

¹Priyanka
Bhatambarekar

²Gayatri Phade

Exploring the Terrain: An Investigation into Deep Learning-Based Fusion Strategies for Integrating Infrared and Visible Imagery



Abstract: - Infrared and visible image fusion technologies influence distinct image features acquired from distinct sensors, preserving complementary information from input images throughout the process of fusion, and utilizing redundant data to enhance the quality of the resulting fused image. Recently, deep learning methods (DL) have been employed by numerous researchers to investigate image fusion, revealing that the application of DL significantly enhances the efficiency of the model and the quality of fusion outcomes. Nevertheless, it is very important to note that DL can be implemented in various branches, and currently, a comprehensive investigation of deep learning-based methods in image fusion is in process. The paper aims to provide an exhaustive review of the evolution of image fusion algorithms grounded in deep learning over the years. Precisely, this paper undertakes a particular exploration of the fusion techniques applied to infrared and visible images through deep learning methodologies. The investigation includes a qualitative and quantitative comparison of extant fusion algorithms using established quality indicators, along with a thorough discussion of diverse fusion approaches. The current research status concerning infrared and visible image fusion is presented, with a forward-looking perspective on potential future directions. This research makes an effort to contribute valuable insights into various image fusion methods developed in recent years, thereby laying a solid foundation for subsequent research goings-on in this domain.

Keywords: Image Fusion, Deep Learning (DL), Convolutional Neural Networks (CNN), Generative Adversarial Networks (GAN)

I. INTRODUCTION

Following diverse image processing domains, image fusion can be broadly categorized into two domains: transform domain and spatial domain. The primary objective of the fusion methodology is to extract relevant information from the input source image and integrate it effectively. Each fusion method encounters three fundamental challenges: image transformation, activity-level measurement, and design of the fusion rule [33]. Image transformation encompasses diverse multiscale decomposition, various sparse representation techniques, non-down sampling methodologies, and combinations of distinct transformations. Activity level measurement aims to quantitatively acquire information for assigning weights from different sources [12]. Fusion rules encompass significant rules and weighted average rules, the essence of which serves to distribute weights [32]. Given the rapid evolution of fusion algorithms in both theory and application, the essential aspect of image fusion lies in the selection of an appropriate feature extraction strategy. Designing a skillful convolution neural network and adjusting parameters for deep learning-based image fusion remains challenging. Particularly in recent years as there is the introduction of adversarial networks for image fusion. While this approach yields a more apparent fusion effect, careful consideration is essential to address the inherent challenges of gradient disappearance and gradient explosion encountered during adversarial training.

Visible images have the same visual qualities as the human eye, displaying a wealth of rich information and edge features [13]. Visible light sensors help to produce richer image spectrum information, including distinct scene details, texturing, and increased spatial resolution. However, in poor settings such as nighttime, camouflage, hidden objects in smoke, background clutter, etc., the target may be difficult to detect in visible images.

¹ Assistant Professor, Electronics and Telecommunication Engineering, Sandip Institute of technology and research Centre Nashik, Maharashtra, India. Email: priyanka.bhatambrekar@gmail.com

² Professor and Head, Electronics and Telecommunication Engineering, Sandip Institute of technology and Research Centre Nashik,

Maharashtra, India. Email: gphade@gmail.com

Copyright©JES2024on-line:journal.esrgroups.org

In contrast, in usual circumstances, objects emit thermal radiation in the form of electromagnetic waves at various frequencies. This phenomenon, known as thermal radiation, remains imperceptible to the human eye [1]. To detect thermal radiation information, the utilization of distinct sensors is requisite, particularly in the processing of infrared images to extract thermal radiation data, thereby enhancing target detection capabilities [11]. Infrared images offer the advantage of justifying external environmental factors such as sunlight, smoke, and other conditions [1,12]. Nevertheless, they present challenges in terms of low contrast, intricate backgrounds, and suboptimal feature performance.

Consequently, the fusion of infrared and visible light technologies combines the respective advantages of the two modalities, retaining a more comprehensive set of infrared and visible feature information in the resultant fusion [14,35]. The universal applicability and mutual complementarity of infrared and visible images have applied the fusion technology into diverse fields, assuming an increasingly crucial role in the domain of computer vision.

The integration technique that combines visible and infrared images has become widely used these days in many different fields, such as electronics, industrial applications, remote sensing detection, image enhancement, target detection, and recognition, among others. Many visible and infrared image fusion techniques have been put forth recently in the image fusion field. However, there are still issues with applying these fusion techniques to different circumstances involving the fusion of visible and infrared image [36,37]s. To improve information richness, a popular fusion technique incorporates prominent features from the source image into the fusion image. On the other hand, the unique properties of infrared heat radiation, which are mostly expressed in terms of pixel intensity, are in opposition to the textural detail information found in visual images that are defined by gradients and edges. Using conventional manually created fusion rules that are based on the same selection criteria as the fused image could lead to a lack of diversity in features and could introduce artifacts into the fused image. Furthermore, manual fusion criteria add to the growing methodological complexity in the context of multi-source image fusion. The comparison between deep learning-based and conventional image processing is explained in Table 1, which is provided below.

| Parameter | Conventional Image processing | Deep learning-based image processing |
|------------------------|--------------------------------------|---|
| Training dataset | Small | Large |
| Computing power | Low | High |
| Training time | Short | Long |
| Algorithm Transparency | High | Low |
| Assumptions | Few | Many |
| Deployment Flexibility | High | Low |
| Expenditure | Low | High |

Table 1 Comparison between deep learning and traditional image processing

Addressing these issues, the image fusion methodology grounded in deep learning employs an adaptive mechanism to assign weights to the model [34]. In comparison to the rule design inherent in traditional methods, this approach substantially reduces computational costs, a critical consideration in numerous fusion rules. As such, this research paper endeavours to conduct a comprehensive review of existing infrared and visible image fusion algorithms based on deep learning, including their future developmental trajectories and associated challenges.

The paper also presents an overview of the theoretical underpinnings of visible and infrared image fusion, as well as the corresponding fusion evaluation indices. To provide a solid basis for future study, a qualitative and quantitative comparison of studies from relevant literature is offered. In the end, the survey summarizes current fusion approaches and examines potential directions for further research in this field.

II. FUSION OF INFRARED AND VISIBLE IMAGES USING DEEP LEARNING

Recent years have witnessed a large number of innovative approaches utilizing deep learning for the fusion of infrared and visible images. These state-of-the-art methodologies find widespread application in diverse domains such as image pre-processing, target recognition, and image classification. Figure 2 provides a schematic representation of the conventional fusion framework, highlighting two essential factors: feature extraction and feature fusion. The primary theoretical methodologies employed in these algorithms include multi-scale transformation, sparse representation, subspace analysis, and hybrid techniques. However, the inherent limitations of these artificially designed feature extraction methods contribute to the complexity of the image fusion algorithm.

In response to the shortcomings of conventional fusion methods, deep learning techniques are incorporated for feature extraction. The evolving landscape of deep learning has led to the emergence of several fusion methodologies, notably based on convolutional neural networks (CNN), generative adversarial networks (GAN), Siamese networks, and autoencoders within the field of image fusion. Table 1 categorically lists the principal fusion methods discussed in this section. While image fusion outcomes derived from deep learning exhibit commendable performance, it is important that many methods also encounter apparent challenges. Consequently, a detailed exposition of each method will be provided to explicate both their merits and challenges. The deep learning-based image fusion process typically involves various steps in which training the data is of prime importance.

Data training involves the collection and pre-processing of input images earmarked for fusion. It is imperative to ensure that the data is suitably annotated if a supervised learning approach is adopted. Network Architecture Design entails the selection of an appropriate deep-learning architecture tailored for image fusion tasks. Convolutional Neural Networks (CNNs) stand as a prevalent choice for this purpose. The architectural design necessitates the accommodation of multiple input images, culminating in the generation of a fused output.

The Training phase goes through the following steps:

1. Dataset division into training and validation subsets. Model training using the training subset, wherein weight adjustments are made based on the disparity between predicted and ground truth fused images. Employment of a loss function to quantify deviations between predicted and actual fusion outcomes.
- 2 Validation phase: Assessment of the model's performance on the validation subset to ascertain its generalization capabilities to unseen data. Potential refinement of the model is contingent upon validation outcomes.
- 3 The Testing/Inference stage: Utilization of the trained model to fuse novel, unobserved images. Application of acquired transformations to amalgamate features from input images into a fused output.
4. Post-processing, while discretionary, may be employed to refine the final fused image in accordance with specific criteria.
- 5 Evaluation: Evaluation of fused image quality using pertinent metrics such as the structural similarity index and peak signal-to-noise ratio, alongside visual scrutiny.

This systematic approach facilitates the acquisition of intricate patterns and relationships within input images by deep learning models, thereby facilitating the generation of high-quality fused images with enhanced performance relative to conventional methodologies.

A. CNN-Based Fusion Methods

In the area of computer vision, convolutional layers are crucial for feature extraction, often yielding richer information compared to conventional manual methods [55,56]. A fundamental challenge in image fusion lies in the extraction of salient features from source images and their integration to produce a organized fused image. However, the application of Convolutional Neural Networks (CNNs) to image fusion encounters three primary obstacles. Firstly, training an effective network demands ample labelled data, yet architectures for image fusion

based on CNNs are often simplistic, with insufficient convolutional layers to adequately extract features, leading to poor fusion performance. Secondly, the integration of manually devised image fusion rules into an end-to-end model network poses challenges, introducing errors during feature reconstruction that compromise image fidelity. Lastly, conventional CNN algorithms overlook valuable information in the final layer, resulting in incomplete preservation of model features. As network depth increases, the loss of crucial features increases, damagingly impacting the overall fusion outcome.

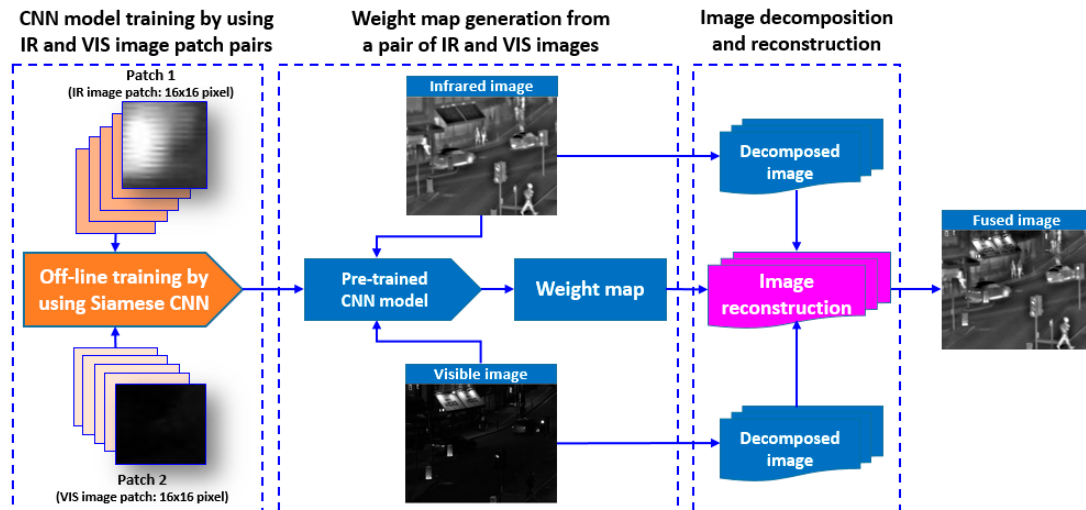


Figure 1. Block diagram of CNN based visible(VIS) and infrared(IR) image fusion algorithm

The study leads the segmentation of source image data into two distinct components: one encompassing low-frequency information and the other comprising texture details. The model under consideration adopts a multilayer fusion strategy inspired by the architecture of the VGG-19 network [24], facilitating the extraction of deep features indicative of intricate content characteristics. In contrast to conventional multiple exposure fusion (MEF) algorithms reliant on manually engineered features for image fusion, the discussed model demonstrates resilience to variations in input conditions without necessitating parameter adjustments, thereby mitigating potential degradation in algorithm robustness and computational overhead associated with processing multiple exposure images. The efficacy of Convolutional Neural Networks (CNNs) is primarily dependent upon the selection of appropriate loss functions. Notably, Prabhakar et al. [25] propose a method obviating the need for parameter tuning in response to input variations. The fusion network architecture comprises three key components: an encoder, a fusion layer, and a decoder, leveraging encoder networks for fusion tasks. Within the CNN paradigm, optimization of loss function parameters enhances the accuracy of result prediction. In a related context, Ma et al. [26] introduce an infrared and visible image fusion approach predicated on minimizing total variation (TV), thereby ensuring alignment of the fused image's pixel intensity with that of the infrared image and its gradient characteristics with those of the visible image. Similarly, Li et al. [27] present a fusion framework integrating deep features and zero-phase component analysis. Initially, a residual network is employed to extract depth features from the source image, followed by normalization utilizing ZCA-zero-phase component analysis and L1-norm to derive an initial weight map. Subsequently, a weighted average strategy is employed for the reconstruction of the fused image.

B. Autoencoder-Based Fusion Method

An autoencoder constitutes a fundamental layer within a neural network framework, wherein the application of backpropagation facilitates the acquisition of an approximation to the identity function, thereby endeavouring to faithfully reproduce its input at its output. Through iterative refinement of network parameters guided by an appropriate cost function, the overarching objective is to discern correlations within high-dimensional datasets, thus engendering improved feature representations thereof. Structurally, stacked autoencoders (SAEs) are composed of multiple layers of such autoencoders, wherein the outputs of each layer are intricately connected to the inputs of subsequent layers, facilitating hierarchical abstraction and representation learning [28].

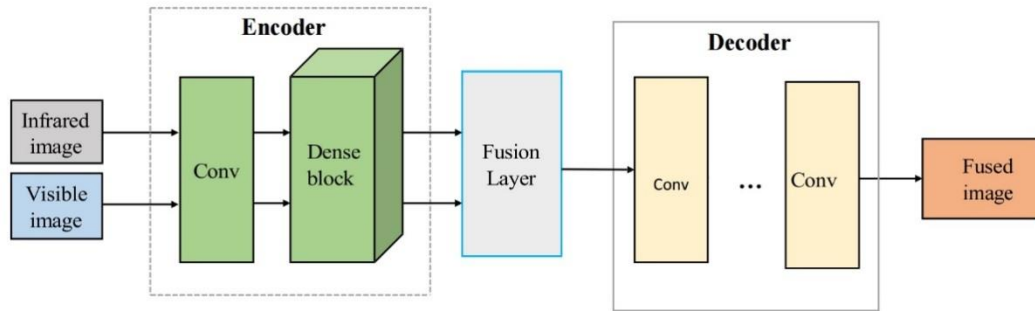


Figure 2. Autoencoder-based infrared and visible image fusion framework [4].

C. Siamese Networks-Based Fusion Methods

One of the inherent challenges in image fusion arises from the disparate imaging modalities employed in capturing infrared and visible images. A pyramid framework is utilized for feature extraction, separate from the infrared and visible images, to guarantee that the final fused image preserves all of the information from both source images.

A Siamese convolutional network designed especially for image fusion tasks was recently proposed by Liu et al. [29]. This network produces a weight map that is essential for the ultimate fusion decision, given two source images as input. To produce an excellent training dataset, a large number of natural photos are processed using random sampling and Gaussian blurring. Notably, this approach emphasizes activity level measurement, with weight assignments concurrently determined by the network. Specifically, convolutional layers and fully-connected layers function analogously to activity level measurement and weight assignment components, respectively, within the context of image fusion.

The proposed model comprises four key steps: passing the infrared and visible images through the convolutional neural network to generate weights, decomposing the source image weights using Gaussian and Laplacian pyramids, and finally fusing the information obtained from each pyramid decomposition via weighted averaging. The operational principle of the Siamese network in the fusion process is elucidated in Figure 3.

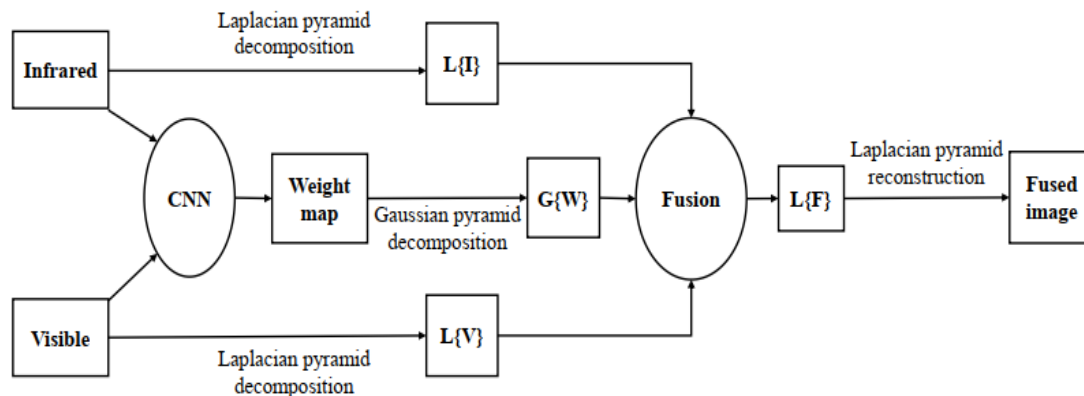


Figure 3. Siamese network-based infrared and visible image fusion scheme [4]

In a related context, Zhang et al. [30] assert the potent feature representation capability of CNNs, albeit acknowledging the time-intensive nature of model training and updating. Hence, they advocate for the utilization of the Siamese network for pixel-level fusion to expedite processing time. The proposed method involves fusing infrared and visible images before subjecting them to the Siamese network for feature tracking.

Moreover, Piao et al. [4] devise an adaptive learning model predicated on the Siamese network, aiming to automatically generate corresponding weight maps based on pixel saliency in the source images. This approach

mitigates the issue of parameter redundancy associated with traditional fusion rules. The proposed methodology employs a three-level wavelet transform to decompose the source image into low-frequency and high-frequency weight maps, subsequently leveraging scaled weight maps for reconstructing the wavelet image and obtaining the fused image.

D. GAN-Based Fusion Methods

Current deep learning-based image fusion methodologies commonly employ Convolutional Neural Network (CNN) models. However, this approach typically necessitates the provision of ground truth data for model training. Yet, in the context of fusing infrared and visible images, establishing definitive fusion image standards proves impractical. Consequently, to avoid reliance on ground truth data, a deep model is trained to assess the degree of blurring in each patch of the source image, subsequently calculating weights accordingly to generate the final fusion image. An alternative approach to addressing these challenges involves employing a generative adversarial network (GAN) for image fusion.

In their work [28], Ma et al. introduced an image fusion method based on generative adversarial network architecture. Here, the generator component primarily focuses on fusing infrared and visible images, while the discriminator's role is to ensure that the fused image preserves more details from the visible image, thereby simultaneously retaining both visible texture information and infrared heat radiation information in the fusion image. Figure 4 illustrates the image fusion framework based on GAN. Despite its effectiveness, fusion GANs encounter limitations wherein vital information from the source images may not be fully retained during the fusion process and considerable computational resources are consumed during convolution operations.

In response to these challenges, the work presented in reference [32] advocates for the adoption of learning group convolution as a means to bolster model efficiency while conserving computational resources. This methodology strives to attain a more optimal equilibrium between model precision and processing velocity. Moreover, the incorporation of residual dense blocks as foundational elements within the network architecture, coupled with the integration of dormant perceptual attributes as features for input content loss, serves to facilitate robust supervision of deep networks.

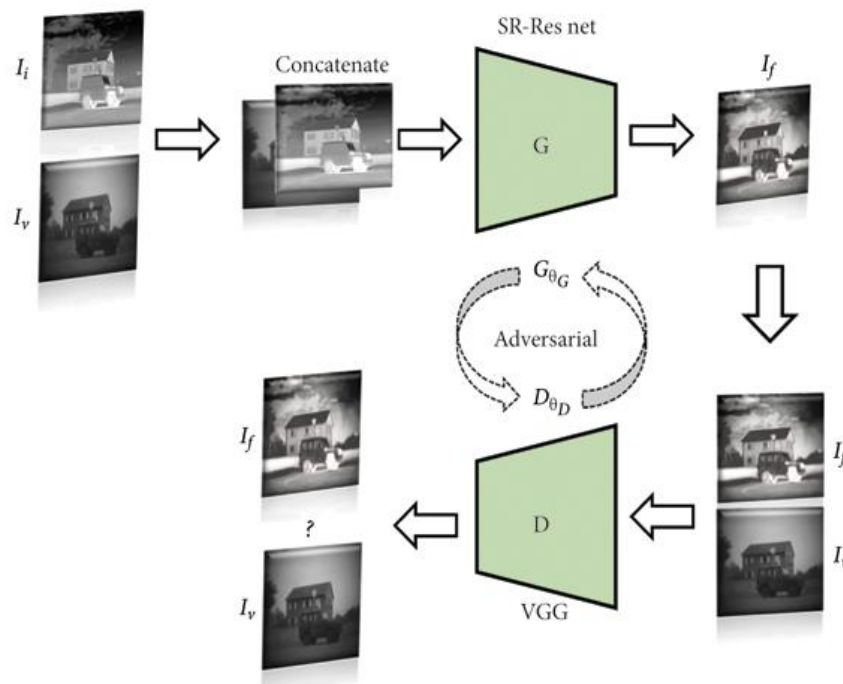


Figure 4. GAN-based infrared and visible image fusion framework

Considering the inherent challenges associated with Convolutional Neural Networks (CNNs), relying solely on adversarial training may lead to the cost of detailed information. Thus, a min-max game is formulated between the generator and discriminator components. The model loss encompasses factors such as detail loss, target edge loss, and confrontation loss.

In a related study, Xu et al. [33] leverage local binary pattern (LBP) analysis to intuitively capture edge information in images. By comparing pixel values between central and surrounding pixels, a fusion image with enhanced boundary delineation is generated. The discriminator in this approach encodes and decodes both the fused image and individual source images, subsequently assessing the disparities in distributions post-decoding.

Additionally, Li et al. [34] adopt a pre-fused image as a reference strategy, enabling the generator to utilize it as a benchmark during the generation process. This ensures that the generated fused image effectively preserves the rich texture from the visible image and thermal radiation information from the infrared image. A comprehensive impression of various image fusion techniques based on genetic algorithms (GA) is presented.

III. EXPERIMENTAL RESULTS

The qualitative examination of experimental outcomes reveals notable distinctions in the fusion performance across various methodologies. Specifically, the fusion results obtained through the guided filtering-based fusion method (GFF) exhibit a pronounced presence of artifacts over a substantial area, coupled with a diminished prominence of the thermal radiation target. In contrast, fusion images generated via fusion GAN demonstrate a heightened presence of texture details, elevated image contrast, and more salient thermal radiation targets compared to alternative approaches. Consequently, a qualitative assessment underscores the imperative for further optimization across the majority of the aforementioned methodologies.

The fusion algorithm can be embedded in various applications and enhance the original method due to the complementary features present in the feature information of infrared and visible images. This fusion technology has been utilized in diverse areas such as target detection, tracking, surveillance, remote sensing, and medical image processing [4].

The qualitative experimental findings reveal notable distinctions among the fusion outcomes of various methods. Specifically, the guided filtering-based fusion method (GFF) exhibits extensive artifacts and lacks the prominence of thermal radiation targets, as depicted in Fig. 5 (c). In contrast, fusion images generated by the dual-discriminator conditional generative adversarial network (DDcGAN) and fusion GAN display heightened texture detailing, increased image contrast, and more pronounced thermal radiation targets, as illustrated in Fig. 5 (e).

Comparison between pairs of infrared images and visible images in the TNO dataset. Quantitative experimental results show that each fusion algorithm has advantages and disadvantages, and different methods show different advantages in different aspects. With deep learning in infrared and visible image fusion, superior new technologies have been continuously emerging to achieve better fusion results. In general, the fusion effect of GFF, fusion GAN, and DDcGAN is better than other methods in terms of brightness, texture detail, and contrast. It can be seen from Figure 5 that DDcGAN was generally, even on individual indexes of several fusion methods, is better than the last. However, through Table 3, it is found that the operation efficiency of DDcGAN is lower than fusion GAN. Considering that the balance between computational complexity and fusion effect is essential for the fusion of infrared and visible images, infrared and visible image features will become a challenging problem. Most image fusion indicators can only reflect the quality of the fused image to a certain extent. Hence, it is necessary to study more effective fusion methods and evaluation indicators to conduct a comprehensive quality evaluation. The table below explains the characteristics of Deep learning based image fusion methods.

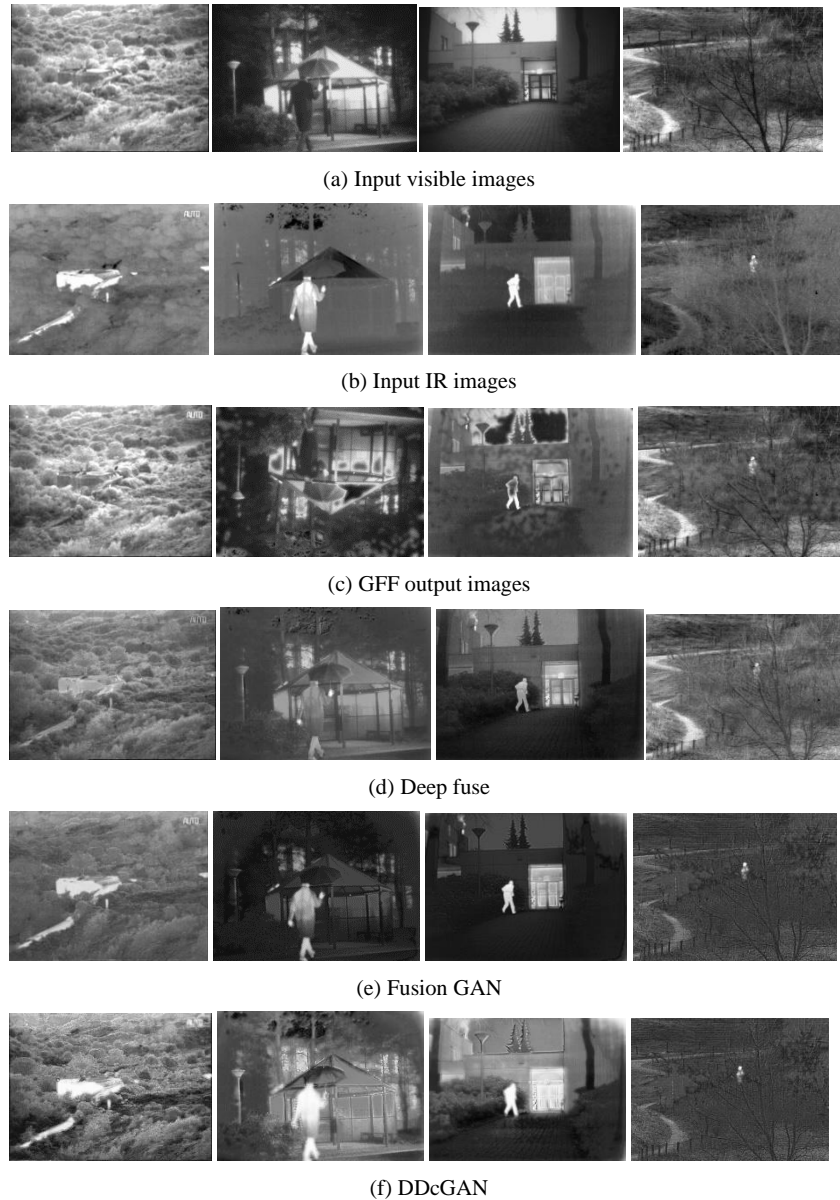


Figure 5. Infrared and visible image fusion results on typical image pairs from the TNOdatabase. From left to right: Bunker, Kaptein_1654, Kaptein_1123 and Sand path. From topto bottom: (a) the visible image, (b) the infrared image, (c) GFF, (d) deep fuse (e) fusion GAN (f) DDcGAN [4].

| DL Model | Major Characteristics |
|----------|---|
| CNN | <ul style="list-style-type: none"> - A fully connected CNN is learned to model the complex process of preventing motion artifacts in dynamic scenes. - Capable of autonomously extracting features and learning representative features from training data without human intervention. - High computational cost associated with the approach. - Utilizes a fully connected Convolutional Neural Network (CNN) to mitigate motion artifacts in dynamic scenes, extracting features autonomously and learning representative features from training data without manual intervention. However, this method imposes a significant computational burden. |
| CSR | <ul style="list-style-type: none"> - Operates by computing a sparse representation of the entire image. - Facilitates a shift-invariant representation. |

| | |
|-----|--|
| | <ul style="list-style-type: none"> - Adept at preserving details and less susceptible to misregistration issues. - Requires a significant amount of training data for efficacy. |
| SAE | <ul style="list-style-type: none"> - Employs an automatic feature extraction strategy combining autoencoder and dense network architectures. - Utilizes separate encoders for RGB and infrared imagery. - Followed by a decoder module to restore feature map resolution. - Incorporates a two-phase training mechanism with potential in scenarios with limited labeled data. - Training process may be prolonged without access to robust GPUs. |
| GAN | <ul style="list-style-type: none"> - Interaction between generator and discriminator - Employing trainable group convolution - Adversarial generation network with dual discriminators - Utilization of local binary pattern representation - Pre-fused image used as the ground truth label |

Table 2Mejor characteristics of DL based Image fusion methods

IV. EVALUATION METRICS

Quantitative experimental findings reveal that each fusion algorithm possesses its own set of advantages and disadvantages, showcasing distinct strengths in different aspects. The advancement of deep learning in infrared and visible image fusion has led to the emergence of superior technologies that consistently deliver improved fusion outcomes. Specifically, in terms of brightness, texture detail, and contrast, fusion GAN generally outperforms other methods. This observation is evident from the results presented in Table 2.

| Fusion Method | SD | SF |
|---------------|---------|--------|
| CVT | 28.700 | 6.9696 |
| Deep Fuse CNN | 44.0915 | 6.9603 |
| Fusion GAN | 48.5708 | 7.4707 |

Table 3 Comparison of Evaluation metrics for various fusion methods

Given the importance of striking a balance between computational complexity and fusion effectiveness in the fusion of infrared and visible images, it becomes evident that extracting meaningful features from these types of images poses a challenging and intricate problem. It is worth noting that most image fusion evaluation indicators only partially gauge the quality of the output fused image. Therefore, research necessitates the exploration of more effective fusion methods and evaluation indicators to facilitate wide-ranging quality assessments. Table 3 given below explains the challenges that need to face while using CNN and CSR methods. Observing that, GAN can be a good choice among other deep learning-based models.

| DL Model | Challenges |
|----------|--|
| CNN | <ul style="list-style-type: none"> - Exclusively applicable to multi-focus image fusion - Utilizes only the final layer for result computation - Valuable information captured by intermediate layers disregarded - Heightened network depth exacerbates information loss - Feature extraction leads to inevitable information loss - Discrepancies in resolution and spectral characteristics across application domains affect fusion outcome accuracy - Specific characteristics of source images require meticulous consideration within each dataset - Accommodating multitude of samples with intricate backgrounds imposes substantial computational burden during model training |

| | |
|-----|--|
| CSR | <ul style="list-style-type: none"> - Selection of image transform - Design of fusion strategy: activity level measurement and fusion rule - Selection of fusion manner: reconstruction-based or non-reconstruction based - Generation of training examples - Setting of learning parameters |
|-----|--|

Table 4 Challenges for Various Fusion Methods

V. CONCLUSION

The basic goal of image fusion is to generate a single fused image. Image fusion entails the amalgamation of two or more images derived from diverse sensors possessing distinct characteristics. This process aims to synthesize pertinent information from the source images into a comprehensive, informative singular image. The resultant fused image typically offers enhanced reliability and accuracy compared to any singular constituent image. Additionally, it exhibits a heightened capacity to discern features within the image. With the expanding adoption of deep learning techniques across various Image tasks, fusion methodologies commonly fall into two categories: deep learning-based and traditional approaches. Deep learning methods, necessitated by the absence of ground truth in image fusion, typically adopt an end-to-end fusion strategy facilitated by specialized loss functions and the integration of fusion strategies within the network's feature map set. Conversely, traditional methods tend to lack the robust feature representation and learning capabilities inherent to deep learning methodologies. To capitalize on the strengths of both paradigms, this study proposes a hybrid approach that amalgamates deep learning and traditional methodologies. Here, deep learning serves to extract texture features from images, which are subsequently leveraged in traditional image-processing tasks to enhance their efficacy.

A notable advancement within the realm of deep learning methodologies, particularly in domains such as computer vision, is the utilization of a technique known as data augmentation. This technique yields improved model performance by enhancing model proficiency and introducing a regularization effect, thereby reducing generalization error. Data augmentation operates by generating new synthetic examples within the problem domain on which the model is trained. Particularly for image data, these techniques often involve rudimentary manipulations such as cropping, flipping, zooming, and other basic transformations applied to existing images within the training dataset. Successful implementation of generative modeling offers a potential alternative that is more tailored to the specific characteristics of the domain for data augmentation purposes. In essence, data augmentation can be perceived as a simplified manifestation of generative modeling.

The integration of Deep Learning (DL) techniques into the fusion of visible and infrared (VI-IR) images has witnessed rapid advancement in recent years. Nonetheless, owing to the intricate nature of application situations and the concurrent chase of computational efficiency and fusion efficiency, various facets of VI-IR image fusion warrant further refinement. Moreover, there exist promising avenues for potential development in this domain. This review critically examines the latest advancements in DL-based image fusion technology, delineating key areas necessitating enhancement in the future. Specifically, the review scrutinizes DL-based fusion methodologies for infrared and visible imagery that have emerged in recent years. These methodologies are broadly categorized into four distinct groups: Convolutional Neural Network (CNN)-based approaches, Generative Adversarial Network (GAN)-based methods, Siamese network-based techniques, and Autoencoder architectures. The utilization of a deep convolutional network facilitates the acquisition of a fused image wherein all objects are accurately focused within a unified foreground and background context. Within this framework, convolutional and max-pooling layers inherent to Convolutional Neural Networks (CNNs) are leveraged to extract high-frequency details from the input source images. In future scope progress can be made in the field, particularly through methodologies such as Fully Learnable Group Convolution (FLGC)-fusion GAN, Dual-Discriminator Conditional Generative Adversarial Network (DDcGAN), and other contemporary approaches discussed herein, it becomes evident that Deep Learning (DL) has progressively evolved and matured in the realm of image fusion. However, within the domain of DL, which finds widespread application in the fusion of infrared and visible imagery, continued attention to fusion efficacy and computational considerations remains imperative.

REFERENCES

- [1] Harpreet Kaur¹ · Deepika Koundal² · Virender Kadyan³, “Image Fusion Techniques: A Survey”, *Archives of Computational Methods in Engineering* (2021) 28:4425–4447
- [2] Gang Xiao · Durga Prasad Baviriseti Gang Liu · Xingchen Zhang, “Image Fusion”, © Springer Nature Singapore Pte Ltd. and Shanghai Jiao Tong University Press 2020
- [3] FuchaoZha,Xiaoming Wu, Shuai Wang, “Object-oriented Vegetation Classification Method based on UAV and Satellite Image Fusion”, *International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI2019)*, 2019, 609-615.
- [4] Changqi Sun 1, Cong Zhang 1, and NaixueXiong 2, “Infrared and Visible Image Fusion Techniques Based on Deep Learning: A Review”, *Electronics* 2020, 9, 2162, www.mdpi.com/journal/electronics.
- [5] Deepak Murugan*, Akanksha Garg*, Tasneem Ahmed†, and Dharmendra Singh*, “Fusion of Drone and Satellite Data for Precision Agriculture Monitoring”, *11th International Conference on Industrial and Information Systems (ICIIS)*, IEEE, 2016, 910-914
- [6] Anish Vijana ,ParthDubeya, Shruti Jaina*, “Comparative Analysis of Various Image Fusion Techniques for Brain Magnetic Resonance Images”, *International Conference on Computational Intelligence and Data Science (ICCIDS 2019)*
- [7] T. Saikumar, S. Venkatesh , Md. Ameen Uddin, “Comparison Techniques of Image Fusion in Image Segmentation”, *Journal of Advances in Computer, Electrical & Electronics Engg.*, Volume 3 , Issue 1; Spl. Issue of IC3T 2014 @ISSN: 2248-9584 Page | 378
- [8] NawarAlseelawi, Hussein TuamaHazim, Haider TH. Salim ALRikabi, University of Misan, Maysan, Iraq, “A Novel Method of Multimodal Medical Image Fusion Based on Hybrid Approach of NSCT and DTCWT”, *iJOE – Vol. 18, No. 03, 2022*
- [9] Ian Goodfellow, NIPS 2016 Tutorial: Generative Adversarial Networks, [rXiv:1701.00160v4 \[cs. LG\]](https://arxiv.org/abs/1701.00160v4) 3 Apr 2017.
- [10] Lisa Corbata, Mohammad Nauvala, Julien Henrieta, Jean-Christophe Lapayre, “A Fusion Method Based on Deep Learning and Case-Based Reasoning which Improves the Resulting Medical Image Segmentations”, 20, 147, pp.113200 (9). [hal-02993849](https://hal.archives-ouvertes.fr/hal-02993849)
- [11] Dwarikanath Mahapatra, Bhavna Antony, Suman Sedai, Rahil Garnavi,” *DEFORMABLE MEDICAL IMAGE REGISTRATION USING GENERATIVE ADVERSARIAL NETWORKS*”, 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), April 4-7, 2018, Washington, D.C., USA
- [12] Li Dong¹, Yongzheng Lin², Yishen Pang, “Medical Image Registration via Similarity Measure based on Convolutional Neural Network”, *2020 5th International Conference on Cloud Computing and Artificial Intelligence: Technologies and Applications (CloudTech) | 978-1-7281-6175-4/20/\$31.00 ©2020 IEEE | DOI: 10.1109/CloudTech49835.2020.9365908*
- [13] Yogesh M. Chukewad, Johannes James ,AvinashSingh,and Sawyer Fuller, “RoboFly: An Insect-Sized Robot With Simplified Fabrication That Is Capable of Flight, Ground, and Water Surface Locomotion”, *IEEE TRANSACTIONS ON ROBOTICS* 1552-3098 © 2021 IEEE.
- [14] Yong Yang, Jiayang Liu, Shuying Huang, Weiguo Wan, Wenying Wen, and Juwei Guan, “Infrared and Visible Image Fusion via Texture Conditional Generative Adversarial Network”, *IEEE Transactions on Circuits and Systems for Video Technology* , Volume: 31, Issue: 12, Dec. 2021 , 4771 – 4783
- [15] Kunfeng Wang, Member, IEEE, Chao Gou, Yanjie Duan, Yilun Lin, Xinhui Zheng, and Fei-Yue Wang, Fellow, IEEE, “Generative Adversarial Networks: Introduction and Outlook”, *IEEE/CAA JOURNAL OF AUTOMATICA SINICA*, VOL. 4, NO. 4, OCTOBER 2017.
- [16] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, SherjilOzairy, Aaron Courville, YoshuaBengioz, “Generative Adversarial Nets”, 10 Jun 2014.
- [17] Mamta Sharma, “A Review : Image Fusion Techniques and Applications”, *(IJCSIT) International Journal of Computer Science and Information Technologies*, Vol. 7 (3) , 2016, 1082-108
- [18] AYUSH DOGRA, BHAWNA GOYAL, AND SUNIL AGRAWAL, “From Multi-Scale Decomposition to Non-Multi-Scale Decomposition Methods: A Comprehensive Survey of Image Fusion Techniques and Its Applications”, *IEEE access*, VOLUME 5, 2017
- [19] C. Pohl & J. L. Van Genderen, “Review article Multisensor image fusion in remotesensing: Concepts, methods and applications”, *International Journal of Remote Sensing*,1998, vol. 19, no. 5, 823± 854
- [20] Shrouk A. Elmasry, Wael A. Awad and Sami A. Abd El-hafeez, “Review of Different Image FusionTechniques: Comparative Study”, Springer Nature Singapore Pte Ltd. 2020
- [21] Omar and 4 others,” A novel perceptual two layer image fusion using deep learning for imbalanced COVID-19 dataset”, Elzeki et al. (2021), *PeerJComput. Sci.*, DOI 10.7717/peerj-cs.364
- [22] Suman Deb and two more, “Application of Image Fusion for Enhancing the Quality of an Image”, Natarajan Meghanathan, et al. (Eds): *SIPM, FCST, ITCA, WSE, ACSIT, CS & IT* 06, pp. 215–221, 2012.
- [23] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [24] Prabhakar, K.R.; Srikar, V.S.; Babu, R.V. DeepFuse: A Deep Unsupervised Approach for Exposure Fusion with Extreme Exposure Image Pairs. *ICCV* 2017, 1, 3.
- [25] Ma, Y.; Chen, J.; Chen, C.; Fan, F.; Ma, J. Infrared and visible image fusion using total variation model.
- [26] *Neurocomputing* 2016, 202, 12–19. [CrossRef]
- [27] Li, H.; Wu, X.J.; Durrani, T.S. Infrared and visible image fusion with ResNet and zero-phase component analysis. *Infrared Phys. Technol.* 2019, 102, 103039. [CrossRef]
- [28] Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Inf. Fusion* 2019, 48, 11–26. [CrossRef]

- [29] Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* 2017, 36, 191–207. [CrossRef]
- [30] Zhang, X.; Ye, P.; Qiao, D.; Zhao, J.; Peng, S.; Xiao, G. Object fusion tracking based on visible and infrared
- [31] images using fully convolutional siamese networks. In *Proceedings of the 2019 22th International Conference on Information Fusion (FUSION)*, Ottawa, ON, Canada, 2–5 July 2019; pp. 1–8.
- [32] Zhang, X.; Ye, P.; Peng, S.; Liu, J.; Gong, K.; Xiao, G. SiamFT: An RGB-infrared fusion tracking method via fully convolutional siamese networks. *IEEE Access* 2019, 7, 122122–122133. [CrossRef]
- [33] Yuan, C.; Sun, C.Q.; Tang, X.Y.; Liu, R.F. FLGC-Fusion GAN: An Enhanced Fusion GAN Model by Importing Fully Learnable Group Convolution. *Math. Probl. Eng.* 2020. [CrossRef]
- [34] Xu, J.; Shi, X.; Qin, S.; Lu, K.; Wang, H.; Ma, J. LBP-BEGAN: A generative adversarial network architecture for infrared and visible image fusion. *Infrared Phys. Technol.* 2020, 104, 103144. [CrossRef]
- [35] Li, Q.; Lu, L.; Li, Z.; Wu, W.; Liu, Z.; Jeon, G.; Yang, X. Coupled GAN with relativistic discriminators for infrared and visible images fusion. *IEEE Sens. J.* 2019, 1. [CrossRef]
- [36] Gayatri Phade, Omkar Vaidya, Rahul Awathankar, Priti Shahane, "Canopy Greenfield Detection System and Pesticides Spraying using Drone", *International Journal of Engineering Trends and Technology* Volume 71 Issue 10, 215-222, October 2023
- [37] Dr. Omkar Vaidya, et. al., "E-Governance System for Traffic Infringement Using Automatic Vehicle Number Plate Recognition", *Proceedings of International Conference on Trends in Electronics and Communication (IC-TELCON-2020)*, ISBN (13): 978-93-90385-79-9, McGraw Hill Education (India) Private Limited, Feb 20
- [38] Dr. Omkar Vaidya, et., "Development of Human Machine Interface using Smart Mirror and Face Recognition Algorithm", *International Journal of Advanced Research Engineering and Technology*, Volume 11, Issue 10, 555-563, October 2020