

Research Article

The Fusion of Unmatched Infrared and Visible Images Based on Generative Adversarial Networks

Yuqing Zhao , Guangyuan Fu , Hongqiao Wang, and Shaolei Zhang

Xi'an Research Institute of High-Tech, Shaanxi 710025, China

Correspondence should be addressed to Yuqing Zhao; zoe_rabbi@126.com

Received 27 December 2019; Accepted 27 February 2020; Published 20 March 2020

Guest Editor: Fangying Song

Copyright © 2020 Yuqing Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Visible images contain clear texture information and high spatial resolution but are unreliable under nighttime or ambient occlusion conditions. Infrared images can display target thermal radiation information under day, night, alternative weather, and ambient occlusion conditions. However, infrared images often lack good contour and texture information. Therefore, an increasing number of researchers are fusing visible and infrared images to obtain more information from them, which requires two completely matched images. However, it is difficult to obtain perfectly matched visible and infrared images in practice. In view of the above issues, we propose a new network model based on generative adversarial networks (GANs) to fuse unmatched infrared and visible images. Our method generates the corresponding infrared image from a visible image and fuses the two images together to obtain more information. The effectiveness of the proposed method is verified qualitatively and quantitatively through experimentation on public datasets. In addition, the generated fused images of the proposed method contain more abundant texture and thermal radiation information than other methods.

1. Introduction

Image fusion involves the use of mathematical methods to comprehensively process important information acquired by multiple sensors to produce a composite image that is easier to understand, thereby greatly improving the utilization rate of the image information and the reliability and automation degree of systems for target detection and recognition. Image fusion technology currently plays an important role in military, remote sensing, medicine, computer vision, target recognition, intelligence acquisition, and other applications. The fusion of visible and infrared images is one of the most useful cases of applying this technology. Infrared imaging sensors capture the heat radiation emitted by objects. On the one hand, infrared images are less affected by dark or severe weather conditions but typically lack sufficient background and contour edge details. On the other hand, visible images obtained by spectral reflection offer high resolution, excellent image quality, and rich background details but cannot detect objects under hidden or low light and night conditions. The advantages of

visible and infrared images can be combined by constructing fused images that retain richer feature information, making them suitable for subsequent processing tasks.

Image fusion is divided into three levels (from lowest to highest level): pixel-, feature-, and decision-level image fusion. Currently, the most highly studied and frequently applied image fusion is performed at the pixel level, and the majority of proposed image fusion algorithms work at this level. According to different image fusion processing domains, image fusion can be roughly divided into two categories: the spatial domain and the transform domain. Image fusion based on the spatial domain is directly conducted on the pixel gray space of an image. Common image fusion methods based on the spatial domain include linear weighted image fusion, false color image fusion, image fusion based on modulation, image fusion based on statistics, and image fusion based on neural networks [1, 2]. Image fusion based on the transformation domain involves transforming multisource images, combining coefficients from the transformation to obtain transformation coefficients of the fused images, and conducting inverse

transformation to obtain the fused images. Common fusion algorithms based on the transform domain include those based on the discrete cosine transform (DCT), the fast Fourier transform (FFT), the multiscale transform [3–5], image subspace technology [6, 7], the saliency method [8, 9], the sparse representation method [10, 11], and others [12–15].

Currently, image fusion based on the transform domain is a widely researched method. Most image fusion algorithms are based on multiscale decomposition and typically use the same transformation or representation for different source images. Since the thermal radiation in infrared images and the texture information in visible images are different, multiscale decomposition methods are not suitable for the fusion of infrared and visible images. To overcome this problem, the developers of FusionGAN [16] proposed an infrared and visible image fusion method based on the novel perspective of generative adversarial networks (GANs) [17]. In FusionGAN, the visible image and the generated fused image are both allowed to enter the discriminator. To “deceive” the discriminator, the fused image retains more visible information and drops its infrared thermal radiation information. It is often difficult to obtain perfectly matched infrared and visible images. To solve the above problems, this paper proposes a new fusion method that generates matching infrared images for visible images and produces fusion images that retain visible light texture details and infrared thermal radiation information, as shown in Figure 1.

The main contributions of this paper are as follows:

A new GAN framework is proposed to retain more information from visible and infrared images in fused images.

For visible images without matching infrared images, approximate infrared images are generated to facilitate the next image fusion.

To verify the feasibility and effectiveness of the proposed method, experiments are conducted on publicly available visible and infrared image datasets, and a comparison of the proposed method with other methods is carried out using several popular evaluation methods.

The remainder of the paper is arranged as follows: Section 2 briefly describes related studies on image fusion and GANs. Section 3 introduces the proposed method. In Section 4, the fusion performance of the proposed method is experimentally evaluated. Section 5 presents the conclusion of the paper.

2. Related Studies

In this section, several methods for the fusion of visible and infrared images are briefly introduced along with GANs.

2.1. Infrared and Visible Image Fusion Using a Deep Learning Framework. A method to generate fusion images of infrared and visible images using a deep learning framework was

proposed by Li et al. [18]. The authors decompose the source image into base and detail content. The base content is fused by weight averaging. A deep learning network is used to extract multilayer features, and then, the $L1$ -norm and a weighted average strategy are used to generate candidates for the fused detail content. The final fused detail content is obtained using a max selection strategy.

2.2. Infrared and Visual Image Fusion through Infrared Feature Extraction and Visual Information Preservation. Zhang et al. [19] proposed an image fusion method that uses infrared feature extraction and visual information preservation. This method uses quadtree decomposition and Bézier interpolation to reconstruct the infrared background and then subtracts the reconstructed background from the infrared image to obtain infrared bright features. The processed infrared features are then added to the visible images to achieve the final fusion image.

2.3. GAN and Its Derivatives

2.3.1. GAN. A GAN [17] consists of a generator G and a discriminator D that perform a minimax game together. The generator attempts to generate realistic images to trick the discriminator, and the discriminator must distinguish all the real images and the images generated by the generator until G and D reach the Nash equilibrium:

$$\min_G \max_D \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))]. \quad (1)$$

2.3.2. Conversion between Matched Images. The Pix2Pix [20] model, based on a conditional GAN (CGAN) [21], realizes the translation task of a variety of matched images. In Pix2Pix, generator G does not require random noise and only accepts one input image X as condition C with translated image Y as the output. Meanwhile, discriminator D accepts an X sample and a Y sample, where Y contains the real sample and the sample generated by the generator and D determines whether X and Y are the actual matched images.

2.3.3. Fusion of Visible and Infrared Images. Recently, Ma et al. proposed FusionGAN [16], which uses a GAN to fuse the thermal radiation information of infrared images with the high resolution and clear texture details of visible images. FusionGAN’s generator produces a fused image with infrared intensity and an additional visible gradient, and the discriminator distinguishes the fused image from the real visible image so that the fused image retains both infrared and visible image information.

3. Method

This section introduces the method proposed in this paper. First, the structural framework of the model is described, and then the model is described in greater detail.



FIGURE 1: Generated fused images.

3.1. Structural Framework of the Model. In this paper, the GAN two-player game is used to fuse visible and infrared images. The structural framework of the training process is shown in Figure 2. Visible image I_V is input as condition C into generator G_1 to generate a fake infrared image I_{RG} . Next, the visible image and the fake infrared image are input into generator G_2 together in a concatenated channel, creating fused image I_F as the output. Discriminator D_1 distinguishes between real visible image I_V and fused image I_F so that the fused image is closer to the visible image and has more visible texture details. Simultaneously, discriminator D_2 distinguishes real infrared image I_R , generated infrared image I_{RG} , and fused image I_F . Through continuous updating, the generated infrared image becomes closer to the real infrared image, and the fused image contains more thermal radiation information. In Figure 3, a visible image is input to obtain a fused image with both visible and infrared radiation information.

3.2. Network Structure. In our model, generator G_1 is the three-part convolutional neural network (CNN), as shown in Figure 4. It contains a downsampling component for convolution, an upsampling component for deconvolution, and a tanh active component. The downsampling component for convolution contains 7 convolution blocks. Except for the first block, each block contains one convolution layer and one LeakReLU active layer. The upsampling component for convolution also contains 7 convolution blocks. The convolution layer adopts a 4×4 filter, with a step size of 2 and “same” padding.

Generator G_2 is the simple five-layer CNN, as shown in Figure 5. The first two layers use a 5×5 filter, layers 3 and 4 use a 3×3 filter, and the last layer uses a 1×1 filter. Each convolution layer has a step size of 1 without padding.

Discriminators D_1 and D_2 adopt the network structure, as shown in Figure 6. The discriminator contains a four-layered CNN and a linear, fully connected layer. The first four convolution layers use a 3×3 filter with a step size of 2, no padding, and the LeakReLU active layer. The last fully connected layer is used for classification.

3.3. Loss Function. The loss function of the proposed method consists of four elements, i.e., loss function L_{G_1} of

generator G_1 , loss function L_{G_2} of generator G_2 , loss function L_{D_1} of discriminator D_1 , and loss function L_{D_2} of discriminator D_2 . The loss function L_{G_1} of generator G_1 is given by equation (2), where the first term represents adversarial loss between the generator and discriminator D_2 and the second term represents the loss of the structural similarity between the input visible image and the output infrared image:

$$\begin{aligned} \mathcal{L}_{G_1} = & \frac{1}{N} \sum_{n=1}^N [\log(1 - D(G_1(I_V^n, z)))] \\ & + \frac{1}{N} \sum_{n=1}^N [\|I_G^n - G_1(I_V^n, z)\|_1]. \end{aligned} \quad (2)$$

The loss function L_{G_2} of generator G_2 contains the adversarial losses between generator G_2 and discriminators D_1 and D_2 and the content loss of the fused image relative to the visible and infrared images:

$$\begin{aligned} \mathcal{L}_{G_2} = & \frac{1}{N} \sum_{n=1}^N (D_1(I_F^n) - c_1)^2 + \frac{1}{N} \sum_{n=1}^N (D_2(I_F^n) - c_2)^2 \\ & + \frac{\lambda}{HW} \left(\|I_F - I_R\|_F^2 + \xi \|\nabla I_F - \nabla I_V\|_F^2 \right). \end{aligned} \quad (3)$$

The loss function L_{D_1} of discriminator D_1 is defined as follows:

$$\mathcal{L}_{D_1} = \frac{1}{N} \sum_{n=1}^N (D_1(I_V^n) - b_1)^2 + \frac{1}{N} \sum_{n=1}^N (D_1(I_F^n) - a_1)^2. \quad (4)$$

The first term represents the classification results of the visible images, and the second term represents those of the fused images. The loss function L_{D_2} of discriminator D_2 is given by equation (5), which includes an additional term to represent the classification results of the generated infrared images:

$$\begin{aligned} \mathcal{L}_{D_2} = & \frac{1}{N} \sum_{n=1}^N (D_2(I_R^n) - b_2)^2 + \frac{1}{N} \sum_{n=1}^N (D_2(I_F^n) - a_2)^2 \\ & + \frac{1}{N} \sum_{n=1}^N (D_2(I_G^n) - d_2)^2. \end{aligned} \quad (5)$$



FIGURE 2: Framework of the proposed method's training process.

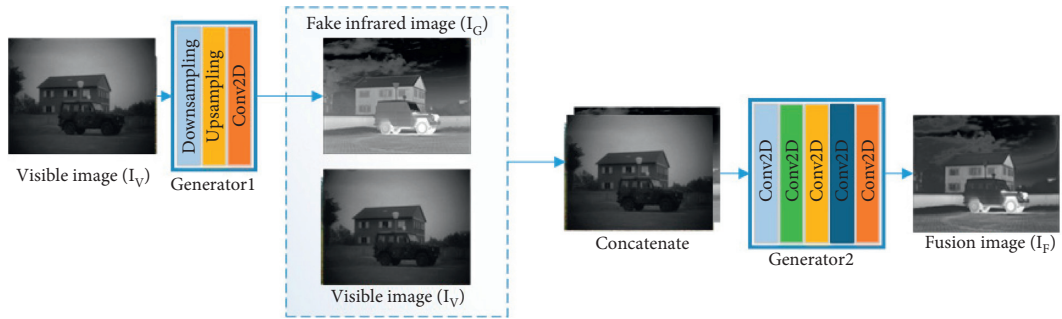


FIGURE 3: Test procedure of the proposed method.

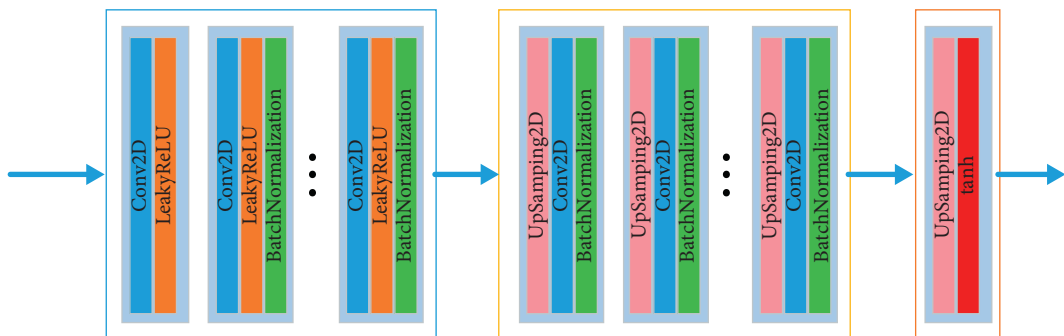


FIGURE 4: Network structure of generator G_1 .

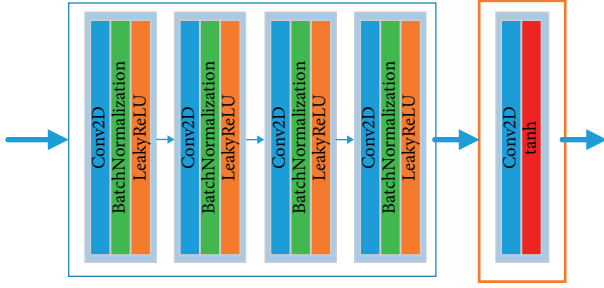
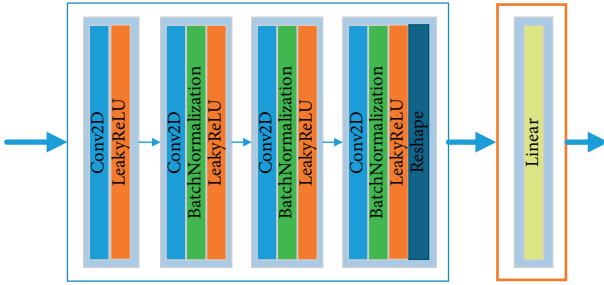
FIGURE 5: Network structure of generator G_2 .

FIGURE 6: Network structure of the discriminator.

4. Experiments

The TNO Image Fusion Dataset, a commonly used infrared and visible fusion image dataset containing night-vision infrared images and visible images of different scenes, was used to train the proposed method. A total of 56 pairs of images from the TNO dataset were selected. After the images were translated, zoomed, and flipped, 25,936 pairs of infrared and visible images were obtained. All the experiments were performed on a desktop computer with a 2.20 GHz \times 40 Intel Xeon(R) Silver 4114 CPU, GeForce GTX 1080 Ti GPU, and 64 GB of internal memory. The training parameters were set to an image batch size of 32 and a learning rate of 10^{-4} , and the generator was trained once for every 2 discriminator training runs. The chosen optimizer was Adam. It took 16.5 hours to train the model. In the first part of this section, several common image fusion evaluation indexes are introduced. In the second part, two datasets are used to validate the effectiveness of the proposed method compared with three popular image fusion methods.

4.1. Common Image Fusion Evaluation Indexes. The evaluation of fused images is performed by combining multiple indexes together. Objective quantitative evaluation methods are mainly divided into two categories: nonreference and reference image evaluation methods. Nonreference image evaluation methods include standard deviation (SD) [22] and information entropy (EN) [23]. Reference image evaluation methods include the correlation coefficient (CC) [24], peak signal-to-noise ratio (PSNR) [25], structural similarity index measure (SSIM) [25], visual information fidelity (VIF) [26], root mean square error (RMSE), and universal image quality index (UIQI) [27] methods. These indexes are defined as follows.

SD reflects the dispersion of the relative mean gray value and is mathematically defined as follows:

$$SD = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (F(i, j) - \mu)^2}, \quad (6)$$

where μ is the average value of the fused images ($M \times N$). A greater SD value indicates a higher contrast in the fused image and a typically better visual effect.

EN is a statistical feature form that reflects the average amount of information in an image. EN is mathematically defined as follows:

$$EN = - \sum_{l=0}^{L-1} p_l \log_2 p_l, \quad (7)$$

where L represents the image's gray levels and p_l represents the proportion of pixels with gray value i in the total pixels. A larger EN means that a greater amount of information exists in the fused images.

The CC measures the degree of linear correlation between a fused image and infrared and visible images and is mathematically defined as follows:

$$CC(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}, \quad (8)$$

where $\text{Cov}(X, Y)$ represents the covariance of X and Y and $\text{Var}(X)$ and $\text{Var}(Y)$ represent the variance of X and Y , respectively. The larger the CC is, the higher the degree of correlation between the fused images and visible and infrared images is, and the higher the similarity is.

The PSNR assumes that the difference between a fused and reference image is noise and is mathematically defined as follows:

$$PSNR = 10 \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right), \quad (9)$$

where MAX represents the maximum value of the image color and MSE is the mean squared error. The larger the PSNR is, the more similar the two images are. The common benchmark is 30 dB, and fused images with PSNR < 30 dB are clearly deteriorated.

The SSIM considers image distortion by comparing changes in image structure information, thereby obtaining an objective quality evaluation. The mathematical definition of SSIM is as follows:

$$SSIM(X, Y) = \left(\frac{2u_x u_y + c_1}{u_x^2 + u_y^2 + c_1} \right)^\alpha \cdot \left(\frac{2\sigma_x \sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2} \right)^\beta \cdot \left(\frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3} \right)^\gamma, \quad (10)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y),$$

where x and y are the reference image and fused image, respectively, u_x , u_y , σ_x^2 , σ_y^2 , and σ_{xy} represent the mean value and variance and covariance of images x and y , respectively, and c_1 , c_2 , and c_3 are small normal numbers to avoid having a



FIGURE 7: Qualitative fusion results of six images from the TNO dataset: (a) visible images; (b) infrared images; (c) DenseFuse fusion results; (d) DeepFuse fusion results; (e) FusionGAN fusion results; (f) proposed method fusion results.

denominator of zero. Parameters α , β , and γ are used to adjust the proportions.

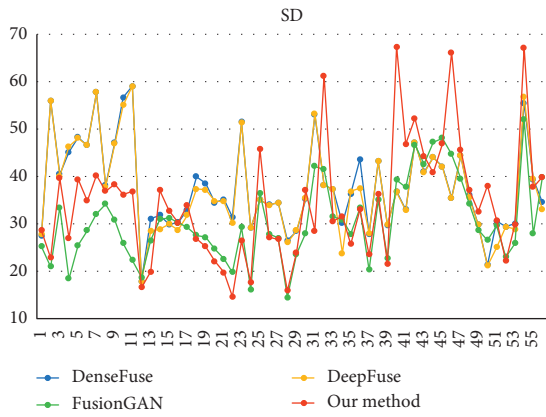
VIF is a reference image evaluation method based on natural scene statistics and the concept of image information extracted from the human visual system. The mathematical definition of VIF is as follows:

$$\text{VIF} = \frac{\sum_{k=1}^K I[(C_r^k; F^k | z_r^k)]}{\sum_{k=1}^K I[(C_r^k; E^k | z_r^k)]}, \quad (11)$$

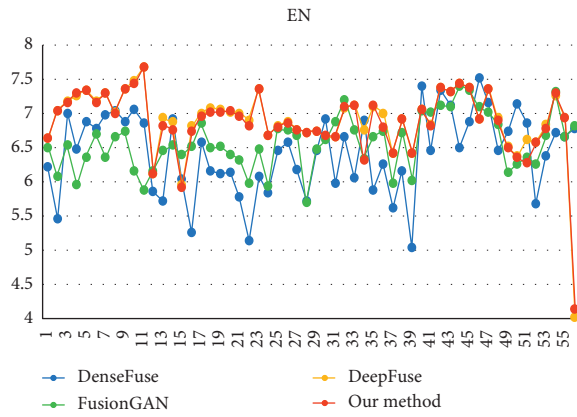
where $I(C; E | z)$ is the reference image information content and $I(C; F | z)$ is the mutual information of the reference and fused images.

The RMSE is defined as follows:

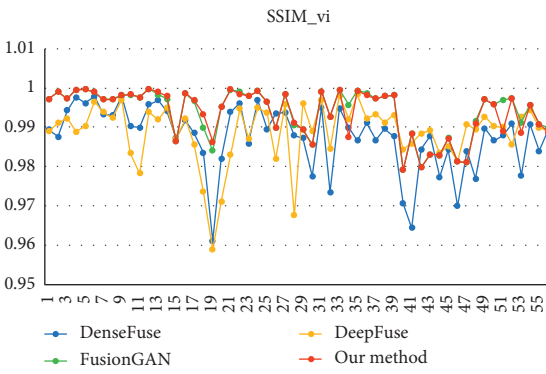
$$\text{RMSE} = \sqrt{\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [R(i, j) - F(i, j)]^2}. \quad (12)$$



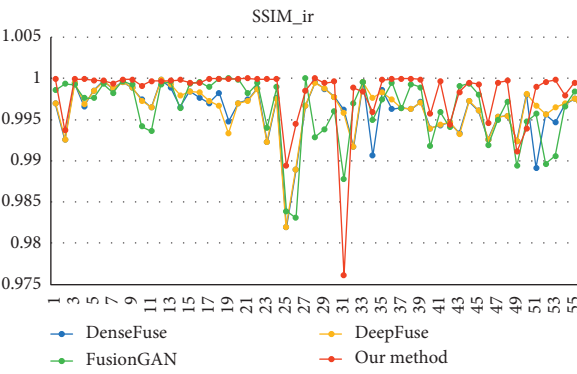
(a)



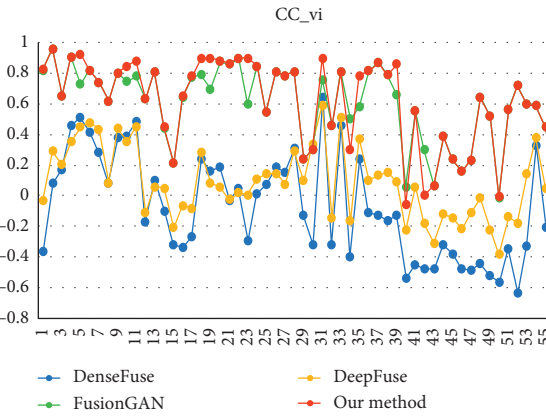
(b)



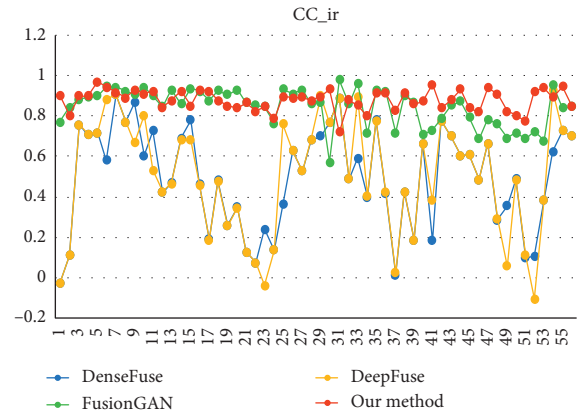
(c)



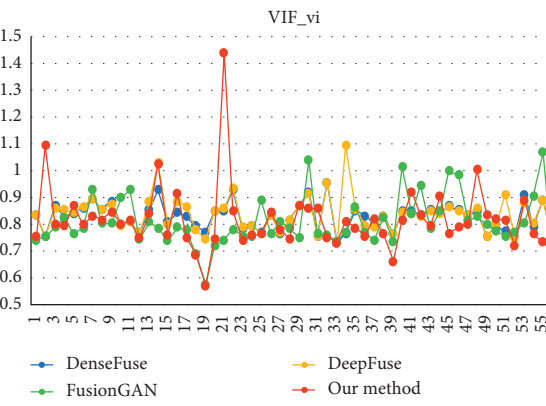
(d)



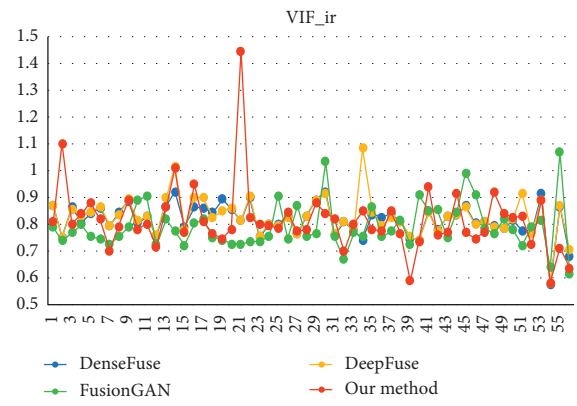
(e)



(f)



(g)



(h)

FIGURE 8: Continued.

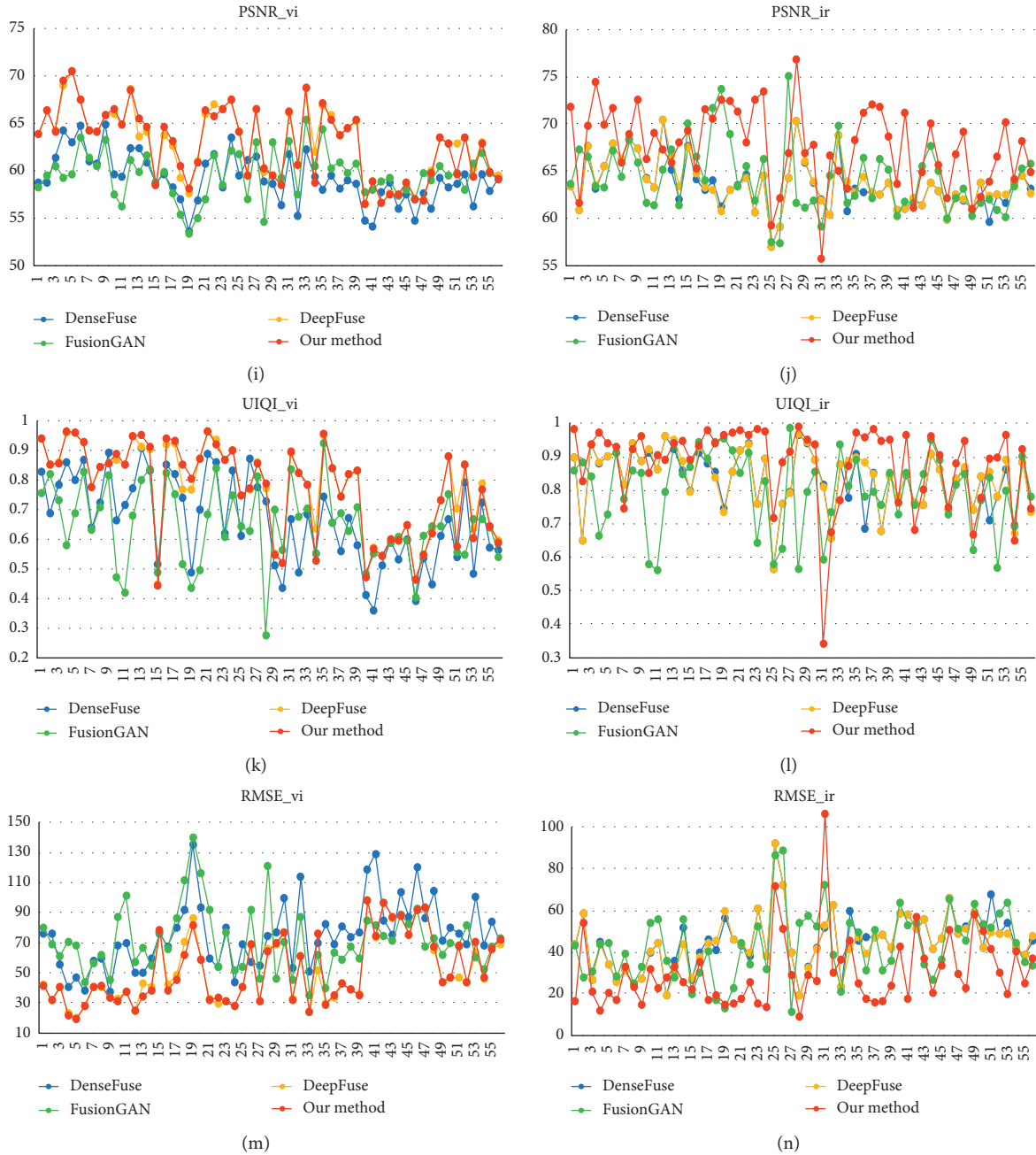


FIGURE 8: Quantitative comparison of the 14 dimensions of 8 indexes for 56 pairs of images from the TNO dataset.

The UIQI measures image distortion via a combination of three factors: loss of correlation, brightness distortion, and contrast distortion.

4.2. Experimental Validation on Fusion Performance

4.2.1. Validation by the TNO Dataset. Qualitative comparison: to provide a more intuitive observation of fusion performance, six representative images were selected for qualitative evaluation. The results of the fusion performance of the proposed method and the other three methods are shown in Figure 7. Figure 7(a) is the visible image,

Figure 7(b) is the infrared image, and Figures 7(c)–7(f) show the fusion results of DenseFuse [28], DeepFuse [29], FusionGAN [16], and the proposed method, respectively. Intuitively, all four methods fuse the texture information of the visible image and the thermal radiation information of the infrared image together to some extent. However, the fusion results of our method are more closely aligned with human visual perception, better preserve visible information, and retain more infrared information, making the image look richer and clearer with higher contrast. In addition, the target area is also more prominent than those of the other three methods.

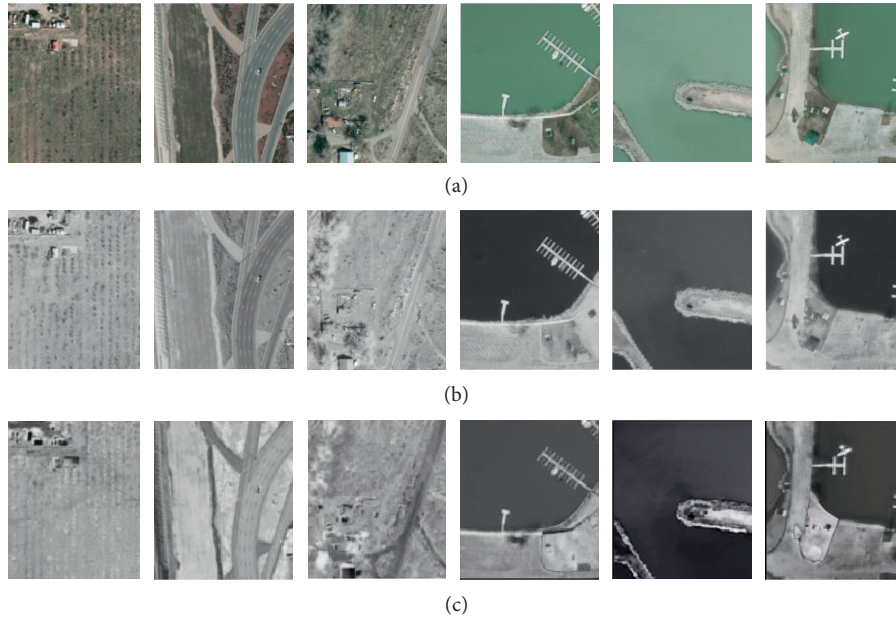


FIGURE 9: Infrared images generated using the proposed method.

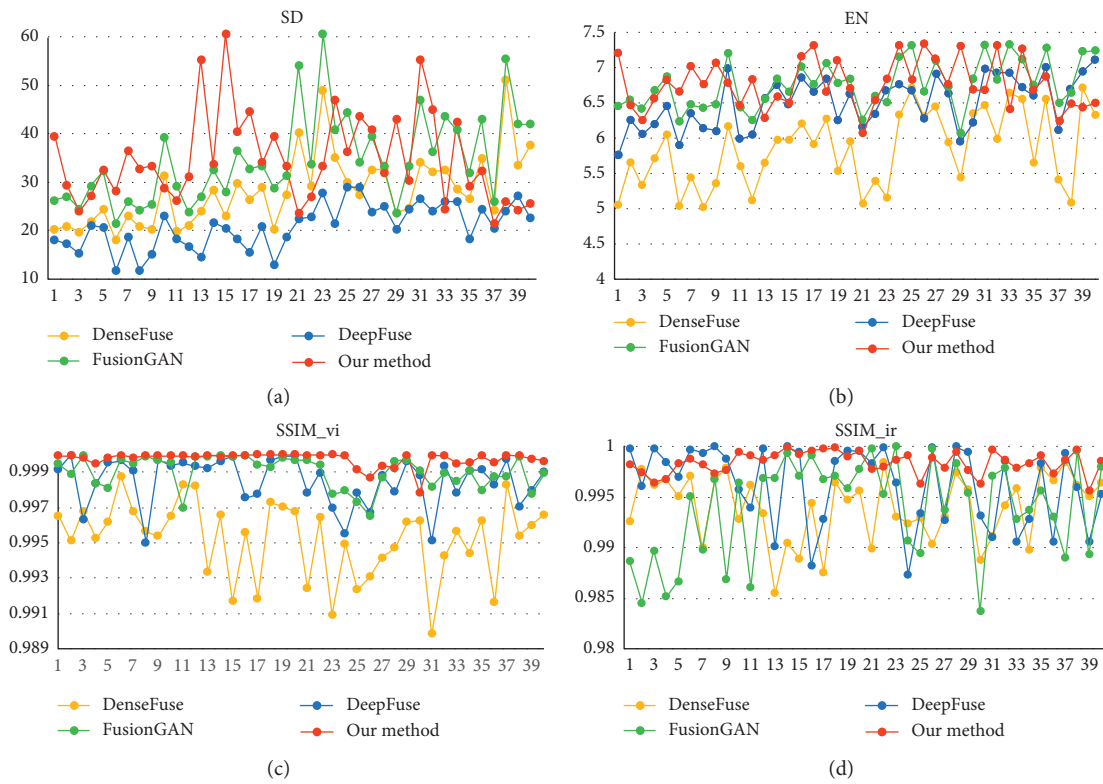


FIGURE 10: Continued.

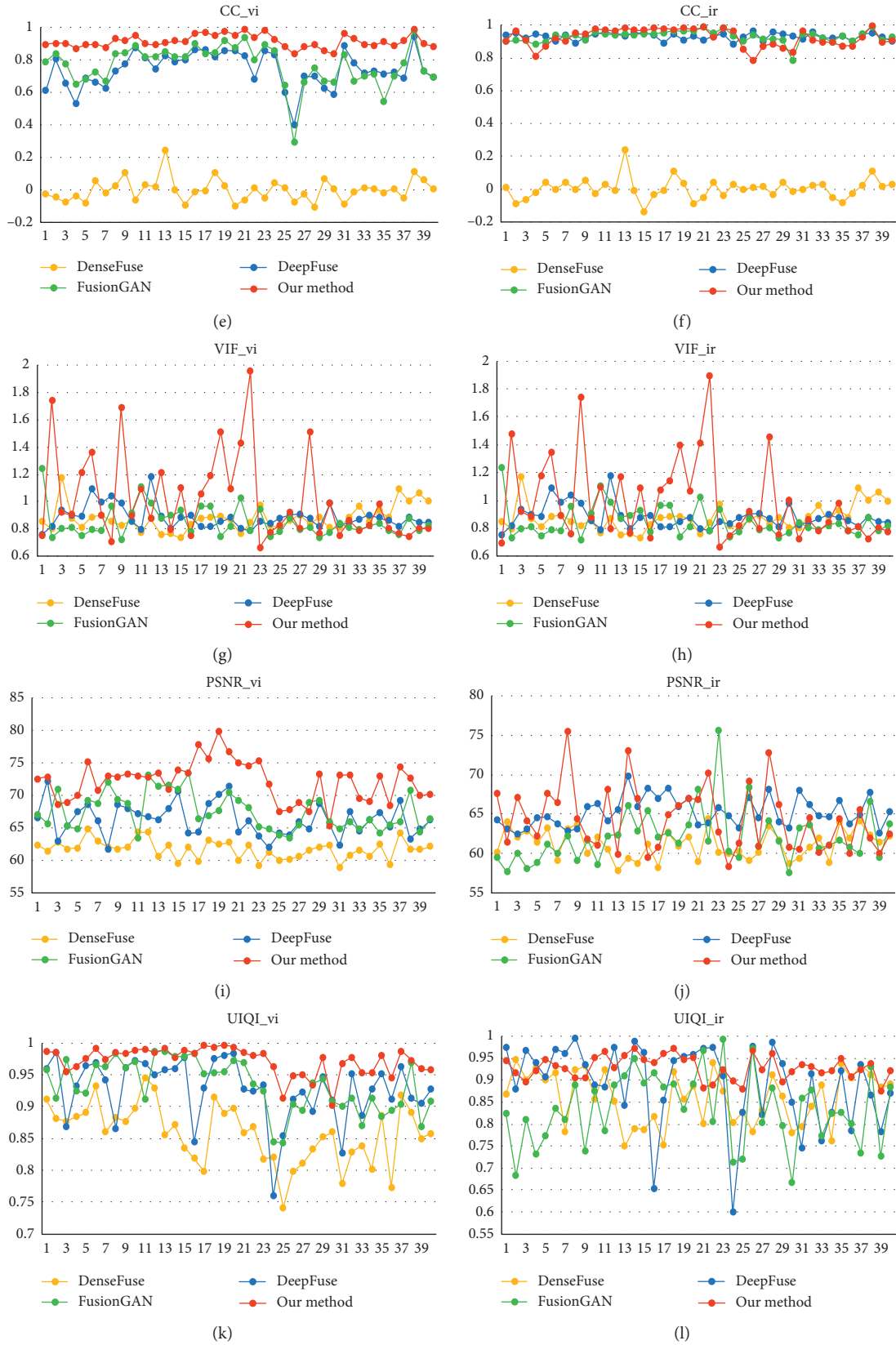


FIGURE 10: Continued.

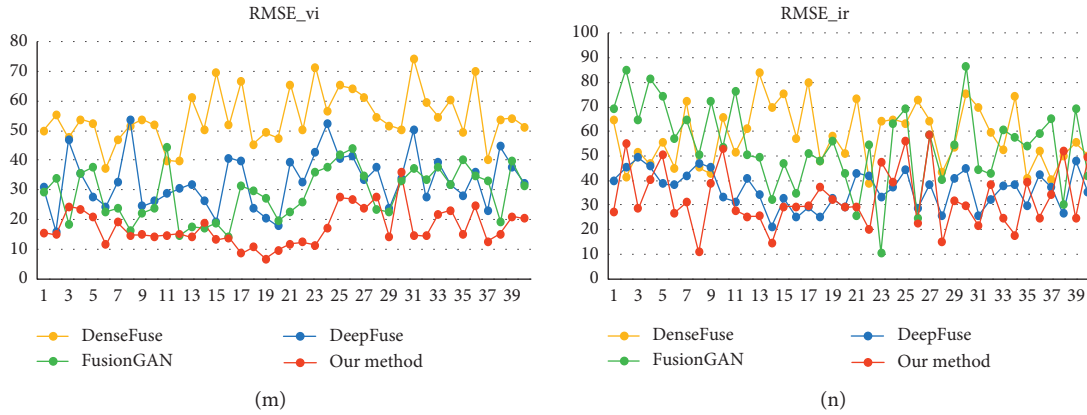


FIGURE 10: Quantitative comparison of the 14 dimensions of 8 indexes for the images from the VEDAI dataset.

Quantitative comparison: the qualitative illustrations in Figure 7 cannot objectively determine the quality of the results. Therefore, the fusion methods were further compared using quantitative methods. Eight indexes were used on 56 pairs of images from the TNO dataset, of which six indexes require a reference (the fused image refers to the corresponding visible and infrared images). The results are shown in Figure 8. The proposed method achieves the best performance for the majority of image pairs, and, for some individual image pairs, the comprehensive fusion index is much higher than that of the other methods. In addition, compared with the other three methods, the proposed method has the best average of the evaluation indexes. Because the proposed method uses two discriminants when referring to visible images, it is comparable to the FusionGAN method. However, when referring to infrared images, the proposed method considerably outperforms the FusionGAN method. This shows that the proposed method retains more infrared thermal radiation information while retaining sufficient visible texture information. Thus, our training framework is both effective and essential.

4.2.2. Validation by the VEDAI Dataset. The Vehicle Detection in Aerial Imagery (VEDAI) dataset [30] contains publicly available aerial orthogonal normalization images from Utah's State Geographic Information Database (SGID) by the Automated Geographic Reference Center (AGRC). These aerial orthogonal normalization images contain a wide variety of vehicles, backgrounds, and obfuscated objects. Each image has three visible channels and one near-infrared channel. In this section, the DenseFuse, DeepFuse, FusionGAN, and proposed method are further tested using the VEDAI dataset. Figure 9 shows the proposed method's generation of infrared images from visible images. Figure 9(a) shows the visible images, Figure 9(b) shows the actual infrared images corresponding to the visible images from the dataset, and Figure 9(c) shows the infrared images generated directly from the visible images using the proposed method. Figure 9 illustrates that the infrared images generated by the proposed method accurately reflect actual

thermal radiation information while maintaining consistency with the real infrared images.

A total of 40 images from the VEDAI dataset were selected for a quantitative comparison. Figure 10 shows a quantitative analysis of the fusion results of the four methods using 8 evaluation indexes. The proposed method achieves the best SSIM, CC, PSNR, UIQI, and RMSE results on the majority of images. Compared with the other three methods, the averages of the other indexes of the proposed method are the highest. Our experiments show that the proposed method generalizes well to other datasets.

5. Conclusion

In this paper, we propose a new fusion method that generates a matched infrared image from a visible image and generates a fused image that retains more visible texture details and infrared heat radiation information than other methods. Experimental evaluations on two public datasets show that the proposed method generates infrared images with thermal radiation information relatively consistent with real infrared images and generates fused images with clearly prominent texture information and rich thermal radiation information. A quantitative analysis of eight evaluation indexes for fused images shows that the proposed method produces better visual effects while retaining more information than other methods.

Data Availability

All data included in this study are available upon request by contact with the corresponding author.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] T. Xiang, L. Yan, and R. Gao, "A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain," *Infrared Physics & Technology*, vol. 69, pp. 53–61, 2015.

- [2] W. Kong, L. Zhang, and Y. Lei, "Novel fusion method for visible light and infrared images based on NSST-SF-PCNN," *Infrared Physics & Technology*, vol. 65, pp. 103–112, 2014.
- [3] S. Li, B. Yang, and J. Hu, "Performance comparison of different multi-resolution transforms for image fusion," *Information Fusion*, vol. 12, no. 2, pp. 74–84, 2011.
- [4] G. Pajares and J. Manuel de la Cruz, "A wavelet-based image fusion tutorial," *Pattern Recognition*, vol. 37, no. 9, pp. 1855–1872, 2004.
- [5] Z. Zhang and R. S. Blum, "A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application," *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1315–1326, 1999.
- [6] D. P. Bavirisetti, G. Xiao, and G. Liu, "Multi-sensor image fusion based on fourth order partial differential equations," in *Proceedings of the 20th International Conference on Information Fusion (Fusion)*, pp. 1–9, Xi'an, China, October 2017.
- [7] W. Kong, Y. Lei, and H. Zhao, "Adaptive fusion method of visible light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization," *Infrared Physics & Technology*, vol. 67, pp. 161–172, 2014.
- [8] X. Zhang, Y. Ma, F. Fan, Y. Zhang, and J. Huang, "Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition," *Journal of the Optical Society of America A*, vol. 34, no. 8, pp. 1400–1410, 2017.
- [9] J. Zhao, Y. Chen, H. Feng, Z. Xu, and Q. Li, "Infrared image enhancement through saliency feature analysis based on multi-scale decomposition," *Infrared Physics & Technology*, vol. 62, pp. 86–93, 2014.
- [10] J. Wang, J. Peng, X. Feng, G. He, and J. Fan, "Fusion method for infrared and visible images by using non-negative sparse representation," *Infrared Physics & Technology*, vol. 67, pp. 477–489, 2014.
- [11] S. Li, H. Yin, and L. Fang, "Group-sparse representation with dictionary learning for medical image denoising and fusion," *IEEE Transactions on Biomedical Engineering*, vol. 59, no. 12, pp. 3450–3459, 2012.
- [12] J. Zhao, G. Cui, X. Gong, Y. Zang, S. Tao, and D. Wang, "Fusion of visible and infrared images using global entropy and gradient constrained regularization," *Infrared Physics & Technology*, vol. 81, pp. 201–209, 2017.
- [13] J. Ma, Z. Zhou, B. Wang, and H. Zong, "Infrared and visible image fusion based on visual saliency map and weighted least square optimization," *Infrared Physics & Technology*, vol. 82, pp. 8–17, 2017.
- [14] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Information Fusion*, vol. 31, pp. 100–109, 2016.
- [15] 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.
- [16] 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.
- [17] I. Goodfellow, "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, pp. 2672–2680, Morgan Kaufmann Publishers, Burlington, MA, USA, 2014.
- [18] H. Li, X.-J. Wu, and J. Kittler, "Infrared and visible image fusion using a deep learning framework," in *Proceedings of the 24th International Conference on Pattern Recognition (ICPR)*, IEEE, Beijing, China, pp. 2705–2710, 2018.
- [19] Y. Zhang, L. Zhang, X. Bai, and L. Zhang, "Infrared and visible image fusion through infrared feature extraction and visual information preservation," *Infrared Physics & Technology*, vol. 83, pp. 227–237, 2017.
- [20] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1125–1134, Honolulu, HI, USA, July 2017.
- [21] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, <https://arxiv.org/abs/1411.1784>.
- [22] Y.-J. Rao, "In-fibre Bragg grating sensors," *Measurement Science and Technology*, vol. 8, no. 4, pp. 355–375, 1997.
- [23] J. W. Roberts, J. A. Van Aardt, and F. B. Ahmed, "Assessment of image fusion procedures using entropy, image quality, and multispectral classification," *Journal of Applied Remote Sensing*, vol. 2, no. 1, Article ID 023522, 2008.
- [24] M. Deshmukh and U. Bhosale, "Image fusion and image quality assessment of fused images," *International Journal of Image Processing (IJIP)*, vol. 4, no. 5, p. 484, 2010.
- [25] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *Proceedings of the International Conference on Pattern Recognition*, IEEE, Istanbul, Turkey, pp. 2366–2369, August 2010.
- [26] Y. Han, Y. Cai, Y. Cao, and X. Xu, "A new image fusion performance metric based on visual information fidelity," *Information Fusion*, vol. 14, no. 2, pp. 127–135, 2013.
- [27] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, 2002.
- [28] H. Li and X.-J. Wu, "DenseFuse: a fusion approach to infrared and visible images," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2614–2623, 2018.
- [29] K. R. Prabhakar, V. S. Srikar, and R. V. Babu, "DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs," in *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 4724–4732, Venice, Italy, December 2017.
- [30] S. Razakarivony and F. Jurie, "Vehicle detection in aerial imagery: a small target detection benchmark," *Journal of Visual Communication and Image Representation*, vol. 34, pp. 187–203, 2016.