

# Infrared and Visible Image Fusion

Subjects: [Others](#)

Contributor: Changqi Sun

Infrared and visible image fusion technologies make full use of different image features obtained by different sensors, retain complementary information of the source images during the fusion process, and use redundant information to improve the credibility of the fusion image.

Infrared and Visible Image Fusion

## 1. Introduction

Under normal conditions, objects will radiate electromagnetic waves of different frequencies, which is called thermal radiation. It is difficult for people to see thermal radiation information with the naked eye<sup>[1]</sup>. It is necessary to use different sensors<sup>[2][3][4][5][6][7][8][9][10]</sup> to process the infrared image to obtain its thermal radiation information, which has good target detection ability<sup>[11]</sup>. Infrared images can avoid the influence of the external environment, such as sunlight, smoke, and other conditions<sup>[1][12]</sup>. However, infrared images have low contrast, complex background, and poor feature performance. Visible images are consistent with the human eye's visual characteristics and contain many edge features and detailed information<sup>[13]</sup>. The use of visible light sensors to obtain image spectral information is richer, scene details and textures are clear, and spatial resolution is high. However, due to the external environment's influence, such as night environment, camouflage, smoke hidden objects, background clutter, etc., the target may not be easily observed in the visible image. Therefore, infrared and visible light fusion technology combines the two's advantages and retains more infrared and visible feature information in the fusion result<sup>[14]</sup>. Due to the universality and complementarity of infrared images and visible images, the fusion technology of infrared and visible images has been applied to more fields and plays an increasingly important role in computer vision. Nowadays, the fusion method of infrared and visible images have been widely used in target detection<sup>[15]</sup>, target recognition<sup>[16]</sup>, image enhancement<sup>[17]</sup>, remote sensing detection<sup>[18]</sup>, agricultural automation<sup>[19][20]</sup>, medical imaging<sup>[21]</sup>, industrial applications<sup>[22][23][24]</sup>.

According to different image fusion processing domains, image fusion can be roughly divided into two categories: the spatial and transform domains. The focus of the fusion method is to extract relevant information from the source image and merge it<sup>[25]</sup>. Current fusion algorithms can be divided into seven categories, namely, multi-scale transform<sup>[26]</sup>, sparse representation<sup>[27]</sup>, neural network<sup>[28]</sup>, subspace<sup>[29]</sup>, saliency<sup>[30]</sup>, hybrid models<sup>[31]</sup>, and deep learning<sup>[32]</sup>. Each type of fusion method involves three key challenges, i.e., image transform, activity-level measurement, and fusion rule designing<sup>[33]</sup>. Image transformation includes different multiscale decomposition, various sparse representation methods, non-downsampling methods, and a combination of different transformations. The goal of activity level measurement is to obtain quantitative information to assign weights from

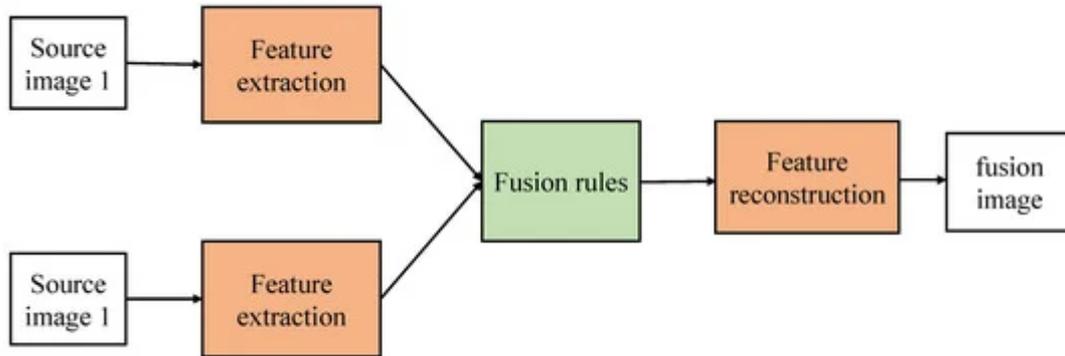
different sources<sup>[12]</sup>. The fusion rules include the big rule and the weighted average rule, the essence of which plays the role of weight distribution<sup>[32]</sup>. With the rapid development of fusion algorithms in theory and application, selecting an appropriate feature extraction strategy is the key to image fusion. It is still challenging to design a suitable convolutional neural network and adjust the parameters based on deep learning image fusion. Especially in recent years, after generating a confrontation network for image fusion, although it brings a clearer fusion effect, it also needs to consider the inevitable gradient disappearance and gradient explosion of the generation confrontation training.

In the field of image fusion, a variety of different infrared and visible image fusion methods have been proposed in recent years. However, there are still some challenges in different infrared and visible image fusion applications. The commonly seen fusion method is to select the same salient features of the source image and integrate them into the fusion image to contain more detailed information. However, the infrared heat radiation information is mainly characterized by pixel intensity, while edges and gradients characterize the visible image's texture detail information. According to the different imaging characteristics of the source image, the selection of traditional manually designed fusion rules to represent the fused image, in the same way, will lead to the lack of diversity of extracted features, which may bring artifacts to the fused image. Moreover, for multi-source image fusion, manual fusion rules will make the method more and more complex. In view of the above problems, the image fusion method based on deep learning can assign weights to the model through an adaptive mechanism<sup>[34]</sup>. Compared with the design rules of traditional methods, this method greatly reduces the calculation cost, which is crucial in many fusion rules. Therefore, this research aims to conduct a detailed review of the existing deep learning-based infrared and visible image fusion algorithms and discuss their future development trends and challenges. Second, this article also introduces the theoretical knowledge of infrared and visible image fusion and the corresponding fusion evaluation index. This survey also makes a qualitative and quantitative comparison of some related articles' experiments to provide a reliable basis for this research. Finally, we summarized the fusion methods in recent years and analyzed future work trends.

## 2. Fusion Methods of Infrared and Visible Images Based on Deep Learning

In this section, we comprehensively review the infrared and visible image fusion methods based on deep learning. Increasing new methods of using deep learning for infrared and visible image fusion have been produced in recent years. These state-of-the-art methods are widely used in many applications, like image preprocessing, target recognition, and image classification. The traditional fusion framework can be roughly summarized in [Figure 2](#). The two essential factors of these algorithms are feature extraction and feature fusion. Their main theoretical methods can be divided into multiscale transformation, sparse representation, subspace analysis, and hybrid methods. However, these artificially designed extraction methods make the image fusion problem more complicated due to their limitations. In order to overcome the limitations of traditional fusion methods, deep learning methods are introduced for feature extraction. In recent years, with the development of deep learning, several fusion methods based on convolutional neural network (CNN), generative adversarial networks (GAN), Siamese network, and

autoencoder have appeared in the field of image fusion. The main fusion methods involved in this section are listed in [Table 1](#) by category. Image fusion results based on deep learning have good performance, but many methods also have apparent challenges. Therefore, we will introduce the details of each method in detail.



**Figure 2.** Traditional image fusion framework.

**Table 1.** The overview of some deep learning (DL)-based fusion methods.

Families of Fusion Methods	Ref.	Innovation
CNN method of DL	[35]	VGG-19; L1 norm; weighted average strategy; maximum selection strategy
	[36]	Dense net
	[37]	Minimize the total change
	[38]	ZCA-zero-phase component analysis; L1-norm; weighted average strategy
	[39]	Elastic weight consolidation
	[40]	Perceptual loss; use two convolutional layers to extract image features; weight sharing strategy
	[41]	Adaptive information preservation strategy
	[42]	MLF-CNN; weighting summation strategy
	[43]	Mixed loss function (M-SSIM loss; TV loss); adaptive VIF-Net
Siamese network of DL	[44]	Fusion strategy of local similarity; weighted average
	[45]	Pixel-level image fusion; feature tracking
	[46]	Dual Siamese network; weight sharing strategy
	[47]	Saliency map; three-level wavelet transform

Families of Fusion Methods	Ref.	Innovation
GAN of DL	[33]	The confrontation between generator and discriminator
	[48]	Learnable group convolution
	[49]	Adversarial generation network with dual discriminators
	[50]	Detail loss; target edge loss
	[51]	Local binary pattern
	[52]	Pre-fused image as the label
Autoencoder of DL	[49]	Automatic coding feature extraction strategy of generator
	[53]	Combination of autoencoder and dense network
	[54]	RGB encoder; infrared encoder; decoder used to restore the resolution of the feature map

## 2.1. CNN-Based Fusion Methods

In computer vision, convolutional layers play an important role in feature extraction and usually provide more information than traditional manual feature extraction methods<sup>[55][56]</sup>. The critical problem of image fusion is how to extract salient features from the source images and combine them to generate the fused image. However, CNN has three main challenges when applied to image fusion. First, training a good network requires much labeled data. However, the image fusion architecture based on the convolutional neural network is too simple, and the convolutional calculation layer in the network framework is less, and the features extracted from the image are insufficient, resulting in poor fusion performance. Second, the artificially designed image fusion rules are challenging to realize the end-to-end model network, and some errors will be mixed in the feature reconstruction process, which will affect the feature reconstruction of the image. Finally, the efficient information of the last layer is ignored in the traditional convolutional neural network algorithm, so that the model features cannot be fully retained. With the deepening of the network, the feature loss will become severe, resulting in a worsening of the final fusion effect.

In<sup>[57]</sup>, Liu et al. proposed a fusion method based on convolutional sparse representation (CSR). In their method, the authors use CSR to extract multilayer features and then use them to generate fusion images. In<sup>[58]</sup>, they also proposed a fusion method based on a convolutional neural network (CNN). They use image patches containing different feature inputs to train the network and obtain a decision graph. Finally, the fusion image is obtained by using the decision graph and the source image. Li et al.<sup>[35]</sup> proposed a simple and effective infrared and visible image fusion method based on a deep learning framework. The article divides the source image information into two parts, the former contains low-frequency information, and the latter contains texture information. The model is based on the multilayer fusion strategy of the VGG-19 network<sup>[59]</sup> through which the deep features of the detailed

content can be obtained. In other multiple exposure fusion (MEF) algorithms, they rely on artificially searched features to fuse images. When the input conditions change, the parameters will follow the change, so the robustness of the algorithm cannot be guaranteed, and processing multiple exposure images will consume a lot. The learning ability of CNN is affected mainly by some loss functions. Prabhakar et al.<sup>[36]</sup>, the proposed method does not need parameter adjustment when the input changes. The fusion network consists of three parts: the encoder, the fusion layer, and the decoder. To combine encoder networks employing encoders. From the perspective of the CNN method, by optimizing the parameters of the loss function learning model, the results can be predicted as accurately as possible. In<sup>[37]</sup>, Ma et al. proposed an infrared and visible image based on the minimization of the total variation (TV) by limiting the fusion image to have similar pixel intensity to the infrared image and similar gradient to the visible image. In<sup>[38]</sup>, Li et al. proposed a fusion framework based on deep features and zero-phase component analysis. First, the residual network is used to extract the depth features of the source image, and then the ZCA-zero-phase component analysis<sup>[60]</sup> and L1-norm are used for normalization to obtain the initial weight map. Finally, the weighted average strategy is used to reconstruct the fused image.

Xu et al.<sup>[39]</sup>, a new unsupervised and unified densely connected network is proposed. The densely connected network (DenseNet)<sup>[61]</sup> is trained to generate a fused image adjusted on the source image in the proposed method. In addition, we obtain a single model applicable to multiple fusion tasks by applying elastic weight consolidation to avoid forgetting what has been learned from previous tasks when training multiple tasks sequentially, rather than train individual models for every fusion task or jointly train tasks roughly. The weight of the two source images is obtained through the weight block, and different feature information is retained. The model generates high-quality fusion results in processing multi-exposure and multi-focus image fusion. In<sup>[40]</sup>, Zhang et al. proposed an end-to-end model divided into three modules: feature extraction module, feature fusion module, and feature reconstruction module. Two convolutional layers are used to extract image features. Appropriate fusion rules are adopted for the convolutional features of multiple input images. Finally, the fused features are reconstructed by two convolutional layers to form a fused image. In<sup>[41]</sup>, Xu et al. believe that an unsupervised end-to-end fusion network can solve different fusion problems, including multimode, multi-exposure, and multi-focus. The model can automatically estimate the importance of the corresponding source image features and provide adaptive information preservation because the model has an adaptive ability to retain the similarity between the fusion result and the source image. It dramatically reduces the difficulty of applying deep learning to image fusion-the universality of the model and the adaptive ability of training weights. Solve the catastrophic forgetting problem and computational complexity.

In <sup>[42]</sup>, Chen et al. used deep learning methods to fuse visible information and thermal radiation information in multispectral images. This method uses the multilayer fusion (MLF) area network in the image fusion stage. In this way, pedestrians can be detected at different ratios under unfavorable lighting (such as shadows, overexposure, or night) conditions. To be able to handle targets of various sizes, prevent the omission of some obscure pedestrian information. In the region extraction stage, MLF-CNN designed a multiscale region proposal network (RPN) <sup>[62]</sup> to fuse infrared and visible light information and use summation fusion to fuse two convolutional layers. In<sup>[43]</sup>, to solve the lack of label dataset, Hou et al. used a mixed loss function. The thermal infrared image and the visible image were adaptively merged by redesigning the loss function, and noise interference was suppressed. This method can

retain salient features and texture details with no apparent artifacts and have high computational efficiency. We make an overview list of some of the image fusion based on CNN in [Table 2](#).

**Table 2.** The overview of some a convolutional neural network (CNN)-based fusion methods.

Ref.	Limitation
[57]	It is only suitable for multi-focus image fusion, and only the last layer is used to calculate the result. Much useful information obtained by the middle layer will be lost. When the network depth increases, the information loss will become more serious.
[36]	Feature extraction will still lose some information.
[37]	In different application fields, the accuracy of the fusion result cannot be guaranteed due to the large difference in resolution and spectrum.
[40]	The specific performance of different source images needs to be considered in a specific dataset.
[42]	A large number of samples with a complex background bring a large amount of calculation to model training.

## 2.2. Siamese Networks-Based Fusion Methods

Part of the difficulty of image fusion is that infrared images and visible images have different imaging methods. In order to make the fusion image retain the relatively complete information of the two source images at the same time, a pyramid framework is used to extract feature information from the infrared image and the visible image, respectively.

Liu et al. [58] recently proposed a Siamese convolutional network, especially image fusion. The network input is two source images, while the output is a weight map for the final decision. Many high-quality natural images are applied to generate the training dataset via Gaussian blurring and random sampling. The main characteristic of this approach is activity level measurement, and weight assignments are simultaneously achieved with the network. In particular, the convolutional layers and fully-connected layers could be viewed as the activity level measurement and weight assignment parts in image fusion, respectively. Again in [44], Liu et al. proposed a convolutional neural network-based infrared and visible image fusion method. This method uses the Siamese network to obtain the network weight map. The weight map combines the pixel activity information of the two source images. The model has mainly divided into four steps: the infrared image and the visible image are passed into the convolutional neural network to generate weights; the Gaussian pyramid is used to decompose the weight of the source image, and the two source images are decomposed by the Laplacian pyramid respectively. The information obtained by the decomposition of each pyramid is fused with coefficients in a weighted average manner. [Figure 3](#) clearly explains the working principle of the Siamese network in the fusion process. In [45], Zhang et al. believe that CNN has a powerful feature representation ability and can produce good tracking performance. Still, the training and updating of the CNN model are time-consuming. Therefore, in this paper, the Siamese network is used for pixel-level fusion to reduce time consumption. First, the infrared and visible images are fused and then put into the

Siamese network for feature tracking. In [46], Zhang et al. used a fully convolutional Siamese network fusion tracking method. SlamFT uses a Siamese network, a visible light network, an infrared network. They are used to process visible and infrared images, respectively. The backbone uses the SiamFC network, the visible light part of the network weight sharing, and the infrared part of the network weight sharing. The operating speed is about [35][36][37][38][58] FPS so that it can meet real-time requirements. In [47], Piao et al. designed an adaptive learning model based on the Siamese network, which automatically generates the corresponding weight map through the saliency of each pixel in the source image to reduce the number of traditional fusion rules. The parameter redundancy problem. This paper uses a three-level wavelet transform to decompose the source image into a low-frequency weight map and a high-frequency weight map. The scaled weight map is used to reconstruct the wavelet image to obtain the corresponding fused image. This result is more consistent with the human visual perception system. There are fewer undesirable artifacts. We make an overview list of some image fusion based on the Siamese network in Table 3.

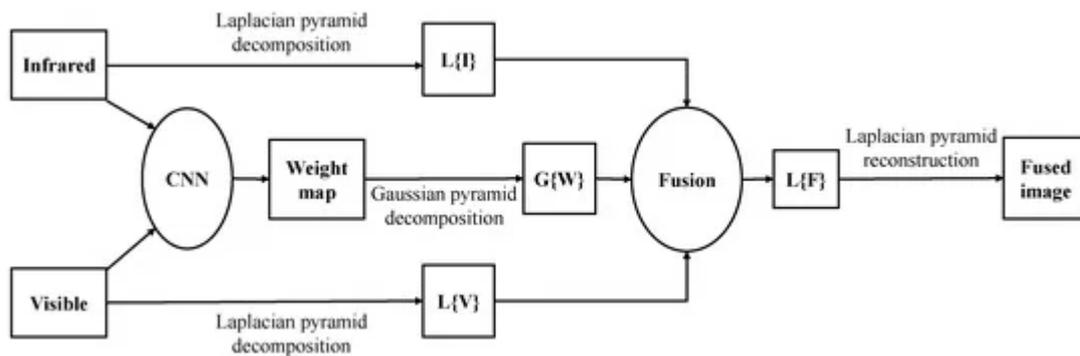


Figure 3. Siamese network-based infrared and visible image fusion scheme (credit to [44]).

Table 3. The overview of Siamese network-based fusion methods.

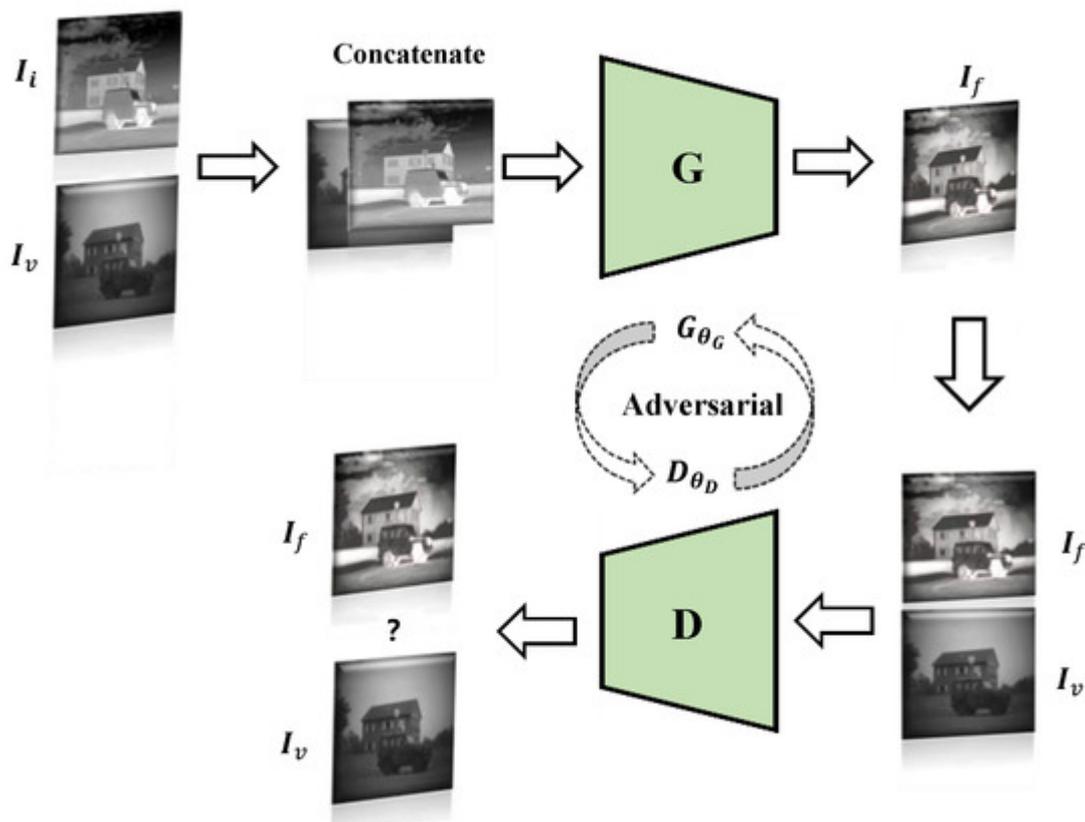
Ref.	Limitation
[45]	The starting point of the article is target tracking. As far as the fusion effect is concerned, the fusion result is slightly blurred.
[44]	It cannot be effectively combined with conventional fusion technology and is not suitable for complex data sets.
[46]	The thermal infrared network training uses visible images, and you can consider using thermal infrared images for better results.
[47]	The CPU is used to train the model, so the computational efficiency of the model is not very prominent. It takes an average of 19 s to process a pair of source images.

## 2.3. GAN-Based Fusion Methods

The existing deep learning-based image fusion technology usually relies on the CNN model, but in this case, the ground truth needs to be provided for the model. However, in the fusion of infrared and visible images, it is

unrealistic to define fusion image standards. Therefore, without considering the ground truth, a deep model is learned to determine the degree of blurring of each patch in the source image, and then the weight is calculated. Map accordingly to generate the final fusion image<sup>[44]</sup>. Using a generative countermeasure network to fuse infrared and visible images can be free from the above problems.

In <sup>[33]</sup>, Ma et al. proposed an image fusion method based on a generative confrontation network, where the generator is mainly for the fusion of infrared images and visible images, and the purpose of the discriminator is to make the fused image have more details in the visible image, which makes the fused image. The infrared heat radiation information and visible texture information can be kept in the fusion image simultaneously. **Figure 4** shows the image fusion framework based on GAN. For fusion GAN, the source image's vital information cannot be retained at the same time during the image fusion process, and too much calculation space is occupied during the convolution process. In<sup>[48]</sup>, learning group convolution is used to improve the efficiency of the model and save computing resources. In this way, a better tradeoff can be made between model accuracy and speed. Moreover, the remaining dense blocks are used as the fundamental network construction unit. The inactive perceptual characteristics are used as the input content loss characteristics, which achieves deep network supervision.



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

In<sup>[49]</sup>, Ma et al. make the fusion image similar to the infrared image by constrained sampling to avoid blurring radiation information or loss of visible texture details. The dual discriminator does not need ground truth fusion images for pre-training, which can fuse images of different resolutions without causing thermal radiation information

blur or visible texture detail loss. Considering the two challenges of CNN, relying only on adversarial training will result in the loss of detailed information. Therefore, a minimax game is established between the generator and the discriminator in<sup>[50]</sup>. The loss of the model becomes the loss of detail, the loss of the target edge, and confrontation loss. In<sup>[51]</sup>, Xu et al., based on local binary pattern (LBP)<sup>[63]</sup>, intuitively reflected the edge information of the image by comparing the values between the central pixel and the surrounding eight pixels to generate a fusion image with richer boundary information. The discriminator encodes and decodes the fused image and each source image, respectively, and measures the difference between the distributions after decoding. In<sup>[52]</sup>, Li et al. used the pre-fused image as the label strategy so that the generator takes the pre-fused image as the benchmark in the generation process so that the image fused by the generator can effectively and permanently retain the rich texture in the visible image and the thermal radiation information in the infrared image. We make an overview list of some of the image fusion based on GAN in [Table 4](#).

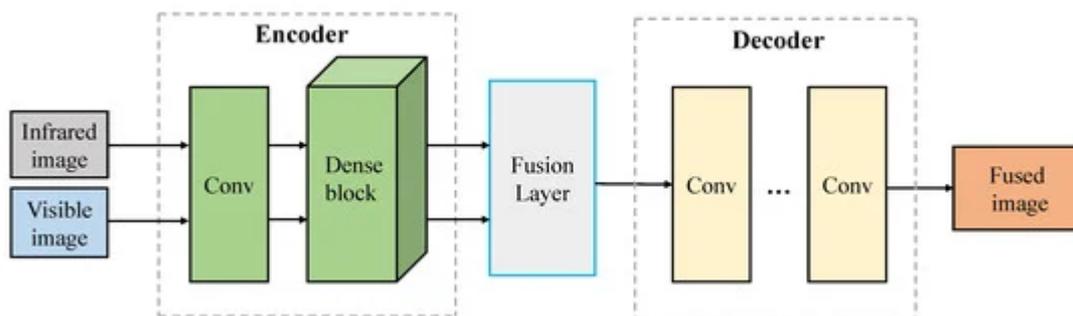
**Table 4.** The overview of some GAN-based fusion methods.

Ref.	Limitation
<sup>[33]</sup>	Reduce the prominence of infrared thermal targets.
<sup>[50]</sup>	The pixel intensity of some fusion image areas is changed, and the overall brightness is reduced.
<sup>[48]</sup>	Some edges of the fused image are a bit blurry.
<sup>[51]</sup>	Unique fusion results have bright artifacts.
<sup>[52]</sup>	In the early stage of model training, it takes some time to label the pre-fused images.

## 2.4. Autoencoder-Based Fusion Methods

In the paper<sup>[36]</sup>, Prabhakaret et al. studied the fusion problem based on CNN. They proposed a simple CNN-based architecture, including two encoding network layers and three decoding network layers. Although this method has good performance, there are still two main shortcomings: (1) The network architecture is too simple, and it may not be able to extract the salient features of the source image correctly; (2) These methods only use the last layer of the encoding network to calculate; as a result, the useful information obtained by the middle layer will be lost. This phenomenon will become sparser when the network is deeper. In the traditional CNN network, as the depth increases, the fusion ability of the model is degraded<sup>[30]</sup>. For this problem, Heet et al.<sup>[64]</sup> introduced a deep residual learning framework to improve the layers' information flow further. Huang et al.<sup>[61]</sup> proposed new architecture with dense blocks in which each layer can be directly connected to any subsequent layer. The main advantages of the dense block architecture: (1) the architecture can retain as much information as possible; (2) the model can improve the information flow and gradient through the network, and the network is easy to train; (3) this dense connection method has a regularization effect, which can reduce overfitting caused by too many parameters<sup>[61]</sup>.

In<sup>[53]</sup>, Li et al. combine the encoding network with the convolutional layer, fusion layer, and dense block, and the output of each layer is connected. The figure shows the working principle of the Autoencoder model in the fusion image. The model first obtains the feature map through CNN and dense block and then fuses the feature through the fusion strategy. After the fusion layer, the feature map is integrated into a feature map containing the significant features of the source image. Finally, the fused image is reconstructed by a decoder. The fusion mechanism of the autoencoder is shown in [Figure 5](#). In<sup>[49]</sup>, Ma et al. considering the existing methods to solve the difference between output and target by designing loss function. These indicators will introduce new problems. It is necessary to design an adaptive loss function to avoid the ambiguity of the results. Most human-designed fusion rules lead to the extraction of the same features for different types of source images, making this method unsuitable for multi-source image fusion. In this paper, a double discriminator is used to pre-train the fused images. An Autoencoder is used to fuse the images with different resolutions to retain the maximum or approximately the maximum amount of information in the source images. In<sup>[54]</sup>, Sun et al. used the RGB-thermal fusion network (RTFNet). RTFNet consists of three modules: RGB encoder and infrared encoder for extracting features from RGB images and Thermal images, respectively, and decoder to restore the resolution of feature images. Where the encoder and decoder are designed regionally symmetric, RTFNet is used for feature extraction, where the new encoder can restore the resolution of the approximate feature map. As this method is mainly used for scene segmentation, the edge of scene segmentation is not sharp.



**Figure 5.** Autoencoder based infrared and visible image fusion framework (credit to<sup>[53]</sup>).

## References

1. Baohua, Z.; Xiaoqi, L.; Haiquan, P.; Ying, Z. A fusion algorithm for infrared and visible images based on saliency analysis and non-subsampled shearlet transform. *Infrared Phys. Technol.* 2015, 73, 286–297.
2. Guo, W.; Xiong, N.; Chao, H.C.; Hussain, S.; Chen, G. Design and analysis of self-adapted task scheduling strategies in wireless sensor networks. *Sensors* 2011, 11, 6533–6554.
3. Zeng, Y.; Sreenan, C.J.; Xiong, N.; Yang, L. T.; Park, J.H. Connectivity and coverage maintenance in wireless sensor networks. *J. Supercomput.* 2010, 52, 23–46.

4. Shu, L.; Zhang, Y.; Yu, Z.; Yang, L.T.; Hauswirth, M.; Xiong, N. Context-aware cross-layer optimized video streaming in wireless multimedia sensor networks. *J. Supercomput.* 2010, 54, 94–121.
5. Guo, W.; Xiong, N.; Vasilakos, A.V.; Chen, G.; Cheng, H. Multi-source temporal data aggregation in wireless sensor networks. *Wirel. Pers. Commun.* 2011, 56, 359–370.
6. Lin, C.; He, Y.X.; Xiong, N. An energy-efficient dynamic power management in wireless sensor networks. In *Proceedings of the 2006 Fifth International Symposium on Parallel and distributed computing Timisoara, Romania, 6–9 July 2006*; pp. 148–154.
7. Huang, K.; Zhang, Q.; Zhou, C.; Xiong, N.; Qin, Y. An efficient intrusion detection approach for visual sensor networks based on traffic pattern learning. *IEEE Trans. Syst. Man Cybern. Syst.* 2017, 47, 2704–2713.
8. Wu, W.; Xiong, N.; Wu, C. Improved clustering algorithm based on energy consumption in wireless sensor networks. *IET Netw.* 2017, 6, 47–53.
9. Zhou, Y.; Zhang, D.; Xiong, N. Post-cloud computing paradigms: A survey and comparison. *Tsinghua Sci. Technol.* 2017, 22, 714–732.
10. He, R.; Xiong, N.; Yang, L.T.; Park, J.H. Using multi-modal semantic association rules to fuse keywords and visual features automatically for web image retrieval. *Inf. Fusion* 2011, 12, 223–230.
11. Jin, H.; Wang, Y. A fusion method for visible and infrared images based on contrast pyramid with teaching learning based optimization. *Infrared Phys. Technol.* 2014, 64, 134–142.
12. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *Inf. Fusion* 2017, 33, 100–112.
13. Cui, G.; Feng, H.; Xu, Z.; Li, Q.; Chen, Y. Detail preserved fusion of visible and infrared images using regional saliency extraction and multiscale image decomposition. *Opt. Commun.* 2015, 341, 199–209.
14. Xu, L.; Cui, G.; Zheng, C.; Zhao, J. Visible infrared image fusion method based on multiscale decomposition and saliency region extraction. *Laser Optoelectron. Prog.* 2017, 54, 111–120.
15. Han, J.; Bhanu, B. Fusion of color and infrared video for moving human detection. *Pattern Recognit.* 2007, 40, 1771–1784.
16. Singh, R.; Vatsa, M.; Noore, A. Integrated multilevel image fusion and match score fusion of visible and infrared face images for robust face recognition. *Pattern Recognit.* 2008, 41, 880–893.
17. Reinhard, E.; Adhikhmin, M.; Gooch, B.; Shirley, P. Color transfer between images. *IEEE Comput* 2001, 21, 34–41.

18. Simone, G.; Farina, A.; Morabito, F.C.; Serpico, S.B.; Bruzzone, L. Image fusion techniques for remote sensing applications. *Inf. Fusion* 2002, 3, 3–15.
19. Mendoza, F.; Lu, R.; Cen, H. Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content. *Postharvest Biol. Technol.* 2012, 73, 89–98.
20. Bulanon, D.M.; Burks, T.F.; Alchanatis, V. Image fusion of visible and thermal images for fruit detection. *Biosyst. Eng.* 2009, 103, 12–22.
21. Hanna, B.V.; Gorbach, A.M.; Gage, F.A.; Pinto, P.A.; Silva, J.S.; Gilfillan, L.G.; Elster, E.A. Intraoperative assessment of critical biliary structures with visible range/infrared image fusion. *J. Am. Coll. Surg.* 2008, 206, 1227–1231.
22. Sanchez, V.; Prince, G.; Clarkson, J.P.; Rajpoot, N.M. Registration of thermal and visible light images of diseased plants using silhouette extraction in the wavelet domain. *Pattern Recognit.* 2015, 48, 2119–2128.
23. Apatean, A.; Rogozan, A.; Bensrhair, A. Visible-infrared fusion schemes for road obstacle classification. *Transp. Res. Part* 2013, 35, 180–192.
24. Eisler, K.; Homma, C.; Goldammer, M.; Rothenfusser, M.; Arnold, W. Fusion of visual and infrared thermography images for advanced assessment in non-destructive testing. *Rev. Sci. Instrum.* 2013, 84, 064902.
25. Wang, J.; Peng, J.; Feng, X.; He, G.; Fan, J. Fusion method for infrared and visible images by using non-negative sparse representation. *Infrared Phys. Technol.* 2014, 67, 477–489.
26. Li, S.; Kang, X.; Hu, J. Image fusion with guided filtering. *IEEE Trans. Image Process.* 2013, 22, 2864–2875.
27. Yang, B.; Li, S. Visual attention guided image fusion with sparse representation. *Opt. Int. J. Light Electron. Opt.* 2014, 125, 4881–4888.
28. Kong, W.; Zhang, L.; Lei, Y. Novel fusion method for visible light and infrared images based on NSST–SF–PCNN. *Infrared Phys. Technol.* 2014, 65, 103–112.
29. Bavirisetti, D.P.; Xiao, G.; Liu, G. Multi-sensor image fusion based on fourth order partial differential equations. In *Proceedings of the 2017 20th International Conference on Information Fusion (Fusion)*, Xi'an, China, 10–13 July 2017; pp, 1–9.
30. Bavirisetti, D.P.; Dhuli, R. Two-scale image fusion of visible and infrared images using saliency detection. *Infrared Phys. Technol.* 2016, 76, 52–64.
31. Liu, Y.; Liu, S.; Wang, Z. A general framework for image fusion based on multiscale transform and sparse representation. *Inf. Fusion* 2015, 24, 147–164.

32. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Inf. Fusion* 2018, 42, 158–173.
33. 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.
34. Shi, W.; Zhu, C.; Tian, Y.; Nichol, J. Wavelet-based image fusion and quality assessment. *Int. J. Appl. Earth Obs. Geoinf.* 2005, 6, 241–251.
35. Li, H.; Wu, X.J.; Kittler, J. Infrared and visible image fusion using a deep learning framework. In *Proceedings of the International Conference on Pattern Recognition 2018 IEEE, Beijing, China, 20–24 August 2018*; pp. 2705–2710.
36. 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.
37. Ma, Y.; Chen, J.; Chen, C.; Fan, F.; Ma, J. Infrared and visible image fusion using total variation model. *Neurocomputing* 2016, 202, 12–19.
38. 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.
39. Xu, H.; Ma, J.; Le, Z.; Jiang, J.; Guo, X. FusionDN: A Unified Densely Connected Network for Image Fusion. *AAAI* 2020, 34, 12484–12491.
40. Zhang, Y.; Liu, Y.; Sun, P.; Yan, H.; Zhao, X.; Zhang, L. IFCNN: A general image fusion framework based on convolutional neural network. *Inf. Fusion* 2020, 54, 99–118.
41. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 1, doi:10.1109/TPAMI.2020.3012548.
42. Chen, Y.; Xie, H.; Shin, H. Multi-layer fusion techniques using a CNN for multispectral pedestrian detection. *IET Comput. Vis.* 2018, 12, 1179–1187.
43. Hou, R.; Zhou, D.; Nie, R.; Liu, D.; Xiong, L.; Guo, Y.; Yu, C. VIF-Net: An unsupervised framework for infrared and visible image fusion. *IEEE Trans. Comput. Imaging* 2020, 6, 640–651.
44. Liu, Y.; Chen, X.; Cheng, J.; Peng, H.; Wang, Z. Infrared and visible image fusion with convolutional neural networks. *International Journal of Wavelets. Multiresolut. Inf. Process.* 2018, 16, 1850018.
45. Zhang, X.; Ye, P.; Qiao, D.; Zhao, J.; Peng, S.; Xiao, G. Object fusion tracking based on visible and infrared 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.

46. 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.
47. Piao, J.; Chen, Y.; Shin, H. A new deep learning based multispectral image fusion method. *Entropy* 2019, 21, 570.
48. 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, doi:10.1155/2020/6384831.
49. Ma, J.; Xu, H.; Jiang, J.; Mei, X.; Zhang, X.P. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. *IEEE Trans. Image Process.* 2020, 29, 4980–4995.
50. Ma, J.; Liang, P.; Yu, W.; Chen, C.; Guo, X.; Wu, J.; Jiang, J. Infrared and visible image fusion via detail preserving adversarial learning. *Inf. Fusion* 2020, 54, 85–98.
51. 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.
52. 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, doi:10.1109/JSEN.2019.2921803.
53. Li, H.; Wu, X.J. DenseFuse: A fusion approach to infrared and visible images. *IEEE Trans. Image Process.* 2018, 28, 2614–2623.
54. Sun, Y.; Zuo, W.; Liu, M. Rtfnet: Rgb-thermal fusion network for semantic segmentation of urban scenes. *IEEE Robot. Autom. Lett.* 2019, 4, 2576–2583.
55. Yan, H.; Yu, X.; Zhang, Y.; Zhang, S.; Zhao, X.; Zhang, L. Single image depth estimation with normal guided scale invariant deep convolutional fields. *IEEE Trans* 2017, 29, 80–92.
56. Li, L.; Zhang, S.; Yu, X.; Zhang, L. PMSC: PatchMatch-based superpixel cut for accurate stereo matching. *IEEE Trans. Circuits Syst. Video Technol.* 2016, 28, 679–692.
57. Liu, Y.; Chen, X.; Ward, R.K.; Wang, Z.J. Image fusion with convolutional sparse representation. *IEEE Signal Process. Lett.* 2016, 23, 1882–1886.
58. Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-focus image fusion with a deep convolutional neural network. *Inf. Fusion* 2017, 36, 191–207.
59. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014, arXiv:1409.1556.
60. Kessy, A.; Lewin, A.; Strimmer, K. Optimal whitening and decorrelation. *Am. Stat.* 2018, 72, 309–314.

61. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 26 July 2017; Volume 1, pp. 4700–4708.
62. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. In Proceedings of the Neural Information Processing Systems 2015, Montreal, QC, Canada, 7–12 December 2015; pp. 91–99.
63. Ojala, T.; Pietikäinen, M.; Harwood, D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit.* 1996, 29, 51–59.
64. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016, pp. 770–778.

---

Retrieved from <https://encyclopedia.pub/entry/history/show/14685>