ISSN: 2710-0588

Enhancing Multimodal Medical Image Fusion Using a Markov Discriminator in Generative Adversarial Networks

Xiaochen Wang[1,*], Zhi Wang[1]

[1]College of Computer Science and Technology, Qingdao University, Qingdao, China

* Corresponding Author: wxchenpro@163.com

Abstract

Multimodal medical images, comprising anatomical and functional images, offer complementary insights into organ structure and metabolism. Anatomical images depict internal organ structures, whereas functional images illustrate metabolic activity but lack detailed structural information. Multimodal image fusion integrates data from different sensors to create images enriched with diverse semantic content, overcoming the limitations of single-modality imaging. Current fusion methods based on generative adversarial networks (GANs) use discriminators that convolve the entire input image, which can reduce efficiency and result in detail loss. To address this, we propose a GAN framework with a Markov discriminator that leverages local (Markov) properties. By redesigning the discriminator and formulating the loss function based on Markov correlation principles, our method focuses on local areas, thereby enhancing network performance and preserving finer details in the fusion images.Experimental results demonstrate that our approach produces fusion images with significantly improved detail retention and superior performance compared to conventional methods.

Keywords

Adversarial generative networks, Multimodal images, Markov discriminator.

1. Introduction

Due to the inherent differences among sensors, no single sensor can capture all the scene's information. Multi-modal image fusion, which combines images from various sensors, has emerged as a crucial research area with significant applications in medical imaging, autonomous vehicles, military detection, and more. For instance, while PET provides functional information about metabolism and blood flow, its low spatial resolution and SNR limit the detection of small lesions; in contrast, MRI offers high resolution and fine tissue details [1]. Similarly, infrared images excel in capturing thermal radiation under challenging conditions, though with lower resolution than visible images that capture rich texture details [2]. Existing fusion methods are categorized into traditional techniques—such as guided filtering, pyramid and wavelet transforms, multi-scale geometric analysis, and sparse representation [3-9]， and deep learning-based approaches, notably CNNs [10] and GANs [11]. However, CNN-based methods require extensive ground truth data, and conventional GANs often lose crucial image details by focusing on global features. To address these limitations, we propose a new fusion model that first employs an encoder-decoder network to fuse source images and then refines the result using a GAN with a Markov discriminator that emphasizes local details and edge structures. Experimental evaluations on publicly available infrared-visible and medical image datasets demonstrate that our approach outperforms state-of-the-art methods in both objective quality metrics and subjective visual assessments, underscoring its effectiveness in generating high-quality fused images.

In this study, we employ an encoder-decoder model to preprocess source images, ensuring the effective integration of distinct features from both inputs in the fused image. Furthermore, we introduce a generative adversarial network (GAN) with a Markov discriminator to refine the preliminary fusion results, enhancing the preservation of local details and edge structures. The proposed model demonstrates strong generalizability and is applicable to both medical image fusion and infrared-visible image fusion tasks. Experimental evaluations on publicly available infrared-visible and medical image datasets indicate that our method outperforms state-of-the-art approaches in both objective quality metrics and subjective visual assessments, underscoring its effectiveness in generating high-quality fused images.

2. Related Work

2.1. Generative Adversarial Network

Generative adversarial networks (GANs) are deep learning models designed to generate synthetic data that closely mimics real data. A GAN comprises two neural networks—a generator and a discriminator [3]. The generator converts a random noise vector (sampled from a uniform or normal distribution) into synthetic "fake" data with the aim of fooling the discriminator [12]. The discriminator, in turn, attempts to differentiate between the real data and the fake data produced by the generator, outputting either a binary or a continuous value. The training process is a zero-sum game: the generator strives to maximize the similarity between its output and the real data, while the discriminator works to minimize the misclassification of fake data as real [13].

2.2. Image fusion methods based on Generative Adversarial Network

GAN used for image fusion can be roughly divided into three categories: classical GAN, dual-discriminators GAN and multi-GAN, as shown in Table 2-1. Ma et al., [14] proposed the FusionGAN, which is the first GAN used for image fusion tasks, and it is also the most classic GAN model used for image fusion. Wang et al, [15] proposed MFIF-GAN to solve the multi-focus image fusion problem, and added attention mechanisms and small area removal (SRR) post-processing operations to the model to refine the fusion results. However, this kind of classical GAN method will make the fusion result more biased to one of the source images, resulting in the loss of some details and feature information of the fused image. Later, Ma et al., [16] proposed DDcGAN, which is a double discriminator GAN that can better retain the respective feature information of two source images. Zhou et al., [17] proposed GIDGAN, adding gradient decision blocks and intensity decision blocks on the basis of dual-discriminator GAN, and applying repeated blur algorithm to solve the problem of multi-task image fusion. Li et al., [18] proposed RCGAN to infrared and visible image fusion. In this model, two groups of GAN models were used, one group was used to check the offset between fusion results and infrared images, and the other group was used to check the offset between fusion results and visible images. Huang et al., [19] proposed MGMDcGAN, where the first set of GAN is used to obtain structural information and the second set of GAN is used to enhance the dense information in the image.

3. Fusion Methods

In this section, the proposed GAN method under Markov discriminator is introduced in detail. Firstly, the fusion framework is presented in Section III-A. Then, the detail of Markov discriminant model is described in Section III-B. The detail of Loss function is described in Section III-C. Finally, we present our novel fusion strategy based on two stages of attention models.

3.1. Fusion Network

Our fusion network (Figure 1) consists of an adversarial generative network with one generator and two discriminators (D1 and D2). Source images 1 and 2 are input to produce a fused image. The key distinction is that D2 is a Markov discriminator, which enhances local detail and preserves both structural and functional information, thereby improving overall detail retention for multimodal image fusion.

Specifically, the process begins by feeding Source images 1 and 2 into a conventional GAN (GAN1), whose generator and discriminator produce an intermediate fused image that captures the semantic information of both inputs and retains structural gradients. This intermediate result is further refined using the dual discriminator framework—particularly leveraging the Markov discriminator—to generate the final high-quality fused image.

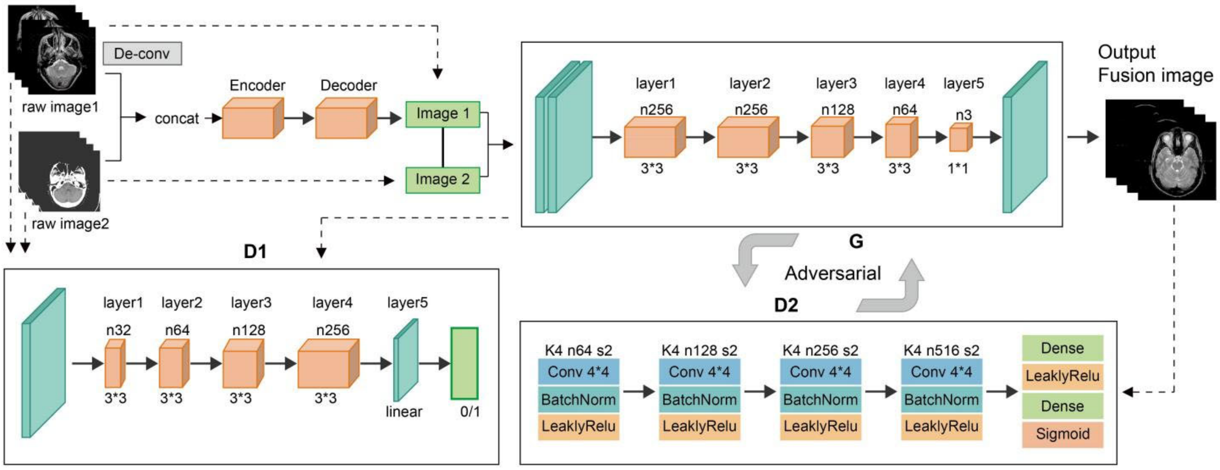

Figure 1: Model Process

In traditional GANs, the discriminator compresses the entire generated image into a single prediction (between 0 and 1) via convolution. Although a pixel gradient-based loss function helps preserve critical gradient information—essential for medical images—the fusion of infrared and visible images typically retains edge details while losing subtle gradients in other regions. To address this, we maintain the original generator and integrate a Markov random field into the discriminator.

In convolutional networks, a neuron's receptive field defines the spatial extent of the input that influences its output; larger fields capture global features, while smaller fields preserve fine details. Conventional discriminators output a single value for the whole image, essentially compressing all details into one pixel. In contrast, our Markov discriminator evaluates local regions using an n×n receptive field. For example, processing a 4×4 patch to produce a 20×20 output matrix means each entry represents a specific 4×4 region of the original image. This localized approach, which exhibits inherent Markov properties due to overlapping receptive fields, not only enhances detail preservation but also improves computational efficiency.

3.2. Markov discriminant model

The Markov discriminator model used in this paper is shown in Figure . The receptor field size is set to 63 and 94 respectively, which is experimentally verified that the discriminator can discriminate the images well.

Input size corresponds to the input size on a particular layer, output size is the output size of that layer, stride is the step size, conv size is the size of the convolution kernel used in that layer. The prediction matrix obtained by Markov discriminator represents the overall similarity

ISSN: 2710-0588

between the generated image and the source image, focusing on the similarity between the fused image and the source image. It should be advantageous in terms of structural similarity (SSIM) when applied to medical fusion images. In addition, there should be some improvement in solving the image distortion problem, like fusing infrared and visible images. These two aspects will be verified during the following experiments.

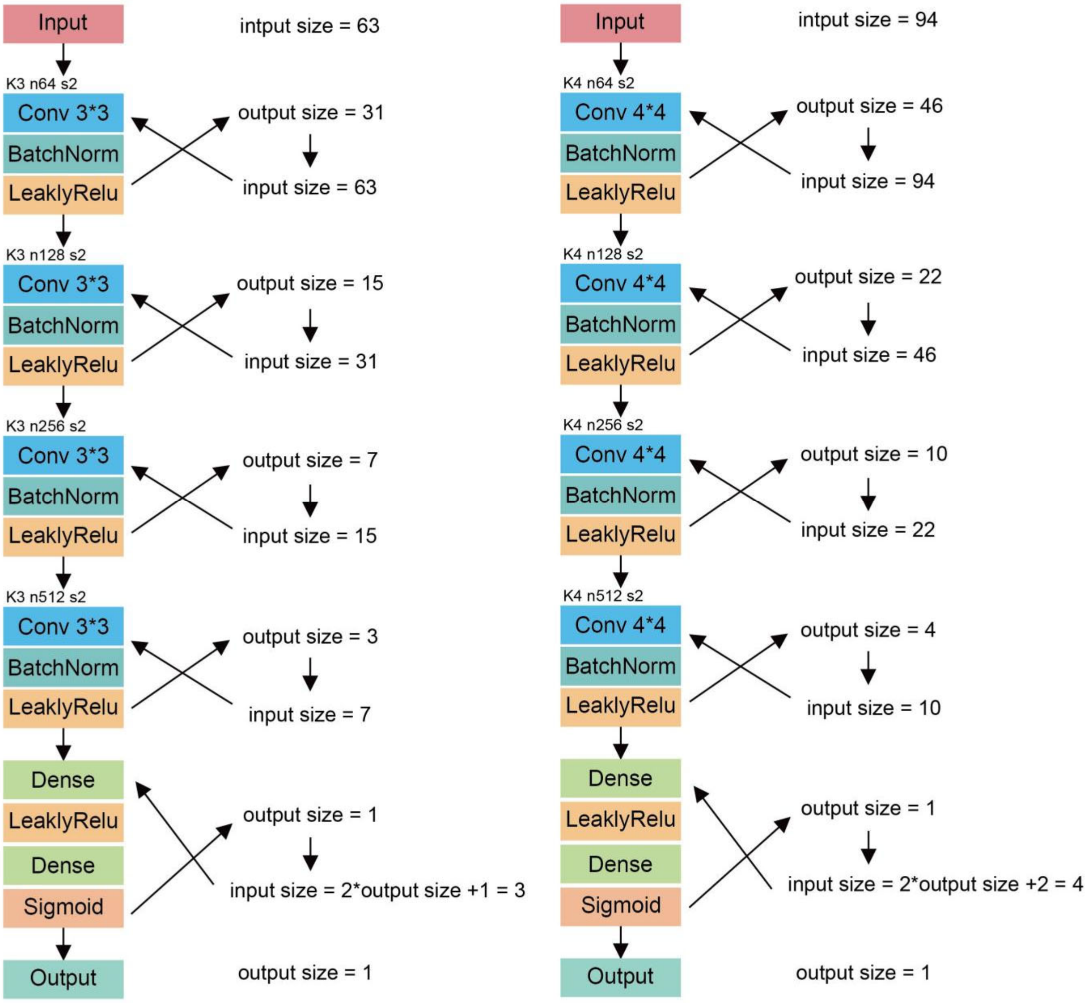

Figure 2: Discriminator network structure and sensory field size derivation process

3.3. Loss function

(1) The loss function between the conventional generator G and the discriminator D1 is based on the WGAN-GP construction, the loss function of the generator L_G is divided into adversarial loss L_{GAN} and content loss L_C.

$$L_G = L_{GAN} + \alpha L_C \tag{1}$$

The loss function of discriminator $L_{(G,D1)}$, gradient penalty term is added to GP.

$$L_{D1} = L_{(G,D1)} + \theta * GP \tag{2}$$

The loss function of the network is:

ISSN: 2710-0588

$$\mathrm{LOSS1} = \mathrm{L_{GAN}} + \alpha \mathrm{L_C} + \mathrm{L_{(G,D1)}} + \theta * \mathrm{GP} \tag{3}$$

(2) In this paper, a Markov discriminator is added to the network as D2, the overall loss function of the network for the generator G and the discriminator D2 is:

$$\mathrm{LOSS2} = \mathrm{L_G} + \mathrm{L_{D2}} \tag{4}$$

For $\mathrm{L_{D2}}$, after the introduction of Markov discriminators ,the existing GAN networks usually use L1 or L2 parametres to constrain the difference between the generated image and the real image to reduce the blurring of the generated image and improve the robustness of the generated image. The L1 parametres are calculated as follows:

$$\mathrm{S} = \sum_{i=1}^{n} |\mathrm{Y_i} - \mathrm{f}(\mathrm{x_i})| \tag{5}$$

Where $\mathrm{Y_i}$ is the input image, and f(xi) is the output image. This constraint has been added to the loss function of the generator network in this paper. Because of the nature of Markov discriminators to discriminate blocks of images (patches) in a perceptual field, the final discriminant is sigmoid for all discriminant values as the discriminant value of image, as shown in (6).

$$\mathrm{I} = \frac{1}{\mathrm{S}}(\mathrm{N} - \mathrm{k} + 2\mathrm{p}) + 1 \tag{6}$$

N is the input image size, k is the size of the sensory field, p is the fill operation, S is the step size. Since the output of the Markov discriminator is progressively determined for each patch, this paper uses structural similarity as a loss function in an adversarial generative network, which should also be computed once for each patch by SSIM and finally summed and averaged, denoted as $\overline{\mathrm{SSIM}}$. The mean and variance tend to vary drastically over the span of the whole image, and the distortion level of different blocks on the image may be different. There are I^2 patches, then $\overline{\mathrm{SSIM}}$ is:

$$\overline{\mathrm{SSIM}} = \frac{1}{\mathrm{I}^2} \sum_{i=1}^{\mathrm{I}^2} \mathrm{SSIM} \tag{7}$$

The overall loss function of the (G,D2) network is:

$$\mathrm{LOSS2} = \mathrm{L_G} + \mathrm{L}_{（\mathrm{G,D2}）} + \partial * \overline{\mathrm{SSIM}} \tag{8}$$

4. Experimental Results

4.1. Experimental Settings

We evaluated our fusion performance against three state-of-the-art methods—NSCT, CNN-based, and GHIS—using identical source images (see Fig. 3). All comparison methods were implemented with publicly available code, with parameters set as described in their respective publications.

Our experiments were conducted using TensorFlow 1.9.0, CUDA 11.5, cuDNN 8.3.0, an Intel i5-12500H CPU, an NVIDIA RTX 3050 GPU, Python 3.6.3, and MATLAB 2018b. The experimental data were obtained from Harvard Medical School, a whole brain atlas slice dataset, and public datasets BraTS2018 and MRBrainS, with infrared and visible images sourced from the TNO public dataset. The training was run for 40 epochs, with separate training for medical and IR-VIS image fusion. In our IR-VIS fusion experiments, extended training led to overfitting—likely due to excessive data augmentation generating overly similar scene samples—which was resolved by reducing the augmentation. Five quality metrics were used to quantitatively compare our fusion method with existing approaches.

4.2. Comparison experiments of medical image

In this section, we analyze the influence of Markov discriminatorto in medical image fusion performance,the result are shown in Figure 3 and Table 1. The comparison of the evaluation indexes proves that the method has more advantages in structural similarity (SSIM) and spatial frequency (SF). So, the overall image similarity can be improved by using Markov discriminator.

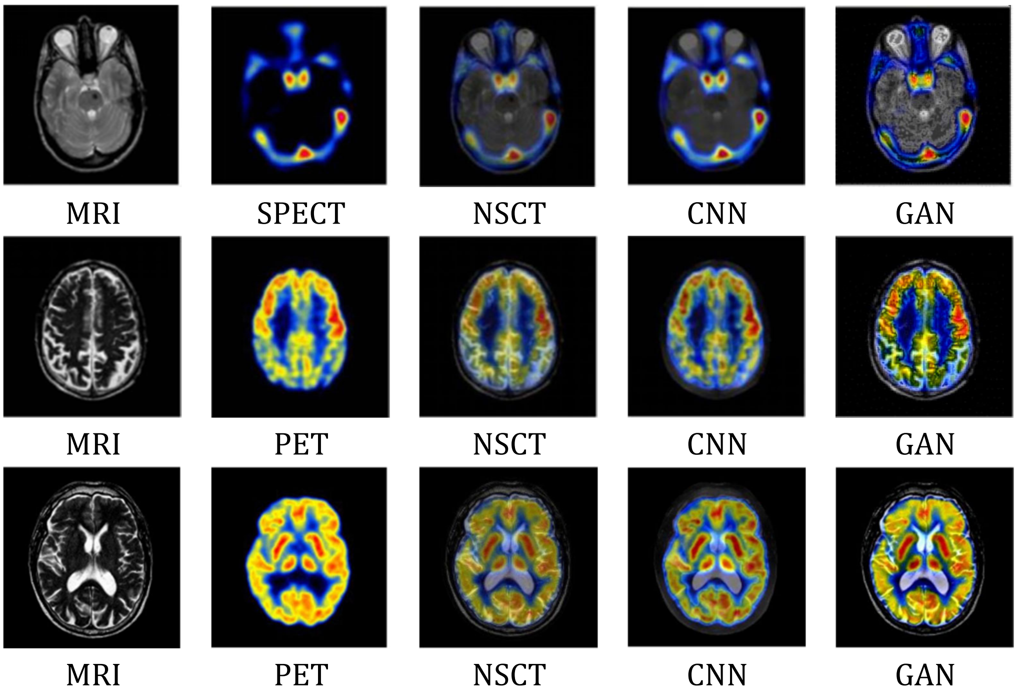

Figure 3. Image fusion results under different methods

Table 1. Evaluation indexes of image fusion results under different methods

Numble	AG	SSIM	EN	SF	QAB/F
NSCT	6.7663	0.5288/0.5947	4.3024	21.4753	0.3362
CNN	6.6526	0.4980/0.6280	4.2617	20.6185	0.3100
GHIS	4.1665	0.1584/0.0549	4.6095	9.9338	0.1562
GAN	6.6065	0.7780/0.4731	4.9159	24.2021	0.6821
NSCT	4.1118	0.6760/0.6836	4.2086	12.4342	0.3680
CNN	3.8452	0.5648/0.7564	4.1678	12.8782	0.2639
GHIS	2.9597	0.1882/0.1087	4.0643	7.95489	0.2425
GAN	4.1473	0.8578/0.5363	4.2050	27.2448	0.6338
NSCT	5.1469	0.5195/0.5319	4.6463	14.4226	0.2113
CNN	4.7181	0.4726/0.6268	4.4571	16.3462	0.1901
GHIS	4.0643	0.1942/0.0713	4.5389	10.5729	0.1474
GAN	4.4784	0.5968/0.5940	4.6200	21.3974	0.3186

4.3. GAN method with Markov discriminant for fusion of IR and visible image

To illustrate that the overall details of the image can be well preserved using Markov discriminators and the effectiveness of this method for IR and visible image fusion. The adversarial generation network is fused for IR and visible images, as shown in Fig.4. (a), (c), (e), (g) are the experimental results of fused images obtained using D1. (b), (d), (f), (h) are the fused images obtained from the experiments using Markov discriminator. the fused images of (b)-(h) obtained by the proposed method have more reasonable luminance information. The image results we obtained have clearer edges and stronger image brightness information.

The image quality is found to be significantly improved by comparison of magnified slices. It is verified that the introduction of Markov discriminator enables the network to generate fused images with high resolution, and get more outstanding image details.

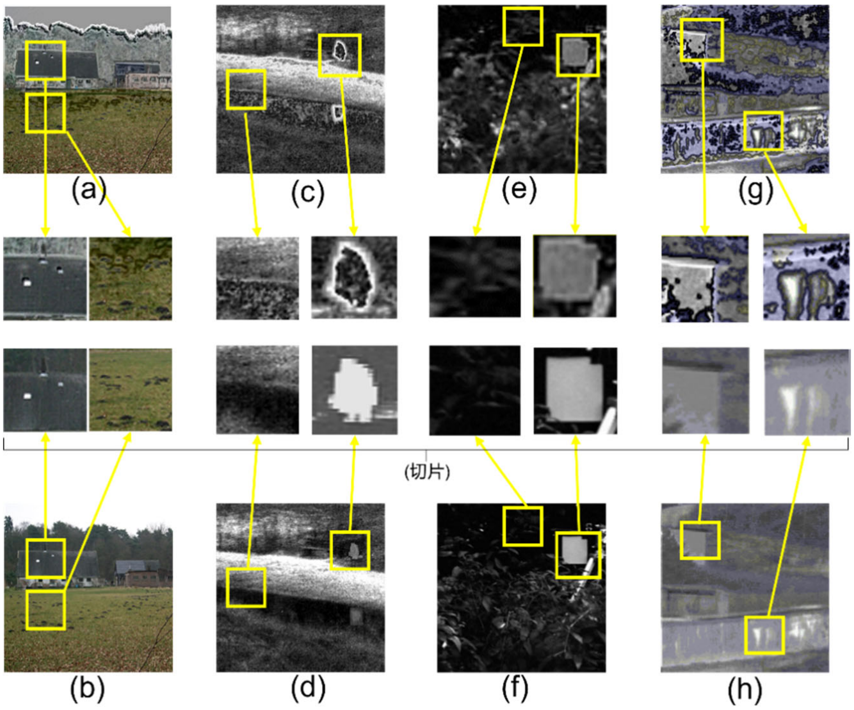

Figure 4. Comparison of IR-VIR image fusion result

5. Conclusion

In this paper, we proposed a novel image fusion architecture for multimodal image fusion tasks. First, an adversarial generative network with one generator and two discriminators is constructed. Then, the Markov discriminator model is used in this model. The receptor field size is set to 63 and 94 respectively, which is experimentally verified that the discriminator can discriminate the images well.

The result shows that adversarial generative networks are a promising approach for multimodal image fusion tasks demonstrates state-of-the-art fusion performance. By comparing the fusion results with other methods for the same source images, research shows that the advantages of fusion networks in multi-modal image fusion have been demonstrated in infrared and visible light image fusion experiments. An additional experiment on the selection of different Markov discriminant size on the fusion results shows that Markov discriminators improves the performance of the network in terms of overall image similarity.

The process of learning based on data loss for different tasks is suitable for a variety of environments.

References

[1] J. Wang, L. Yu, S. Tian, W. Wu, and D. Zhang, "AMFNet: An attention-guided generative adversarial network for multi-model image fusion," Biomedical Signal Processing and Control, vol. 78, p. 103990, 2022/09/01/ 2022, doi: https://doi.org/10.1016/j.bspc.2022.103990.

[2] Y. Fu, X.-J. Wu, and T. Durrani, "Image fusion based on generative adversarial network consistent with perception," Information Fusion, vol. 72, pp. 110-125, 2021/08/01/ 2021, doi: https://doi.org/10.1016/j.inffus.2021.02.019.

[3] T. Zhou, Q. Cheng, H. Lu, Q. Li, X. Zhang, and S. Qiu, "Deep learning methods for medical image fusion: A review," Computers in Biology and Medicine, vol. 160, p. 106959, 2023/06/01/ 2023, doi: https://doi.org/10.1016/j.compbiomed.2023.106959.

[4] S. Li, X. Kang, and J. Hu, "Image Fusion With Guided Filtering," IEEE Transactions on Image Processing, vol. 22, no. 7, pp. 2864-2875, 2013, doi: 10.1109/TIP.2013.2244222.

[5] X. Yang, H. Huo, J. Li, C. Li, Z. Liu, and X. Chen, "DSG-Fusion: Infrared and visible image fusion via generative adversarial networks and guided filter," Expert Syst. Appl., vol. 200, no. C, p. 17, 2022, doi: 10.1016/j.eswa.2022.116905.

[6] L. Kou, L. Zhang, K. Zhang, J. Sun, Q. Han, and Z. Jin, "A multi-focus image fusion method via region mosaicking on Laplacian pyramids," PLOS ONE, vol. 13, no. 5, p. e0191085, 2018, doi: 10.1371/journal.pone.0191085.

[7] H. Li, B. S. Manjunath, and S. K. Mitra, "Multisensor Image Fusion Using the Wavelet Transform," Graphical Models and Image Processing, vol. 57, no. 3, pp. 235-245, 1995/05/01/ 1995, doi: https://doi.org/10.1006/gmip.1995.1022.

[8] Y. Liu, S. Liu, and Z. Wang, "A general framework for image fusion based on multi-scale transform and sparse representation," Information Fusion, vol. 24, pp. 147-164, 2015/07/01/ 2015, doi: https://doi.org/10.1016/j.inffus.2014.09.004.

[9] Y. Liu, X. Chen, R. K. Ward, and Z. J. Wang, "Image Fusion With Convolutional Sparse Representation," IEEE Signal Processing Letters, vol. 23, no. 12, pp. 1882-1886, 2016, doi: 10.1109/LSP.2016.2618776.

[10] J. Gu et al., "Recent advances in convolutional neural networks," Pattern Recognition, vol. 77, pp. 354-377, 2018/05/01/ 2018, doi: https://doi.org/10.1016/j.patcog.2017.10.013.

[11] I. Goodfellow et al., "Generative adversarial networks," Commun. ACM, vol. 63, no. 11, pp. 139–144, 2020, doi: 10.1145/3422622.

[12] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative Adversarial Networks: An Overview," IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53-65, 2018, doi: 10.1109/MSP.2017.2765202.

[13] -. K. Wang, -. C. Gou, -. Y. Duan, -. Y. Lin, -. X. Zheng, and -. F.-Y. Wang, "- Generative Adversarial Networks:Introduction and Outlook," - IEEE/CAA Journal of Automatica Sinica, vol. - 4, no. - 4, pp. - 588, - 2017, doi: - 10.1109/jas.2017.7510583.

[14] 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, 2019/08/01/ 2019, doi: https://doi.org/10.1016/j.inffus.2018.09.004.

[15] Y. Wang, S. Xu, J. Liu, Z. Zhao, C. Zhang, and J. Zhang, "MFIF-GAN: A new generative adversarial network for multi-focus image fusion," Signal Processing: Image Communication, vol. 96, p. 116295, 2021/08/01/ 2021, doi: https://doi.org/10.1016/j.image.2021.116295.

[16] 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, doi: 10.1109/TIP.2020.2977573.

[17] H. Zhou, J. Hou, Y. Zhang, J. Ma, and H. Ling, "Unified gradient- and intensity-discriminator generative adversarial network for image fusion," Inf. Fusion, vol. 88, no. C, pp. 184–201, 2022, doi: 10.1016/j.inffus.2022.07.016.

[18] Q. Li et al., "Coupled GAN With Relativistic Discriminators for Infrared and Visible Images Fusion," IEEE Sensors Journal, vol. 0, p. 0, 06/10 2019, doi: 10.1109/JSEN.2019.2921803.

[19] J. Huang, Z. Le, Y. Ma, F. Fan, H. Zhang, and L. Yang, "MGMDcGAN: Medical Image Fusion Using Multi-Generator Multi-Discriminator Conditional Generative Adversarial Network," IEEE Access, vol. PP, pp. 1-1, 03/19 2020, doi: 10.1109/ACCESS.2020.2982016.