhance the representation of pictures by combining the unique qualities of the infrared and visible wavelengths. Though they can run into problems like feature redundancy, complexity, and computational inefficiency, conventional fusion methods rely on hand-crafted algorithms to find and combine important features. Recent developments in deep learning have significantly improved fusion performance by means of adaptive feature extraction and the elimination of computing overhead. Still, there is a clear lack of a thorough assessment of DL-based fusion techniques and their impact on the quality and efficiency of picture fusion. Emphasizing deep learning approaches, this study presents a thorough investigation of IVIF methods. This paper addresses major issues including data compatibility, perceptual accuracy, and evaluation constraints while exploring a variety of fusion methods and assessing their advantages and limitations. While identifying potential paths for research—such as simplified network design and hardware-efficient fusion techniques, the report underlines the merits and shortcomings of present methods. Establishing a basis for future development in IVIF technology, this work is a vital tool for academics and professionals.
Keywords--- CNN, GAN, Image Fusion, Infrared Images, Transformers, Visible Images

1. Introduction

Image fusion involves processing varied photos captured by various kinds of devices using defined techniques to create a fused image that incorporates various data from multiple inputs [1]. This image is more readily perceived by the human visual system and facilitates subsequent advanced vision tasks, including Image partitioning, object recognition, and tracking. The redundancy and complementarity features of the merged image allow it to surpass the main image's resolution, physical qualities, and data volume limits. In addition to improving the aggregation degree of picture data, it can reduce noise [2]. It is now very challenging to acquire and transmit high-quality photographs owing to the extensive impact of factors such as weather, environment, camera equipment, and others. So, there are a lot of noises in the pictures, and it's hard to tell what the items are, which impacts a few uses. One of the methods to enhance picture quality is image fusion, which uses a set of principles to process numerous images in a way that improves image quality the most. It also understands how different images complement one another. Image fusion has long been a popular area of study, and for good reason: it offers many benefits over competing methods [2]. Improving the efficacy of picture feature extraction, categorization, and object recognition can be achieved by making the most valuable feature information to convey the scene in a more succinct and redundant way. When it comes to image fusion, the goals and needs of various applications vary, necessitating the development or selection of novel fusion methods [2]. Figure 1 shows the various areas where image fusion is being adopted in wide range.

Conventional techniques to imagine fusion fall into one of five broad classes: those based on digital signal processing, sparse representation, multi-scale transformation, subspace analysis, or hybrid approaches. Image fusion can be classified depending to several sensors, including visual, infrared, polarization, lowlight, hyperspectral, multispectral, and panchromatic imaging, as well as Synthetic-aperture radar (SAR). Additionally, spatial and transform domains (such as time and frequency) can be used to group various domains. There are three degrees of picture fusion, categorized by the amount of information abstraction: pixel, feature, and decision [2]. Figure 2 displays the categorization of Image Fusion on the basis of various criteria.

By merging many source images into one, IVIF may boost picture clarity and eliminate extraneous data. This method enhances the visual capabilities of both humans and robots and is commonly used in imaging systems. Natural objects emit thermal radiation, which consists of electromagnetic waves of varying frequencies and is

undetectable to the naked eye. Improving the quality of visuals for use in computer vision and human observation is the primary goal of integrating IR and visible images. Despite being vulnerable to low light and bad weather, IR sensors can nevertheless detect objects' thermal radiation. But IR pictures don't always do a good job of showing the background, while visible pictures have better spatial resolution, more detail, and texture data. To get precise picture of the scene, IR and visible data are combined into one image. As illustrated in Figure 2, major categories of image fusion algorithms are: classical, deep learning-based, and hybrid [3] [4]. All three types have seen substantial advancements in recent years. The most popular and extensively researched traditional fusion methods for image fusion are those that use multi-scale transformation [5] [6]. Still, IVIF applications still faces several obstacles. To improve the fusion image's level of detail, the most common method is to combine the source image's most prominent features with those of the target image. But IR heat radiation data is mostly defined by pixel intensity, whereas visual picture texture detail data is determined by edges and gradients. Due to the diverse imaging properties of the input image, the fused image must be represented by carefully chosen classical fusion rules that are manually developed. When using the same method, there's a chance that the fused image will have artifacts due to the lack of diversity in the extracted features. In addition, using manual fusion criteria in the context of combining images from many sources usually makes the process more complicated. Considering these concerns, the deep learning-based picture fusion method can use an adaptive mechanism to assign model weights. This strategy drastically reduces the computational expense, which is crucial in many fusion rules, in contrast to the design principles of conventional systems [7]. The main objective of this work is to thoroughly scrutinize the existing IVIF techniques, as well as to discuss possible future advancements and difficulties in this area [8] [9].

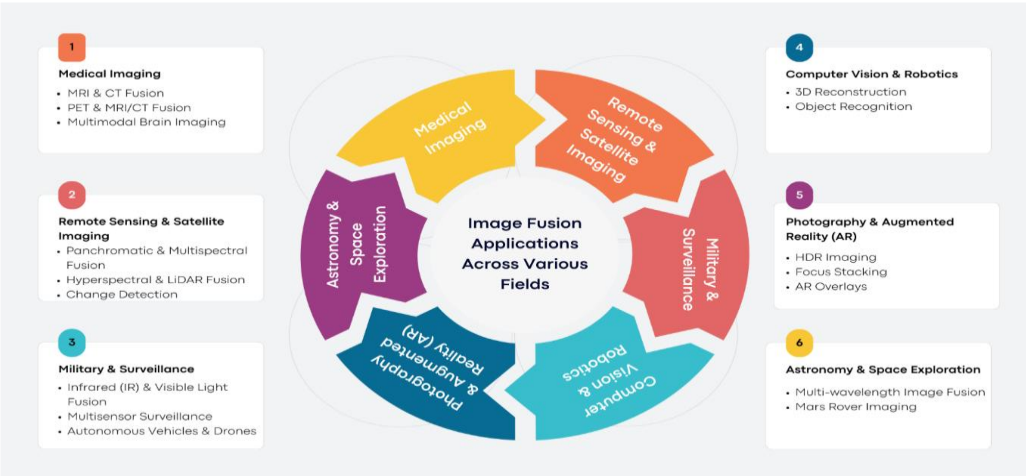

FIGURE 1. IMAGE FUSION APPLICATIONS

The upcoming section discusses overview of IVIF Techniques along with relative comparison of IVIF techniques. Section 3 shows the various performance metrices which are being used for evaluation of fused images. Section 4 discusses open issues and future directions for IVIF techniques.

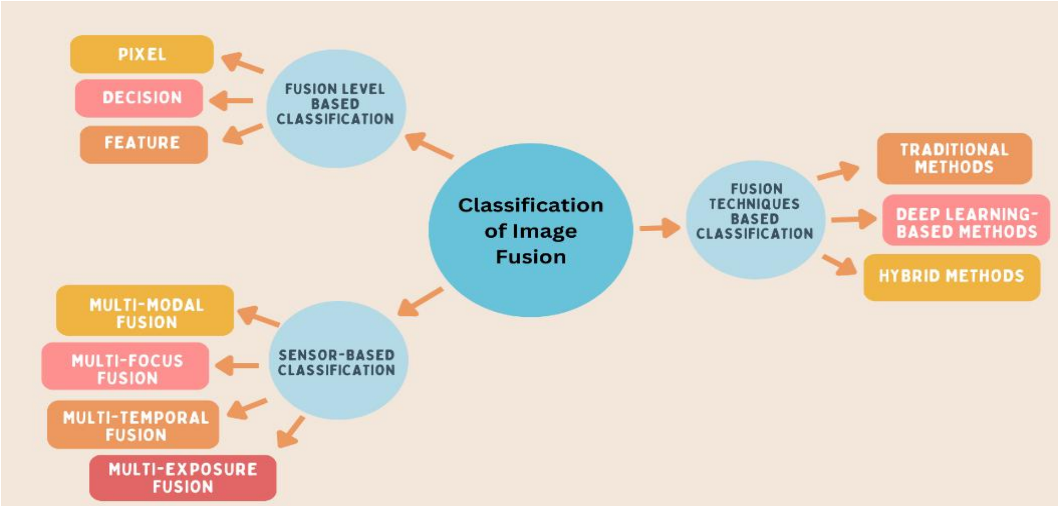

FIGURE 2. CLASSIFICATION OF IMAGE FUSION

2. INFRARED AND VISIBLE IMAGE FUSION (IVIF) TECHNIQUES

Merging IR and visual images means putting together several images into a one-cohesive image. The goal is to improve the quality of the image generally and get rid of unnecessary data. Many imaging technologies use this technique to enhance the visual capabilities of both people and robots. Objects in the natural environment create electromagnetic waves at various frequencies, called thermal radiation, which the human eye remains blind to [10] [11]. An IR sensor collects IR pictures by seeing and recording the heat radiation emitted by various items. The pictures have use in seeing ground items and evaluating surface qualities including the finding of hidden targets and the acknowledgement of camouflage [12]. IR pictures' characteristics would offset the effects of different environmental conditions, smog, and sunlight as well as other influences [13]. Visual images fail to properly represent things and locations with particular IR thermal characteristics outside the bounds of IR photography [14] [15]. Visual pictures are collected to record the visible reflected characteristics of the spectrum data of objects. The pictures offer a thorough representation of a scene in line with human visual qualities by showing different item features and noticeable edges. Night-vision imaging equipment commonly uses IVIF methods to enhance the nighttime capabilities for both people and robots [16].

As discussed in previous section, there are a few broad types of traditional picture fusion methods: a variety of hybrid methods based on digital signal processing, sparse representation, multi-scale transformation, subspace analysis, and other similar techniques The traditional techniques make use of domain-specific previous information. Nevertheless, there are significant limits to older methods in order to enable image fusion from various perspectives. To begin with, they involve the hand-design of complicated feature extraction and fusion rules, which increases algorithmic complexity and calls for a lot of skill. When designing fusion rules, designers frequently rely on their own knowledge, experiences, and experimental judgments without providing any theoretical foundation. Secondly, the model's capacity to generalize is negatively impacted since the features that were manually built are mainly local and at a lower level, which makes it difficult to capture the high-level semantic information that is in the images [1].

Recent years have seen progress in the area of image processing with neural network technology. Because of their great feature extraction capabilities, many different types of neural network algorithms have been used for image fusion. These methods, such as, PCNN methods, Convolutional Neural Network fusion techniques, GAN methods, Autoencoder strategies, Transformer methods, and other framework fusion techniques can be categorized into various categories. Convolutional neural networks (CNNs) achieve better image quality and feature extraction. Moreover, the integration of IR and optical data with LiDAR and hyperspectral photography is becoming increasingly popular to improve environmental awareness and application versatility. Figure 3 shows the evolution and future of IVIF.

Additionally, refining evaluation criteria is essential for objectively assessing fusion quality and performance across various domains, which can significantly benefit advancements in surveillance, medical diagnosis, and autonomous systems. These interconnected themes aim to optimize image processing capabilities for complex scenarios and real-time applications. Numerous techniques have emerged in recent years within the realm of image fusion to merge high-quality and rapid images. Table 1 shows the comparative analysis of various image fusion techniques.

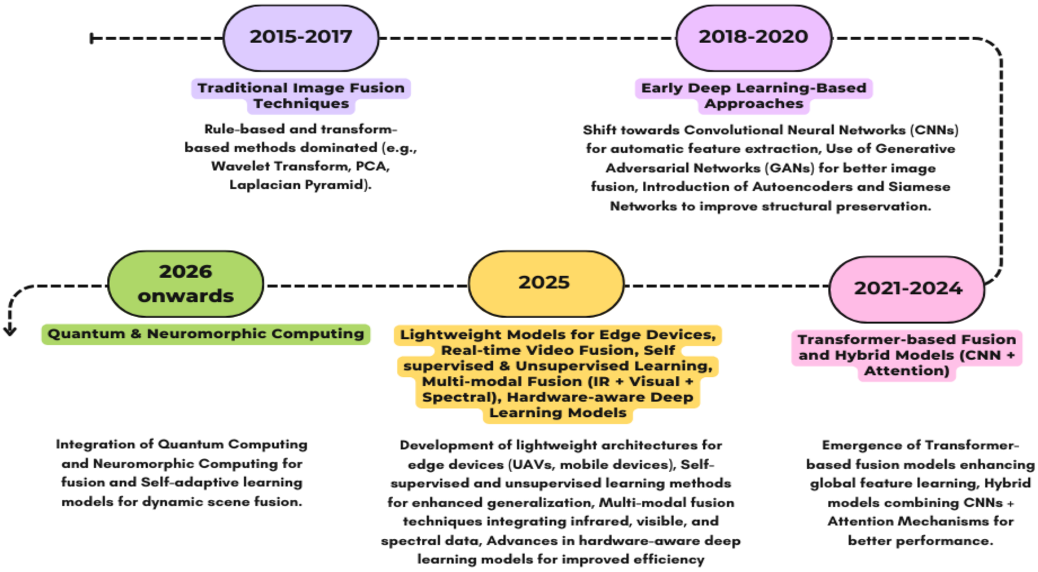

FIGURE 3 EVOLUTION AND FUTURE OF IR AND VISIBLE IMAGE FUSION TECHNIQUES

TABLE 1 COMPARATIVE ANALYSIS OF VARIOUS IMAGE FUSION TECHNIQUES

Reference	Technique	Working	Advantages	Challenges	Suitable Performance Metrices
[17], [18], [19], [20], [21]	CNN based Technique	•Apply convolutional layers to extract features. •Automatically learn texture and spatial properties. •Use VGG-19 and other pre-trained models frequently. •Two fusion strategies are maximal selection and weighted averages.	•The feature extraction process is quite efficient. •Works well for maintaining texture. •Simple incorporation utilizing time-honored fusion methods.	•Awareness of the global context is lacking. •Network depth can potentially lead to performance degradation. •Expensive to compute for photos of a large scale.	•SF •EN •CC •SSIM •Q_AB/F
[18], [19], [22]	Siamese Networks	•Use two networks that share weights to handle pairs of inputs. •Pay close attention to maintaining regional similarities. •A The use of weighted averages is common in fusion techniques.	•Keeps similar features at the local level. Successfully pairs features from IR and visible photos. •Completes tasks involving data from multiple sources efficiently.	•Misalignment is very noticeable. •Training on huge datasets is necessary. •Not as adaptable to datasets with a lot of variation.	•SSIM •PSNR •MSE
[18], [23], [24], [25]	GAN-Based Methods	•Make use of discriminator and generator models in adversarial training. •Produce merged photos that are visually identical to the	•Generates realistic fused images of superior quality. •Automatically learns fusion rules from data.	•Instability in training because of the combative aspect. •A significant amount of training data is needed.	•EN •SSIM •Q_AB/F •

		originals. •Make use of loss functions such as detail loss and target edge loss.	•Retains intricate structural elements.	•Artifacts could be introduced into the fusion findings.	
[26], [27], [25], [28]	Autoencoders	•Make use of latent representations to encode input images. •Then, decode them to create fused images. •It is commonly employed alongside CNNs and DenseNet.	•Efficient fusion is facilitated by dimensionality reduction. •Records both surface-level and deep-level characteristics. •Works well for tasks that do not involve supervision.	•Possible deterioration of features in more complex layers. •Decoder architectures that are complex may be necessary.Hyperparameter adjustment has an impact.	•MSE •SSIM •PSNR •EN
[26], [29], [30], [31], [32]	Transformer-Based Methods	•To capture long-range dependencies, use self-attention methods. •Represent inter-feature connections on a global scale. •Can run hybrid feature extraction with convolutional neural networks (CNNs).	•Proficient understanding of international contexts. •Efficiently manages merging of large-scale images. •Fusion decisions are now easier to understand.	•Excessive computational difficulty. •To avoid overfitting, training must be done using large amounts of data. •Performance issues with inference when used in real-time scenarios.	•EN •SSIM •CC •SF
[33], [34]	Hybrid Methods	•To make the most of each architecture's characteristics, combine them (for instance, convolutional neural networks (CNNs) with autoencoders or transformers). •Try to get good shots of both nearby and faraway features. •A Frequently crafted to tackle certain fusion obstacles.	•A stronger approach to fusion. •This feature extraction method strikes a good balance between local and global extraction. •The capacity to generalize is enhanced.	•Model complexity has increased. •Integrating various architectures must be done with great care. •It is more difficult to optimize.	

3. PERFORMANCE METRICES FOR IMAGE FUSION TECHNIQUES

Recently, IR and visible light picture fusion has become all the rage in the image fusion community. The evolution of image fusion technology has made it crucial to evaluate several suggested approaches linked to image fusion. Though, different fusion processes' practical use shows significant variation and these techniques themselves show particular characteristics [35] [36]. The two primary types into which the present integration indicators fall are subjective and objective assessment [18] [37].

The subjective assessment can be further categorized into the absolute evaluation and relative evaluation, which is carried out using a commonly accepted five-level quality scale and obstacle scale [38]. Human eye observation is the most direct kind of subjective assessment. By looking at picture details, contrast, image integrity, and level of distortion, humans compare several fusion techniques [39]. The assessment standards allow subjective judges to immediately score qualitatively on the combined image.

Still, various people have varied assessment standards for the same picture, which personal choice, surroundings, and other elements easily influence, hence generating erroneous reactions to the fused image. The quality of this approach is bad and the timeliness is low, which does not support a multidimensional assessment of fusion pictures. To assess the fusion outcomes fairly, one must integrate several objective evaluation criteria [39].

Objective assessment helps prevent issues in subjective assessment. Researchers can accurately compare image fusion and provide quantitative reference for evaluation criteria with various emphases. Defined by certain computable mathematical formulas, numerous types of objective fusion indicators exist. By replicating the human visual perception system, the target evaluation is to create a similar mathematical model and then compare the performance of each fusion method using statistical data [40] [41]. Objective image quality measurement comprises the use of non-reference standard images as well as reference standard images. Below are the various evaluation metrices [18] [1] [42]:

- Entropy (EN): ENs are a kind of statistical feature that might reflect the usual data density of an image.
- Spatial Frequency (SF): SF is derived from the picture gradient; it then shows the degree of detail and texture sharpness in the image. Two subsets of SF are spatial column frequency (CF) and spatial row frequency (RF).
- Similarity (COSIN): COSIN transforms the related picture into a vector and computes the cosine similarity between the vectors.
- Correlation Coefficient (CC): The CC gauges the degree of linear correlation between a fused image and IR and visible images.
- Standard Deviation (SD): A measure of the depth of image detail, standard deviation (SD) is a statistical tool. Standard deviation allows one to measure the pixel intensity fluctuation in the composite image.
- Structural Similarity Index Measure (SSIM): In order to provide an overall similarity measure, SSIM first models picture loss and distortion, then compares and assesses the similarity in three elements of image brightness, contrast, and structure. The image's distortion in three areas is seen as an objective evaluation based on the change of image structural information.
- Mutual Information (MI): MI allows one to gauge the data flow from the original image to the fusion image or the data the fusion image has acquired from the original image. MI shows the statistical dependency of two independent variables viewed via information theory.
- Average Gradient (AG): AG could gauge the capacity to express the sharpness of the picture as well as the texture and detail of the fused image.
- Mean Squared Error (MSE): Mean Squared Error (MSE) is a metric representing the difference between two images by means of the error between the fused and source images.
- Gradient-Based Fusion Performance (Q_AB/F): It assesses the level of edge information retention in the fusion image using local metrics from the source image.
- Peak Signal-to-Noise Ratio (PSNR): As name suggests, it is used to determine the fusion image's peak-to-noise ratio.

4. OPEN ISSUES AND FUTURE DIRECTIONS

Supervised learning neural network models have difficulty classifying fused images, which limits their application in image fusion. It is also challenging to get numerous precisely matched paired training samples. Public datasets of IR and visible pictures such as TNO, RoadScene, MSRS, LLVIP, and M3FD lack alignment precision, resolution, scene categories, and image count, which hinders fusion network accuracy and generalization [1] [43] [44] [45] [46].

The majority of algorithms, whether they use classic methods or neural network approaches, have trouble dealing with input images that aren't aligned properly between visible and IR light. Even little errors or misalignments in one modality might lead to ghosting or other geometric abnormalities in the combined image. Geometric distortion, rotation, scale, and disparities in viewpoint are obstacles that must be overcome in order to attain temporal and spatial coherence [1] [47] [48] [49].

Designing complex loss functions, activity level metrics, and fusion rules for unsupervised neural network models calls for great prior knowledge on the part of researchers. Establishing fusion network frameworks is now unguided; academics are creating them by trial and error, which is a time-consuming, labor-intensive, and unpredictable process. Connecting the best solution to the network design by combining conventional fusion techniques or using their knowledge to direct network development could help to enhance image fusion results [1] [45] [48].

The image fusion evaluation index still has some serious flaws. Despite being plagued by issues like noise and information redundancy, picture fusion frequently displays a stellar performance in certain objective indices. Consequently, in order to raise the bar, additional research into the objective evaluation indicators is required [2] [43] [50] [48].

Objective metrics cannot accurately assess image quality. Despite having many white borders and blurring, fusion images with noise—such as those from DDcGAN—outperform other methods in EN and SD. Evaluating fusion quality could be done using distance metric learning. In this field, future studies on an appropriate goal evaluation system are urgently needed [1] [43] [46] [42].

Algorithms typically operate on static images while fusing them, meaning they aren't always up to the task of processing moving sceneries or films. Using the various Spectral and Polarization Imaging Fusion (SPF), fusion strategies can be enhanced in the future to adapt to dynamic situations captured in low light, haze, or adverse weather [2][42] [48].

The practical use of combining IR and visible images depends on efficient functioning. Although a bigger neural network generates better-quality pictures, it also demands more processing power, hence restricting its practical application. Still required are strong GPUs for optimization techniques such as Atrous Convolution (AC), and Neural Architecture Search. Future studies should concentrate on lightweight network architecture for devices with limited resources, such handheld systems and UAVs, thereby solving these challenges. More efficient designs and fusion algorithms that are friendly to hardware are required to fit practical limits. Deployment and practical use depend on algorithms being compatible with hardware, particularly if they are developed for server-based systems like as PyTorch and TensorFlow [43][44] [45].

5. CONCLUSION

By merging IR and visible picture data, which enhance one another, IVIF enhances scene representation. Though deep learning has improved fusion performance, some issues remain, including misaligned datasets, inadequate evaluation standards, and high computational needs. Traditional and neural network-based methods suffer from problems with geometric distortions, real-time adaptability, and feature redundancy.

Future research should focus on lightweight designs for devices with limited resources, fusion techniques that complement hardware, and improved dataset alignment. Using multi-modal fusion techniques and creating enhanced evaluation criteria can help IVIF to be more efficient and practical as well. Dealing with these challenges will help to build more practical and scalable fusion systems.

References

[1] K. Yang, W. Xiang, Z. Chen, J. Zhang and Y. Liu, "A review on infrared and visible image fusion algorithms based on neural networks," Journal of Visual Communication and Image Representation, vol. 101, p. 104179., 2024.

[2] S. Karim, G. Tong, J. Li, A. Qadir, U. Farooq and Y. Yu, "Current advances and future perspectives of image fusion: A comprehensive review," Information Fusion, vol. 90, pp. 185-217, 2023.

[3] F. Zhao, W. Zhao and H. Lu, "Interactive feature embedding for infrared and visible image fusion," IEEE Transactions on Neural Networks and Learning Systems, 2023.

[4] Y. Liu, X. Chen, J. Cheng, H. Peng and Z. Wang, "Infrared and visible image fusion with convolutional neural networks," International Journal of Wavelets, Multiresolution and Information Processing, vol. 16, no. 3, p. 1850018, 2018.

[5] K. Xiao, X. Kang, H. Liu and P. Duan, "MOFA: A novel dataset for multi-modal image fusion applications," Information Fusion, vol. 96, pp. 144-155, 2023.

[6] Y. Liu, B. Xu, M. Zhang, W. Li and R. Tao, "Sub Regional Infrared Visible Image Fusion Using Multi Scale Transformation," Journal of Beijing Institute of Technology, vol. 31, no. 6, pp. 535-550, 2022.

[7] L. Tang, H. Zhang, H. Xu and J. Ma, "Deep learning based image fusion: a survey," Journal of Image and Graphics, vol. 28, no. 1, pp. 3-36, 2023.

[8] J. Ma, Y. Ma and a. C. Li, "Infrared and Visible Image Fusion Methods and Applications: A Survey," Information fusion, vol. 45, pp. 153-178, 2019.

[9] X. Bi, Y. Li, W. Ji, J. Wang, R. Fu and Z. Miao, "FIVFusion: Fog free Infrared and Visible Image Fusion," Journal of King Saud University Computer and Information Sciences, vol. 37, no. 9, pp. 1-25, 2025.

[10] B. Zhang, X. Lu, H. Pei and Y. Zhao, "A fusion algorithm for infrared and visible images based on saliency analysis and non-subsampled Shearlet transform," Infrared Physics & Technology, vol. 73, pp. 286-297, 2015.

[11] H. Jin and Y. Wang, "A fusion method for visible and infrared images based on contrast pyramid with teaching learning based optimization," Infrared Physics & Technology, vol. 64, pp. 134-142, 2014.

[12] Z. Liu, H. Yin, B. Fang and Y. Chai, "A novel fusion scheme for visible and infrared images based on compressive sensing," Optics Communications, vol. 335, pp. 168-177, 2015.
[13] S. Li, X. Kang, L. Fang, J. Hu and H. Yin, "Pixel-level image fusion: A survey of the state of the art," Information Fusion, vol. 33, pp. 100-112, 2017.
[14] G. Cui, H. Feng, Z. Xu, Q. Li and Y. Chen., "Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition.," Optics Communications , vol. 341, pp. 199-209, 2015.
[15] X. Fan, P. Shi, J. Ni and M. Li, "A thermal infrared and visible images fusion based approach for multitarget detection under complex environment," Mathematical Problems in Engineering , vol. 1, p. 750708, 2015.
[16] G. Bhatnagar and Z. Liu, "A novel image fusion framework for night-vision navigation and surveillance," Signal, Image and Video Processing, vol. 9, pp. 165-175, 2015.
[17] H. Li, X.-J. Wu and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in 2018 24th international conference on pattern recognition (ICPR), 2018.
[18] C. Sun, C. Zhang and N. Xiong, "Infrared and visible image fusion techniques based on deep learning: A Review," Electronics, vol. 9, no. 12, p. 2162, 2020.
[19] H. Li, X.-j. Wu and T. S. Durrani, "Infrared and visible image fusion with ResNet and zero-phase component analysis," Infrared Physics & Technology , vol. 102, p. 103039, 2019.
[20] X. Xu, G. Yang, X. Zhang, J. Zhang, X. Chen and a. A. Liu, "ICFusion: Multi-modal Image Fusion via Invertible Feature Enhancement and Cross-modal Feature Balance," IEEE Sensors Journal, 2025.
[21] Y. Wang, Z. Li, J. Wang, L. Yang, B. Dong, H. Zhang and a. J. Liu, "MFF: A Deep Learning Model for Multi-Modal Image Fusion Based on Multiple Filters," IEEE Access, 2025.
[22] W. Xing, D. Chen, M. A. Islam and a. J. Zhou, "Bidfuse: Harnessing Bi-Directional Attention with Modality-Specific Encoders for Infrared-Visible Image Fusion," in IEEE International Conference on Image Processing (ICIP), 2024.
[23] J. Ma, W. Yu, P. Liang, C. Li and J. Jiang, "FusionGAN: A generative adversarial network for infrared and visible image fusion," Information fusion, vol. 48, pp. 11-26, 2019.
[24] J. Ma, P. Liang, W. Yu, C. Chen, X. Guo, J. Wu and Jiang, "Infrared and visible image fusion via detail preserving adversarial learning," Information Fusion, vol. 54, pp. 85-98, 2020.
[25] J. Ma, H. Xu, J. Jiang, X. Mei and X.-P. Zhang., "DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion," IEEE Transactions on Image Processing, vol. 29, pp. 4980-4995, 2020.
[26] H. Wang, L. Li, C. Li and X. Lu, "Infrared and visible image fusion based on autoencoder composed of CNN-transformer," IEEE Access, vol. 11, pp. 78956-78969, 2023.
[27] Z. Zhao, S. Su, J. Wei, X. Tong and W. Gao., "Lightweight infrared and visible image fusion via adaptive DenseNet with knowledge distillation," Electronics, vol. 12, no. 13, p. 2773, 2023.
[28] D. Wu, X. T. H. Wang, Z. Zhao, Y. Ge, S. Hu and a. R. Liang, "DMF-LP: Enhancing image quality through multimodal medical image fusion using diffusion and Laplacian techniques," Biomedical Signal Processing and Control, vol. 106, p. 107890, 2025.
[29] Z. Zhao, L. Deng, H. Bai, Y. Cui, Z. Zhang, Y. Zhang, H. Qin, D. Chen, J. Zhang and P. Wang, "Image fusion via vision-language model," arXiv preprint arXiv:2402.02235, 2024.
[30] C. Xia, H. Gao, W. Yang and a. J. Yu., "MSDT: Multiscale Diffusion Transformer for Multimodality Image Fusion," IEEE Transactions on Emerging Topics in Computational Intelligence, 2025.
[31] M. Zhu, X. Chen, X. Wei, N. Wang and X. Gao, "TeSG: Textual Semantic Guidance for Infrared and Visible Image Fusion," arXiv preprint arXiv:2506.16730 , 2025.
[32] R. Zhang, X. He, K. Cao, L. Liu, L. Zhang, M. Zhou and a. J. Zhang, "Distilling Textual Priors from LLM to Efficient Image Fusion," arXiv preprint arXiv:2504.07029, 2025.
[33] D. Guan, Y. Wu, T. Liu, A. C. Kot and Y. Gu, "Rethinking the Evaluation of Visible and Infrared Image Fusion," arXiv preprint arXiv:2410.06811, 2024.
[34] Y. Liu, C. Wu, J. Cui and B. Jiang, "VPCI: Self-Supervised Visual Prompt-Guided Cross-Domain Interactive Image Fusion Framework.," in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2025.
[35] H. Kaur, D. Koundal and V. Kadyan, "Image Fusion Techniques: A Survey," Archives of computational methods in Engineering, vol. 28, no. 7, pp. 4425-4447, 2021.
[36] S. Singh, N. Mittal and H. Singh, "Review of various image fusion algorithms and image fusion performance metric."," Archives of Computational Methods in Engineering , vol. 28, no. 5, pp. 3645-3659, 2021.
[37] V. Singh, "Image Fusion Metrics–A Systematic Study," in Proceedings of the 2nd International Conference on IoT, Social, Mobile, Analytics & Cloud in Computational Vision & Bio-Engineering (ISMAC-CVB 2020), 2020.
[38] X. Jin, Q. Jiang, S. Yao, D. Zhou, R. Nie, J. Hai and K. He., "A survey of infrared and visual image fusion methods," Infrared Physics & Technology, vol. 85, pp. 478-501, 2017.
[39] P. Jagalingam and A. V. Hegde, "A Review of Quality Metrics for Fused Image," Aquatic Procedia, vol. 4, pp. 133-142, 2015.
[40] C. S. Xydeas and V. Petrovic, "Objective image fusion performance measure.," Electronics letters, vol. 36, no. 4, pp. 308-309, 2000.
[41] H. B. Mitchell, Image fusion: theories, techniques and applications., Springer Science & Business Media, 2010.
[42] SI, M. MA and E.-T. GS, "Multimodal medical image fusion algorithm based on pulse coupled neural networks and nonsubsampled contourlet transform.," Medical & Biological Engineering & Computing, vol. 61, no. 1, pp. 155-177, 2023.

[43] J. Liu, G. Wu, Z. Liu, D. Wang, Z. Jiang, L. Ma, W. Zhong and X. Fan, "Infrared and visible image fusion: From data compatibility to task adaption," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 47, no. 4, pp. 2349-2369, 2024.
[44] Y. Luo and Z. Luo., "Infrared and visible image fusion: Methods, datasets, applications, and prospects," Applied Sciences , vol. 13, no. 19, p. 10897, 2023.
[45] B. Xu and J. Li, "Design and Evaluation of Deep Learning-Based Dual-Spectrum Image Fusion Methods," arXiv preprint arXiv:2506.07779 , 2025.
[46] R. M. Neamah and T. A. Al-Asadi, "The Deep Learning Methods for Fusion Infrared and Visible Images: A Survey.," Revue d'Intelligence Artificielle , vol. 38, no. 2, 2024.
[47] D. Wang, J. Liu, X. Fan and R. Liu, "Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration," arXiv preprint arXiv:2205.11876, 2022.
[48] W. Ma, K. Wang, J. Li, S. X. Yang, J. Li, L. Song and a. Q. Li, "Infrared and visible image fusion technology and application: A review," Sensors, vol. 23, no. 2, p. 599, 2023.
[49] C. Vipparla, T. Krock, K. Nouduri, J. Fraser, H. AliAkbarpour, V. Sagan, J.-R. C. Cheng and P. Kannappan, "Fusion of Visible and Infrared Aerial Images from Uncalibrated Sensors Using Wavelet Decomposition and Deep Learning," Sensors, vol. 24, no. 24, p. 8217, 2024.
[50] M. N. Narsaiah, S. Vathsal and a. D. V. Reddy, "A survey on image fusion Requirements, techniques, evaluation metrics, and its applications," Int. J. Eng. Technol , vol. 7, no. 2, pp. 260-266, 2018.