Infrared and Visible Image Fusion Using Two-layer Generative Adversarial Network

Lei Chen[1], Jun Han[1], Feng Tian[2]

[1]Xi'an Technological University, Xi'an, Shaanxi Province, 710021, China

[2]Bournemouth University, Poole, BH12 5BB, UK

ABSTRACT

Infrared images can distinguish targets from their backgrounds based on difference in thermal radiation, whereas visible images can provide texture details with high spatial resolution. The fusion of the infrared and visible images has many advantages and can be applied to applications such as target detection and recognition. This paper proposes a two-layer generative adversarial network (GAN) to fuse these two types of image. In the first layer, we generate fused images using two GANs: one uses the infrared image as input and the visible image as ground truth, and the other with the visible as input and the infrared as ground truth. In the second layer, we transfer one of the two fused images generated in the first layer as input and the other as ground truth to GAN to generate the final fused image. Experiment results have demonstrated that the proposed approach is able to achieve better performance against existing methods on preserving both texture details from visible images and thermal information from infrared images.

Keywords: Infrared and visible images; Image fusion; Generative adversarial network; Deep learning

1. Introduction

The fusion of infrared (IR) and visible images is to integrate inherent properties of both images and merge their salient information to produce a fused image. As shown in Fig. 1 (a)(b), visible images usually have the characteristics of high resolution, and rich texture information, while IR images that are captured by infrared sensors are normally of high contrast and more robust against external factors such as weather[1]. The fusion of advantages of both IR and visible images has made success on target detection in military and civilian applications[2,3]. In the past decades, based on the fusion strategies and theories[4], many fusion algorithms have been proposed, such as hybrid models[5,6], multi-scale transform[7,8], saliency-based methods[9,10], sparse representation[11,12], neural network[13,14], subspace[15,16], and others[17,18]. Traditional methods, like those based on multi-scale transform or sparse representation, make pixel-level operations directly on source images (IR and visible), though it's challenging to establish direct correlation between the pixels of the source images that have completely different imaging principles.

In recent years, research on deep learning (DL) has become more and more extensive, especially in image processing, and it starts to apply to image fusion by extracting deep features automatically. Liu et al. [19,20] ~~initial~~ proposed the methods to use CNN for IR and visible image fusion. The method has a good performance in image generation, but it is difficult to control the process of generating images, and some information had lost in the process. Recently, some scholars presented other DL frameworks to fuse IR and visible images[21,22]. They proposed a CSR-based framework and densely connected convolutional networks for image fusion, respectively. However, DL-based methods are not end-to-end models, and they need to be trained beforehand. No matter whether the network weights are generated by training or provided by other feature extraction models, some transforms or operations, like those in siamese convolutional network and VGG-network, are still needed to accomplish the final fusion process[23,24]. To address this issue, Ma eta. [25] proposed a novel IR and visible image fusion method, FusionGAN, based on a generative adversarial network (GAN). FusionGAN works just like an adversarial game, between retaining the thermal radiation information and the appearance texture information[25]. The network model of FusionGAN is effective and relatively simple, and the fusion results can be further improved by adjusting the network and loss functions. Since the discriminator in FusionGAN has no ground truth to determine whether the data is true or false, FusionGAN uses the visible image as the ground truth image, which leads to the fused image tend to have more texture details and less thermal radiation information.

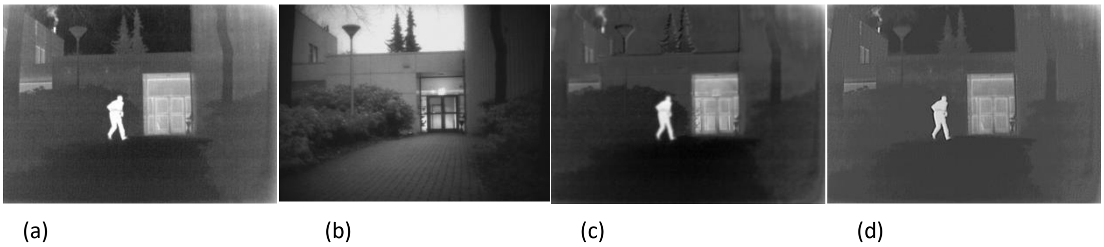


Figure 1. (a) IR image; (b) Visible image; (c) Fused image by FusionGAN; (d) Fused image by our method, showing clear texture of the trees and the building .

Motivated by FusionGAN, in this paper we propose a two-layer GAN network for the IR and visible image fusion, as shown in Fig. 2. In the first layer, we feed IR image or visible image to two generators (G) to generate fused images respectively, and use the other image as ground truth. Then in the second layer we feed one of the two images generated by the first layer to G to generate fused image, and feed the other to discriminator (D) as the ground truth image.

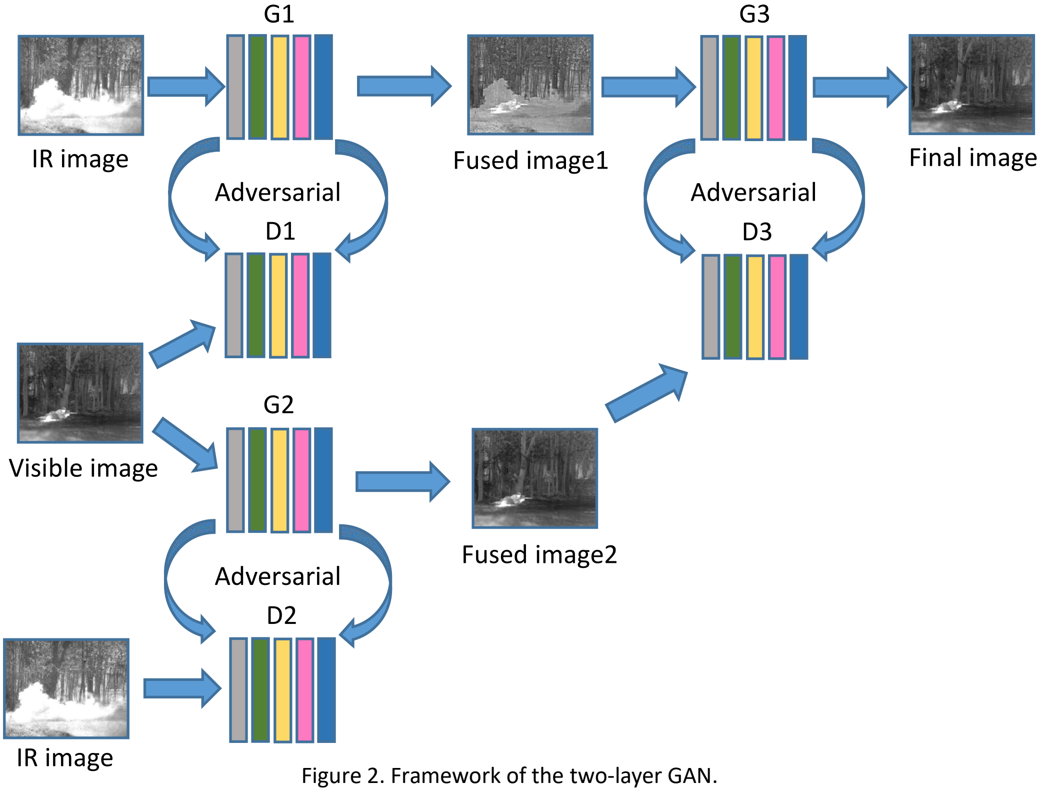


Figure 2. Framework of the two-layer GAN.

Our experiments have demonstrated that our proposed network is able to produce fused images with improved performance against the state-of-arts by preserving both texture details from visible image and thermal radiation information from IR image. Fig. 1 (c)(d) shows an example of fused images by FusionGAN and our method.

Among the remaining sections of this paper, Section 2 reviews traditional and deep learning based fusion methods. Section 3 presents the proposed approach, including the network structure, loss function and training flow of our model. Section 4 gives our experimental results, including qualitative illustration and quantitative analysis and comparison to the state-of-arts. We conclude the paper with future work in Section 5.

2. Related works

2.1 Conventional Fusion Methods

The conventional methods mainly include sparse representation based methods[26,27], the multi-scale decomposition (MSD)-based methods[28,29], hybrid models-based methods[5,30], and neural network-based methods[31,32]. MSD-based methods resolve the source images into components of different scales, and each component indicates sub-images of different scales. Then, the methods fuse the sub-images at different scales according to the given rules and obtain the fusion image through the corresponding inverse multi-scale transforms. Sparse representation based methods tend to build an over-complete dictionary from vast high-quality natural images. Then, the source images can be sparsely represented by the learned dictionary. Most of the image fusion methods based on neural network mainly adopt pulse-coupled neural network (PCNN) or its variants. PCNN based methods extract local details through its own biological characteristics and gain better fusion results when the gradient and phase information are considered beforehand.

2.2 Deep learning based image fusion

In the last few years, convolutional neural networks (CNNs) have attained great success in many image processing applications. Prabhakar et al.[33] presented a deep learning framework for fusing static multi-exposure images. This method proposed a new idea for information fusion by CNNs. For the image fusion, Liu et al.[34] presented a Siamese convolutional network to get the weight graph and solve the multi-scale problem by the image pyramid. Moreover, another fusion approach, "DenseFuse", was proposed by Li et al.[35], adopt dense blocks to store more information from the middle layers. Obviously, DL based methods have made a great breakthrough in the image fusion. However, the method based CNN must satisfy a critical precondition that is, the ground truth should be available beforehand. On this case, the CNN methods for the image fusion build a deep model to determine every patches' fused degree in the source images, and then calculate a weight map for generating the final image[34]. Li et al.[35] proposed Dense Net based CNN for make full use of each convolution layer and get good results. However, the aspects like network architecture can still be further improved. Recently, Ma et al.[25] innovatively proposed the method with GAN, and formulated the fusion as an adversarial game between keep the thermal radiation information and visible texture information. Instead of pre-training, the method used IR and visible image patches to train the network.

2.3 Generative adversarial networks

The GAN was initial proposed by Goodfellow et al.[36] In 2014, and it has attracted wide attention in deep learning. The algorithm is based on minimax game and provided a easy and effectual method for evaluating target distribution and generating new samples. GAN framework include two models: generative model(G) and discriminative model(D). The model build an adversarial game between the G and D, G take the noise P_z as input, and try to generate a different sample data to fool D. D determine whether the input is from the real data distribution. Finally, the samples be generated by generator which cannot recognize by D as the final data. The Schematic diagram as follows:

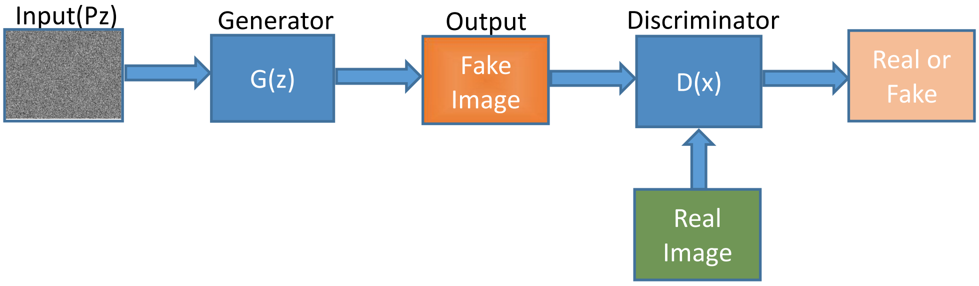


Figure 3. The schematic diagram of the generative adversarial network

The adversarial formula as follows:

$$\min_{G} \max_{D} V_{GAN}(G, D) = E_{x\sim Pdata(x)}[logD(x)] + E_{z\sim Pz(z)}[log(1 - D(G(z)))] \quad (1)$$

In the function $V_{GAN}(G,D)$, the first item signifies the entropy of data from real distribution judged by D. D tries to maximize the mark. The second item is the entropy of the data generated by G from random inputs discriminated by D. D tries to make this item bigger and equivalently minimize the $D(G(z))$ to assurance its ability to distinguish untruth from truth. In short, D wants to maximize the function $V_{GAN}(G,D)$ and minimize the function $V_{GAN}(G,D)$.

Compared to the traditional fusion method, GAN has many advantages such as no need to use Markov chain or expand approximate reasoning network for training and sample generation[36]. In this paper we take these advantages propose a two-layer GAN network for fusing IR and visible images.

3. Our Fusion Framework

The structure of our proposed two-layer GAN is illustrated in Fig. 2. It's worth mentioning that, during the experiment, we have also changed the input of G in the second layer, i.e. taking the second fusion result of the first layer as the input, and the first fusion result as the ground truth. The final fusion results were quite similar, thanks to the symmetric features of the proposed structure.

3. 1 The Structure of G

Our model has three generators that have the same structure as shown below in Fig. 4.

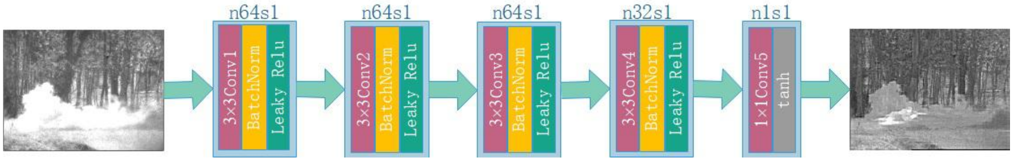


Figure 4. The network structure of the generator.

G is a convolution neural network and has five layers, the first four layers are set to 3×3 filters, and 1×1 filter in the last layer. Every layer's stride is set to 1, and no padding operation in convolution. In order to keep the source's details, only introduced convolutional layer and no downward sampling, which also keeps the same size of the input and output images [37]. Moreover, to avoid the disappearing gradient, we adopt the rules of deep convolutional GAN[38] for batch normalization and activation function. To surmount the sensitivity to data initialization, we used batch normalization in the first four layers, which can make our model more stable and also help the gradients to back propagate to each layer effectively. For the activation function, we use leaky ReLU activation function in the first four layers, and the tanh activation function in the last layer.

3. 2 The Structure of D

Our model has three discriminators which have the same structure but different input images and ground truth images, as illustrated in Fig. 2. The network structure of the generators is as follows:

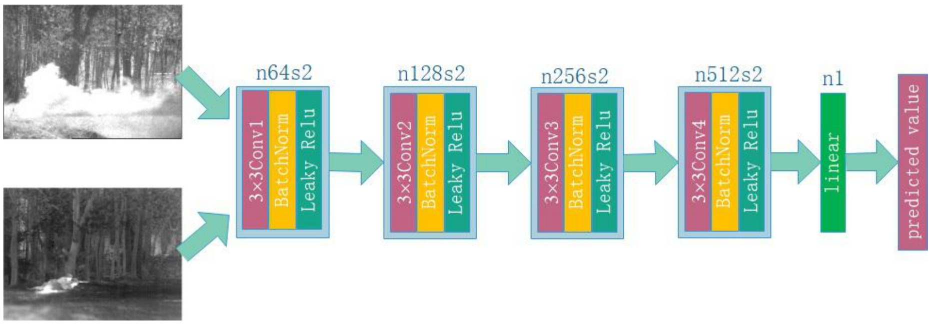


Figure 5. The network structure of the discriminator.

D is a five-layer convolution neural network which has 3×3 filters in the first four layers, and the stride set as 2 without padding. This is dissimilar from G, for the D is a classifier, which first extracts feature map from the input image and then classifies it. Hence, it works in the same way as the pooling layer, setting the stride as 2. Want not to introduce noise, we do not pad the input image. We used the batch normalization layer from the layer 1 to layer 4. In addition, we adopt the Leaky ReLU activation function in the first four layers. The last layer is a linear layer for classification.

The core of GAN algorithm is the adversarial loss, through which we can establish the adversarial relationship between G and D. Below we introduce the adversarial loss of our model in detail.

3. 3 The Loss Function of G

For the network have three generators and three discriminators, we introduce the loss functions, respectively. Motivated by FusionGAN[25], we split the loss function into two terms: the adversarial loss and content loss for the generators, and the positive loss and negative loss for the discriminators.

G1: The first G's input is the IR image, and output is the fused image $\mathrm{I}_{\mathrm{IRV}}$. The loss function consists of two terms:

$$G\mathcal{L}_1 = \mathcal{L}_1^{\mathrm{A}} + \mathcal{L}_1^{C} \tag{2}$$

Where $G\mathcal{L}_1$ denotes the G's total loss. The first term $\mathcal{L}_1^{\mathrm{A}}$ denotes the adversarial loss between the generator G and discriminator D, defined as:

$$\mathcal{L}_1^{\mathrm{A}} = \frac{1}{N}\sum_{n=1}^{N} (D(I_{IRV}) - \ell_1))^2 \tag{3}$$

Where $\mathrm{I_{IRV}}$ and N denote the fused image with $\mathrm{n} \in \mathbb{N}_\mathrm{n}$ and the number of fused images, respectively. ℓ_1 is the value that generator tricks discriminator to believe in fake data, which in our method is set as 0.8.

The second term $\mathcal{L}_1^{\mathrm{C}}$ represents the content loss, since the thermal radiation information of infrared image is characterized by its pixel intensities. The edge information in the image is particularly important for target detection, so it is an important basis for the loss function, and we added the weight of the edge information in the formula. We also enforce the fused image $\mathrm{I_{IRV}}$ to have similar intensities as $\mathrm{I_{IR}}$ specifically. Let $\mathrm{I'_{IRV}}$ as the edge points of the fused image, $\mathrm{I'_{IR}}$ as the points that corresponds to $\mathrm{I'_{IRV}}$ in the image of $\mathrm{I_{IR}}$. The content loss is then defined as:

$$\mathcal{L}_1^{C} = \frac{1}{S}(\|I_{IRV} - I_{IR}\|)_F^2 + \lambda\frac{1}{N}\sum_{i=0}^{N} (I'_{IRV} - I'_{IR})^2 \tag{4}$$

Where S represents the area of the input images, $\|\cdot\|_\mathrm{F}$ stands for the matrix Frobenius norm, and λ is the weight which we set as 2 in our method.

G2: The second G's input is the visible image, and the output is the fused image $\mathrm{I_{VIR}}$. Similar to G1, the loss function consists of two terms:

$$G\mathcal{L}_2 = \mathcal{L}_2^{\mathrm{A}} + \mathcal{L}_2^{C} \tag{5}$$

$$\mathcal{L}_2^{\mathrm{A}} = \frac{1}{N}\sum_{n=1}^{N} (D(I_{VIR}) - \ell_2)^2 \tag{6}$$

$$\mathcal{L}_2^{C} = \frac{1}{S}(\|\nabla I_{VIR} - \nabla I_V\|)_F^2 \tag{7}$$

The explanation of these three formulas is the same as above for the generator G1. ∇ means the gradient operator.

G3: The third G's input is the fused image $\mathrm{I_{IRV}}$ from the first layer and the output is the final fused image $\mathrm{I_{ff}}$. The loss function consists of two terms:

$$G\mathcal{L}_3 = \mathcal{L}_3^{\mathrm{A}} + \mathcal{L}_3^{C} \tag{8}$$

$$\mathcal{L}_3^{\mathrm{A}} = \frac{1}{N}\sum_{n=1}^{N} (D(I_{ff}) - \ell_3)^2 \tag{9}$$

$$\mathcal{L}_3^{C} = \frac{1}{S}((\|I_{ff} - I_{IRV}\|)_F^2 + \lambda\frac{1}{N}\sum_{i=0}^{N} (I'_{ff} - I'_{IRV})^2 + \zeta\|\nabla I_{ff} - \nabla I_{VIR}\|)_F^2 \tag{10}$$

The explanation of these three formulas is the same as above. ζ is a positive parameter controlling the trade-off between two terms, which in our method is set as 100. The larger this value, the more information of the fused image is retained by the second fused image.

3. 4 The Loss Function of D

The loss function contains two terms: the deviation degree of the ground truth image from the expectation and the deviation degree of fused image from the expectation. The three D have same loss function which consists of two terms:

$$DL = \frac{1}{N}\sum_{n=1}^{N} (D(I_{ground\,truth}) - \alpha)^2 + \frac{1}{N}\sum_{n=1}^{N} (D(I_{fusion}) - \beta)^2 \tag{11}$$

where α and β denote the parameters of the fused image I_{fusion} , and the ground truth image $I_{ground\,truth}$ respectively, while $\mathrm{D(x)}$ denote the classification results of the x. We set the parameter α as 0.8, since we regard the ground truth images as the real image thus making it close to 1. On the contrary, we set the parameter β as 0.2, since we regard fused image as the fake images thus making it close to 0. This setting is to balance the loss function. While optimizing D, we try to minimize $D(I_{fusion})$, so that D is always able to distinguish the fake data from the truth. The smaller DL mean that more details of the ground truth image are transferred to the elementary fused image generated by G. The adversarial game between G and D gradually complete the fusion process so that the final fused images have comprehensive information from source images.

4. Experiments and Results

In our experiments, we adopted 40 pairs of IR and visible images from the TNO database[39] which includes visual, near-infrared, and long-wave infrared or thermal, night time imagery. All images have already been pre-aligned. We divided 40 pairs into two parts: 30 pairs for training and 10 pairs for testing. For 30 pairs of images are not enough to train a good model, we crop each image by setting the stride to 12, and each patch is of the same size 128 × 128. As a result, we combined 35,283 pairs of images together as the training data set.

4. 1 Details of training

During the training, we set iteration number k as 10, step number as 2, and cropped the training images into 128 × 128 batches without overlapping. We fed the image batches to G, which then generates fused image batches. We fed IR batches and fused image batches to D, which output the loss $\mathcal{L}_D$ and $\mathcal{L}_{G.}$ At the end of iteration, G generates the fused image. Fig. 6 shows our GAN's training process.

We first train the discriminator k times, and use the optimizer solver as that in[40]. Then we train the generator until the end of iterations.

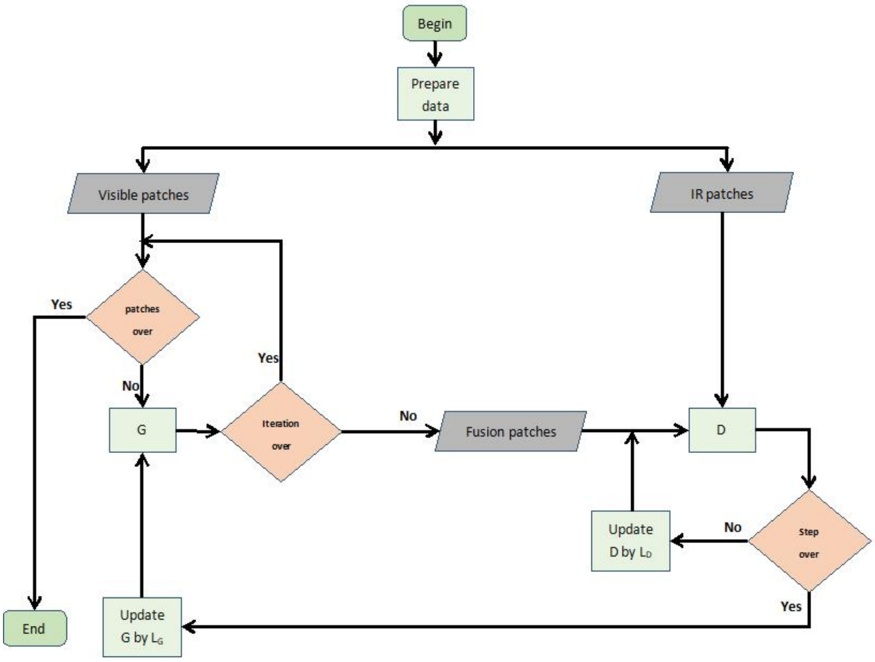


Figure 6. Training process of GAN.

In the training process, we will train three pairs of G and D. The first pair takes visible images as input and IR images as ground truth. Through the training, the network with the first pair can generate fused image that contains more thermal radiation information. The second pair takes IR images as input and visible images as ground truth. The network with the second pair, after training, can generate fused image that contains more texture details. The third pair takes the images that are generated by first network as input and the images that generates by second network as ground truth for training. After these trainings, the final fused images will then retain salient details from both IR and visible images.

4. 2 Objective Evaluation Metrics

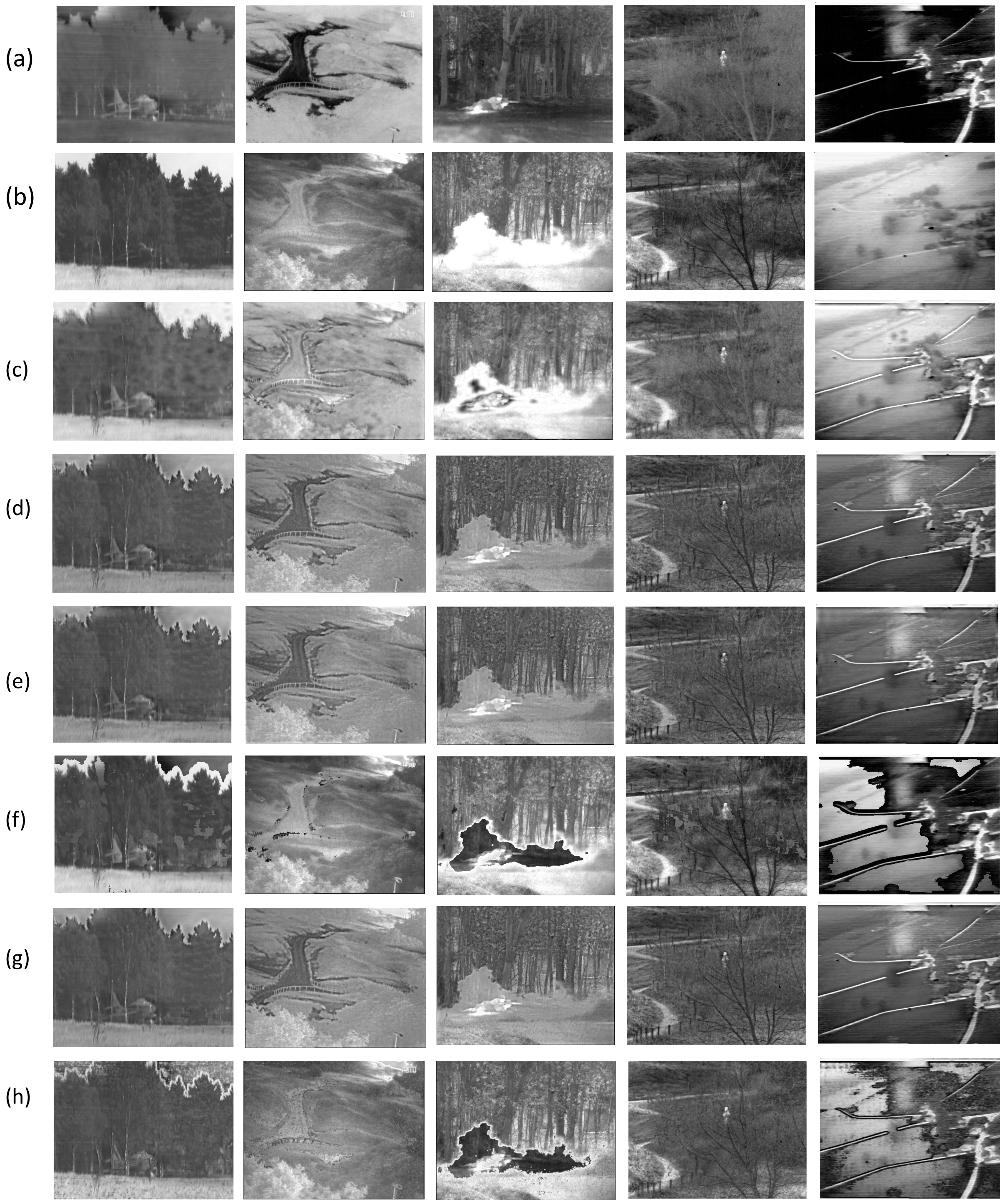
In the task of infrared and visible image fusion, there is no real ground truth, so it is difficult to conduct the objective evaluation. As a result, researchers take a reasonable way to apply several fusion metrics to make an overall evaluation[41]. In our experiments, we used eight objective fusion metrics to evaluate the fusion results. The eight metrics were entropy (EN) which measures the amount of information the fused image contains. Mutual information (MI) which is used to evaluate the mutual information of fused images. Structural similarity (SSIM) measures the mean structural similarity between the source images and fused image. Spatial frequency (SF) [42] measures the spatial frequency of the fused image. Standard deviation (SD)[43] measures the contrast of the fused image which influences the visual attention. The sum of the correlation of differences (SCD)[44] is an independent index for judging the amount of information transmitted from source images to the fused image.

The feature mutual information (FMI)[45] is based on information theory and measures the mutual information between image features. QABF[46] is a local measure used to estimate the degree of retention of significant information in fused images. For all eight metrics, the larger value means the better fusion performance.

4. 3 Subjective Evaluation

To elaborate and compare the fusion effects of different methods clearly, the fusion results obtained by different methods of ten pairs are shown in Fig. 7 and Fig. 8.

The comparing methods in our experiment include NSST-PAPCNN[47], nonsubsampled contourlet transform (NSCT)[48], curvelet transform(CVT)[49], convolutional sparse representation (CSR)[50], dual-tree complex wavelet transform (DTCWT)[51], cross bilateral filter(CBF)[52], Latent Low-Rank Representation(LATLRR)[53], weighted least square(WLS)[54], convolutional neural network based fusion (CNN)[55], and a GAN based method (FusionGAN)[25]. We used the codes provided by the authors or a well-known toolbox to generate the fused images from the source images, i.e. the IR and visible images.



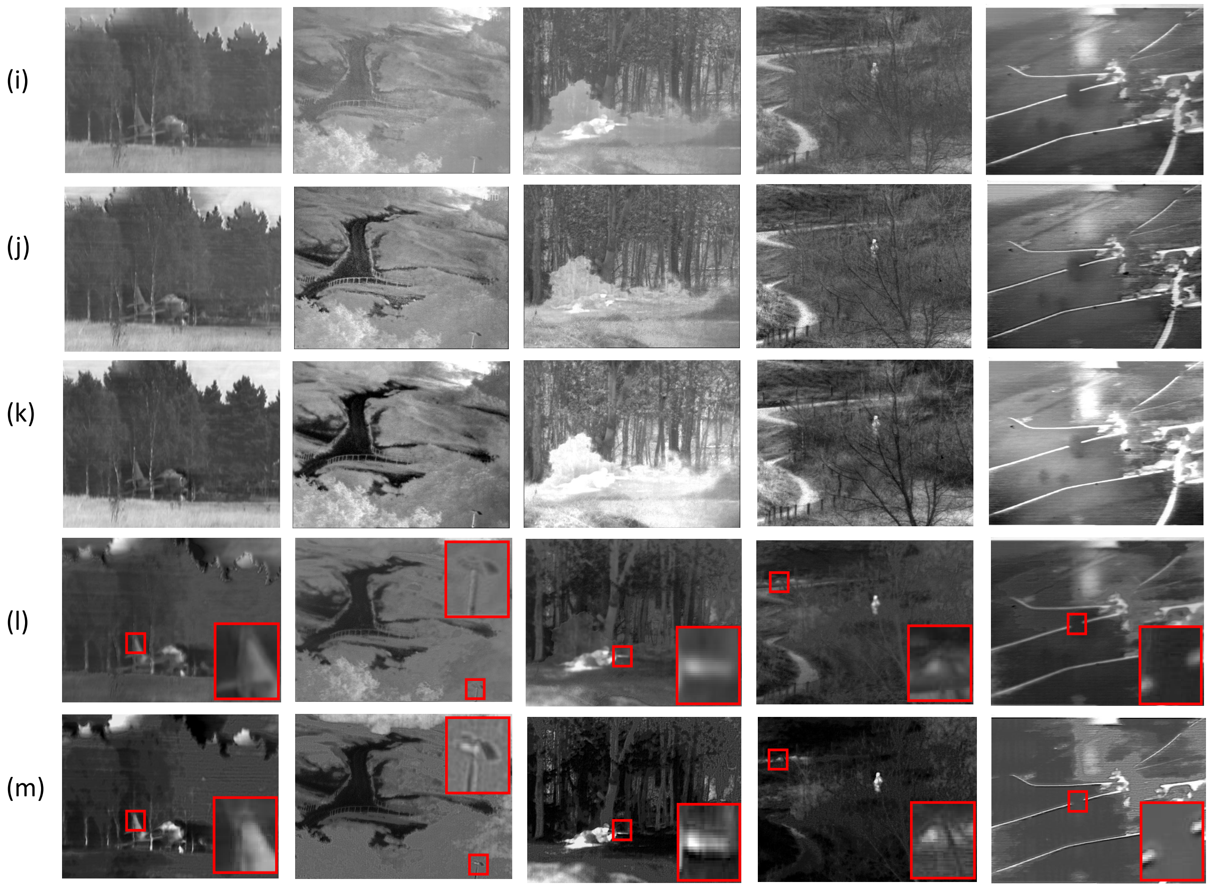
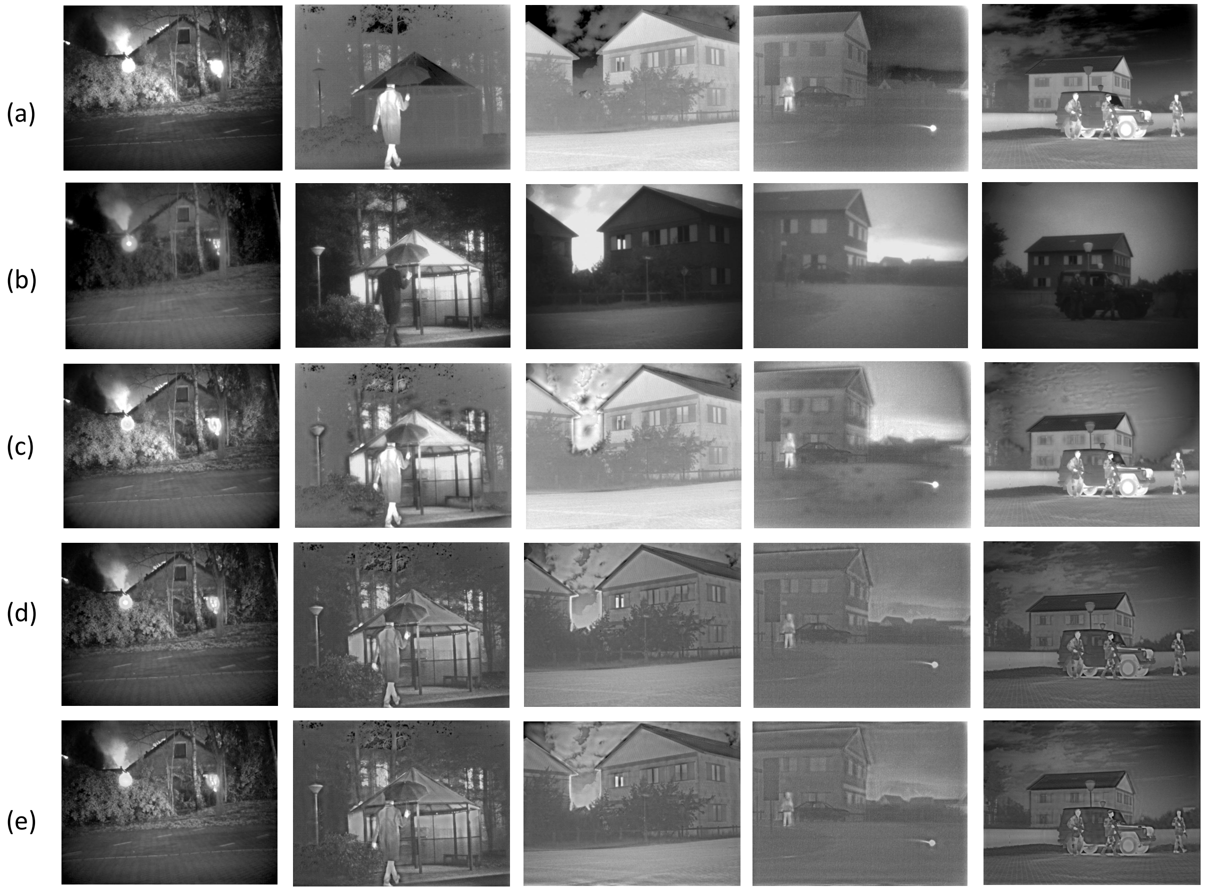


Fig. 7. The results for the first 5 groups. (a)IR image, (b)VIS Image, (c)NSST-PAPCNN, (d)NSCT, (e)CVT, (f)CSR, (g)DTCWT, (h)CBF, (i)LATLRR, (j)WLS, (k)CNN, (l)FGAN, (m)OURS. Note that in the last two rows, for clear comparison we select a small region (i.e., the red box) in each fused image, and then enlarge and put it in the right corner.



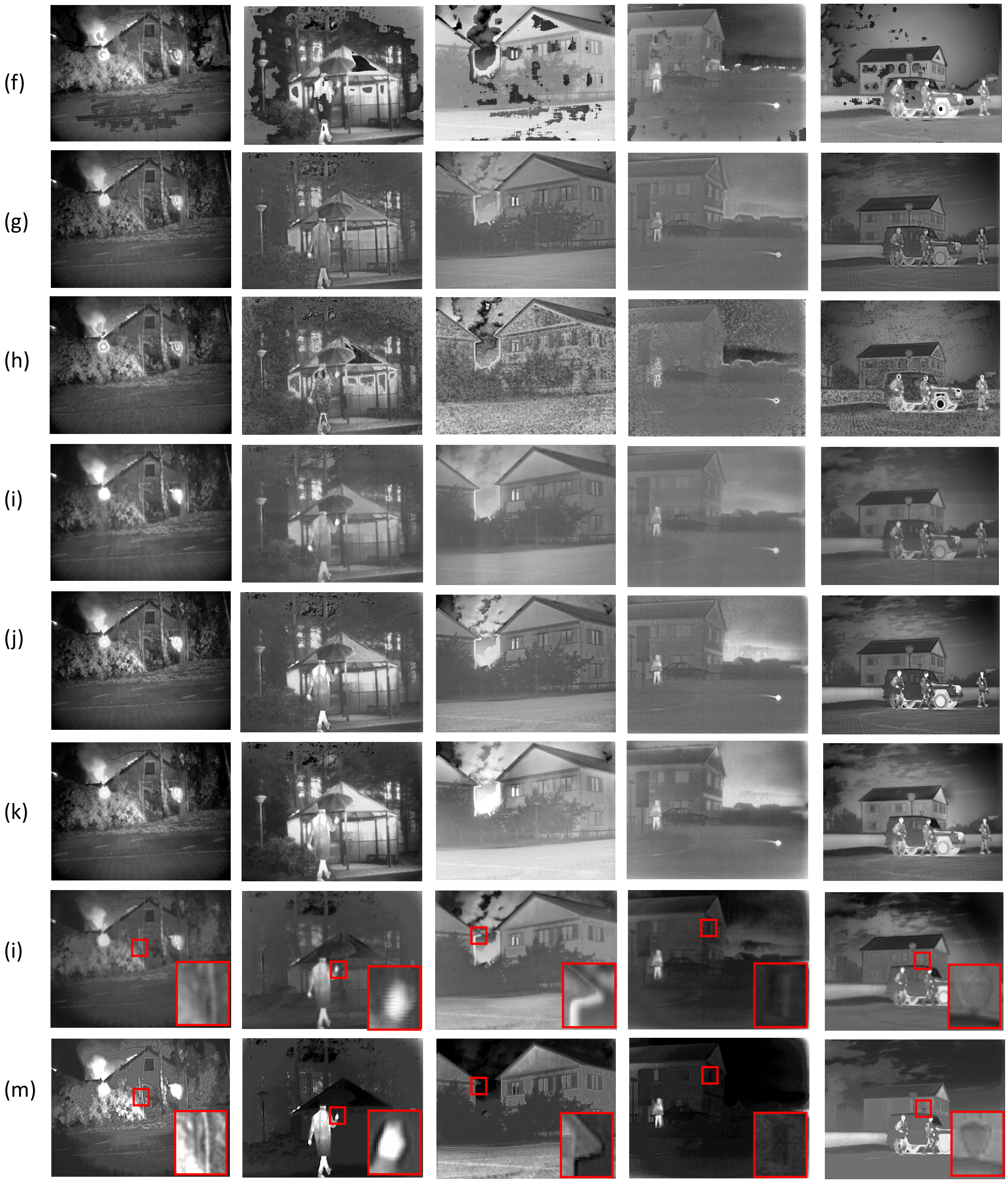


Fig. 8. The results for the last 5 groups. (a)IR image, (b)VIS Image, (c)NSST-PAPCNN, (d)NSCT, (e)CVT, (f)CSR, (g)DTCWT, (h)CBF, (i)LATLRR, (j)WLS, (k)CNN, (l)FGAN, (m)OURS. Note that in the last two rows, for clear comparison we select a small region (i.e., the red box) in each fused image, and then enlarge and put it in the right corner.

The first two rows in Fig. 7 and 8 represent IR images and visible images, and the last row is the fusion results of our method. Overall, the results show that all the methods can fuse the information of visible image and IR image well to some extent. However, it can be seen that, compared to other methods, FusionGAN and ours make the target area (such as buildings, people and cars) more prominent in the fused images, which is conducive to automatic target detection and localization. This could be attributed to the fact that FusionGAN and ours are able to preserve more thermal radiation information in the IR images, while other comparing methods focus more on exploiting the texture details in the visible images.

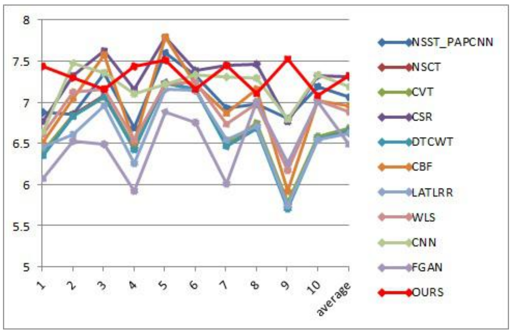
Between FusionGAN and our method, we can see that our results contain slightly more abundant details, and they are more suitable for human visual perception, as shown in the red boxes in Fig. 7 and 8. For example, the solider in the fourth column of Fig. 7 and the hand of the umbrella bearer in the second column of Fig. 8 are presented more clearly by ours than that by FusionGAN. In the second column of Fig. 8, the hand of the umbrella bearer is fused more appropriately and more clearly by ours than FusionGAN. And in the third column of Fig. 8, the corner of the roof highlighted in the box fused by FusionGAN is fuzzy, while our result is sharper. This demonstrates our method’s excellent performance in terms of simultaneously preserving thermal radiation information and texture details.

For quantitative comparisons, we evaluated all methods through the above-mentioned eight metrics. The results are plotted in Fig. 9 and the averages of 10 fused images for 8 metrics are listed in Table 1.

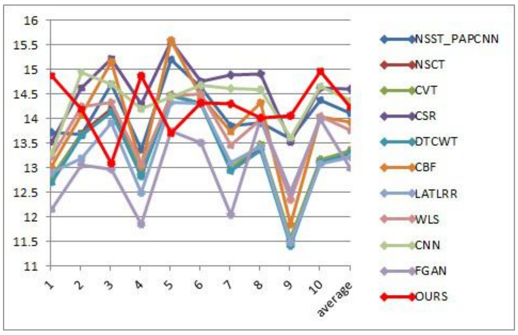
Table 1. The average values of 10 fused images for the 8 metrics.

Methods	EN	MI	QABF	FMI	SSIM	SD	SF	SCD
NSST_PAPCNN	7.062735	14.12547	0.349003	0.893138	0.6545	41.70837	0.487205	1.44104
NSCT	6.67294	13.34588	0.442946	**0.899258**	0.678554	30.47823	0.559955	1.56964
CVT	6.684161	13.36832	0.4151	0.895826	0.664334	30.46991	0.515316	1.55008
CSR	7.308307	**14.61661**	0.441369	0.867851	0.607427	49.43395	0.502694	1.06805
DTCWT	6.648422	13.29684	0.410618	0.894699	0.664047	30.05842	0.562504	1.54710
CBF	6.96852	13.93704	0.354673	0.857701	0.547844	37.91702	0.511339	1.27858
LATLRR	6.618689	13.23738	0.408749	0.891143	**0.730806**	30.45902	0.583359	1.60302
WLS	6.887689	13.77538	0.413503	0.88595	0.674388	38.43261	0.527757	**1.69113**
CNN	7.195607	14.39121	0.44262	0.89764	0.662346	48.1034	0.498179	1.63320
FusionGAN	6.49529	12.99058	0.21183	0.880279	0.639393	29.35145	0.630014	1.32043
Ours	**7.321885**	14.25377	**0.443773**	0.871807	0.726497	**49.78707**	**0.630342**	1.56132

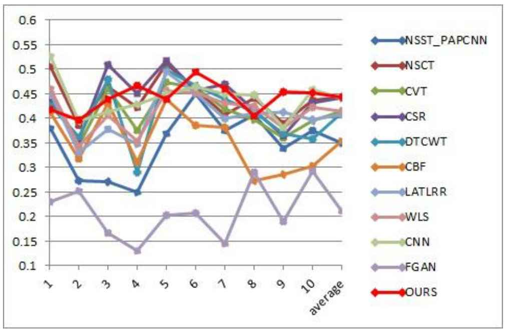
In Table 1， the best values for each metric are presented in bold face. It can be seen that our method achieves the best performance in EN, QABF, SD, and SF. For other metrics, the performance of our method is not far from the best. High EN and SD values indicate that our fused images have higher contrast, while high QABF means our fused image is superior in conspicuousness. Also, a high SF indicates the images fused by the proposed method contain more texture details. However, our method has slightly lower SCD, which will be our future work to optimize the network structure and loss function further.



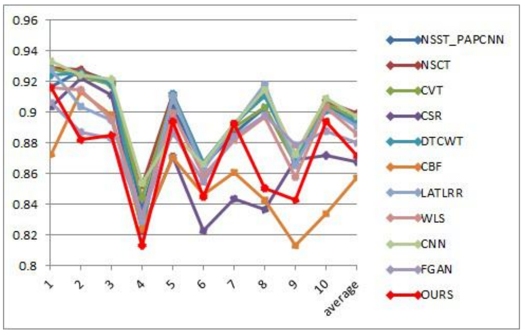
EN



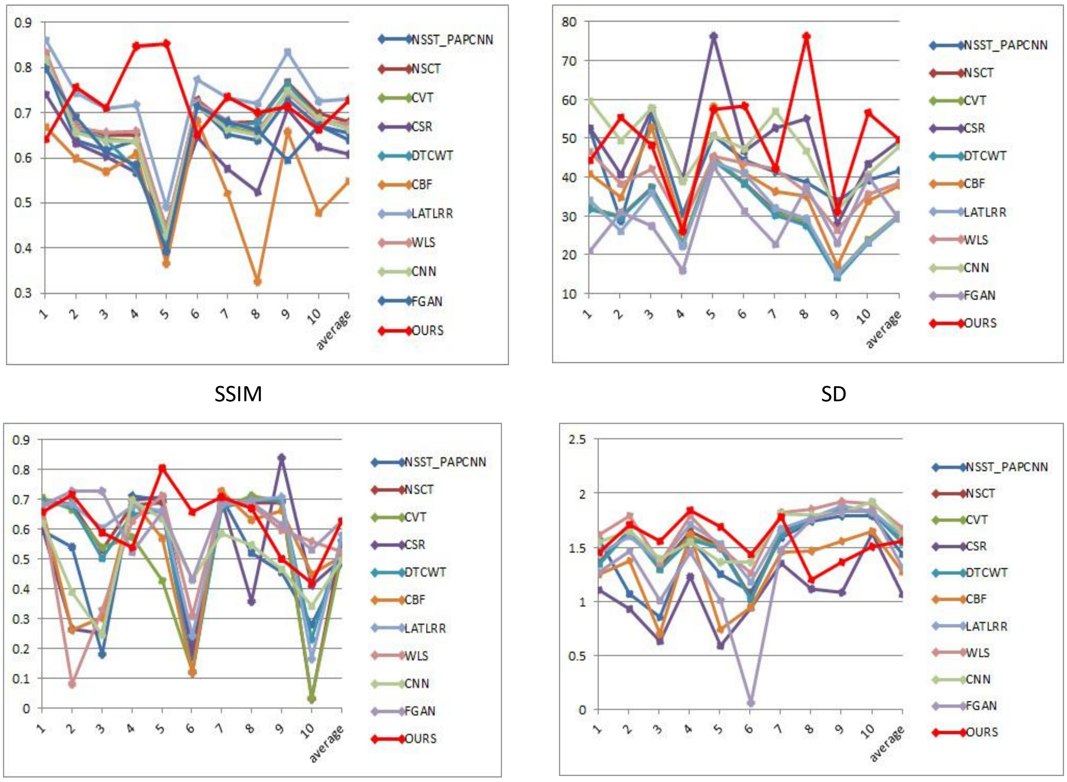
MI



QABF



FMI



SSIM

SD

SF

SCD

Figure 9. Quantitative values of eight metrics.

5. Conclusions

Inspired by FusionGAN, we propose a two-layer generative adversarial network for the fusion of infrared and visible images. The proposed network can simultaneously retain the thermal radiation information from infrared image and the texture details from visible image. Our experiments demonstrate that compared to FusionGAN and other existing approaches, our fusion results can highlight salient information in the images such as potential targets more clearly, which is important for target detection and recognition applications. The quantitative comparisons with the state-of-the-arts on eight evaluation metrics also reveal that our method can not only produce better visual effects, but also keep more details existing in the source images. In our future work, we will optimize the structure and the loss functions of our framework so that the fused images have more texture details and target radiation information.

Acknowledgment

This work was supported in part the by the National Natural Science Foundation of China under Grant 61572392 and in part by the National and Local Funds for New Networks and Measurement and Control Laboratories under Grant GSYSJ2017001.

References

[1] Zhishe Wanga, Jiawei Xub, Xiaolin Jiang,Xiaomei Yan. Infrared and image fusion via hybrid decomposition of NSCT and morphological sequential toggle operator. Optik,2020.

[2] Cheng Zhao, Yongdong Huang, Shi Qiu. Infrared and image fusion algorithm based on saliency detection and adaptive double-channel spiking cortical model. Infrared Physics & Technology, 2019.

[3] Xin Jin, Qian Jiang, Shaowen Yao,et al. Infrared and visual image fusion method based on discrete cosine transform and local spatial frequency in discrete stationary wavelet transform domain. Infrared Physics & Technology,2018,88

[4] Ma, J.; Ma, Y.; Li, C. Infrared and image fusion methods and applications: A survey. Inf. Fusion 2018,45, 153–178.

[5] Latreche B , Saadi S , Kious M , et al. A novel hybrid image fusion method based on integer lifting wavelet and discrete cosine transformer for visual sensor networks[J]. Multimedia Tools and Applications, 2019, 78(8):10865-10887.

[6] Ma, J.; Zhou, Z.; Wang, B.; Zong, H. Infrared and image fusion based on visual saliency map and weighted least square optimization. Infrared Phys. Technol. 2017, 82, 8–17.

[7] Guo, Chen, Li,et al. Weighted sparse representation multi-scale transform fusion algorithm for high dynamic range imaging with a low-light dual-channel camera.. Optics Express, 2019.

[8] B, Jun Chen A , et al. Infrared and image fusion based on target-enhanced multiscale transform decomposition. Information ences, 2020, 508:64-78.

[9] Zihui L , Yuxing W , Jianlin Z , et al. Image Fusion of Infrared and Images Based on Saliency Map. Infrared Technology, 2019.

[10] Zhai-Sheng D , Dong-Ming Z , Ren-Can N , et al. Infrared and image fusion using residual network and visual saliency detection. Journal of Yunnan University (Natural Sciences Edition), 2019.

[11] Yubin Q , Mei Y , Hao J , et al. Multi-exposure image fusion based on tensor decomposition and convolution sparse representation. Opto-Electronic Engineering, 2019.

[12] Xinxiang L I , Longbo Z , Lei W , et al. Image fusion method based on convolutional sparse representation and morphological component analysis. Intelligent Computer and Applications, 2019.

[13] Mustafa H T , Yang J , Zareapoor M . Multi-scale convolutional neural network for multi-focus image fusion. Image and Vision Computing, 2019, 85(MAY):26-35.

[14] Yang Y , Nie Z , Huang S , et al. Multi-level Features Convolutional Neural Network for Multi-focus Image Fusion. IEEE Transactions on Computational Imaging, 2019:1-1.

[15] Kong, W.; Lei, Y.; Zhao, H. Adaptive fusion method of light and infrared images based on non-subsampled shearlet transform and fast non-negative matrix factorization. Infrared Phys. Technol.2014, 67, 161–172.

[16] WANG Wei-zhe, DAI Ye-yong. Improvement of the Edge Fusion Algorithm for Subspace of Remote Sensing Image. ence & Technology of West China, 2015.

[17] Wei, Tan, Huixin,et al. Infrared and image perceptive fusion through multi-level Gaussian curvature filtering image decomposition.. Applied optics, 2019.

[18] Farahnakian F , Poikonen J , Laurinen M , et al. and Infrared Image Fusion Framework based on RetinaNet for Marine Environment. 22th International Conference on Information Fusion (FUSION). IEEE, 2020.

[19] Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-focus image fusion with a deep convolutional neural network. Inf. Fusion 2017, 36, 191–207.

[20] Liu, Y.; Chen, X.; Cheng, J.; Peng, H.; Wang, Z. Infrared and image fusion with convolutional neural networks. Int. J. Wavelets Multiresolution Inf. Process. 2018, 16, 1850018.

[21] Y. Liu, X. Chen, R.K. Ward, Z.J. Wang, Image fusion with convolutional sparse representation, IEEE Signal Process. Lett. 23 (12) (2016) 1882–1886.

[22] G. Huang, Z. Liu, L. van der Maaten, K. Q. Weinberger, Ieee, Densely connected convolutional networks, in: 30th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2261–2269.

[23] Y. Liu, X. Chen, J. Cheng, H. Peng, Z. Wang, Infrared and visible image fusion with convolutional neural networks,International Journal of Wavelets Multiresolution and Information Processing 16 (3) (2018).

[24] Hui Li, Xiao-Jun Wu, and Josef Kittler. Infrared and visible image fusion using a deep learning framework. arXiv preprint arXiv:1804.06992, 2018.

[25] Jiayi, M.; Wei, Y.; Pengwei, L.; Chang, L.; Junjun, J. FusionGAN: A generative adversarial network for infrared and image fusion. Inf. Fusion 2018, 48, 11–26.

[26] Liu Y , Chen X , Wang Z , et al. Deep learning for pixel-level image fusion: Recent advances and future prospects. Information Fusion, 2018, 42:158-173.

[27] S. Li, H. Yin, L. Fang, Group-sparse representation with dictionary learning for medical image denoising and fusion, IEEE Trans. Biomed. Eng. 59 (12) (2012) 3450–3459.

[28] Li H , Ma K , Yong H , et al. Fast Multi-Scale Structural Patch Decomposition for Multi-Exposure Image Fusion. IEEE Transactions on Image Processing, 2020, PP (99):1-1.

[29] S. Li, B. Yang, J. Hu, Performance comparison of different multi-resolution transforms for image fusion, Inf. Fusion 12 (2) (2011) 74–84.

[30] J. Ma, Z. Zhou, B. Wang, H. Zong, Infrared and visible image fusion based on visual saliency map and weighted least square optimization, Infrared Phys. Technol. 82(2017) 8 – 17.

[31] Rout M , Nahak S , Priyadarshinee S , et al. A Deep Learning Approach for SAR Image Fusion. 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT). IEEE, 2020.

[32] T. Xiang, L. Yan, R. Gao, A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking pcnn in nsct domain, Infrared Phys. Technol.69 (2015) 53 – 61.

[33] Prabhakar K R , Babu R V . GHOSTING-FREE MULTI-EXPOSURE IMAGE FUSION IN GRADIENT DOMAIN.IEEE International Conference on Acoustics. IEEE, 2016.

[34] Y. Liu, X. Chen, J. Cheng, H. Peng, Z. Wang, Infrared and visible image fusion with convolutional neural networks, Int. J. Wavelets Multiresolution Inf. Process. 16 (3) (2018) 1850018.

[35] Li, H.; Wu, X.-J. DenseFuse: A Fusion Approach to Infrared and Images. ITIP 2019, 28, 2614–2623.

[36] Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Bing, X.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y.Generative Adversarial Nets. Proc. 27th Int. Conf. Neural Inf. Process. Syst. 2014, 2, 2672–2680.

[37] F. Yu, V. Koltun, Multi-scale context aggregation by dilated convolutions, arXiv:1511.07122v1 (2015).

[38] A. Radford, L. Metz, S. Chintala, Unsupervised representation learning with deep convolutional generative adversarial networks, arXiv:1511.06434v1 (2015).

[39] A. Toet, E.M. Franken, Perceptual evaluation of different image fusion schemes, Displays 24 (1) (2003) 25–37 https://figshare.com/articles/TNO_Image_Fusion_Dataset/ 1008029.

[40] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv:1412.6980v1(2014).

[41] Liu, Z.; Blasch, E.; Xue, Z.; Zhao, J.; Laganiere, R.; Wu, W. Objective Assessment of Multiresolution Image Fusion Algorithms for Context Enhancement in Night Vision: A Comparative Study. IEEE Trans. Pattern Anal. Mach. Intell. 2011, 34, 94–109.

[42] Eskicioglu, A.M.; Fisher, P.S. Image quality measures and their performance. IEEE Trans. Commun. 1995, 43,2959–2965.

[43] Xing Su-xia, Chen Tian-hua, Li Jing-xian. Image fusion based on regional energy and standard deviation. International Conference on Signal Processing Systems. IEEE, 2010.

[44] Aslantas, V.; Bendes, E. A new image quality metric for image fusion: The sum of the correlations of differences. AEU Int. J. Electron. Commun. 2015, 69, 1890–1896.

[45] Haghighat, M.; Razian, A. Masoud Fast-FMI: Non-reference image fusion metric. In Proceedings of the 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT),Astana, Kazakhstan, 15–17 Octorber 2014; pp. 1–3.

[46] Piella G , Heijmans H . A new quality metric for image fusion[C]. International Conference on Image Processing. IEEE, 2003.

[47] Ming Yin, Xiaoning Liu, Yu Liu*, Xun Chen "Medical Image Fusion With Parameter-Adaptive Pulse Coupled Neural Network in Nonsubsampled Shearlet Transform Domain", IEEE Transactions on Instrumentation and Measurement, in press, 2018.

[48] Zhang, Q.; Guo, B.-l. Multifocus image fusion using the nonsubsampled contourlet transform. SIGPR 2009,89, 1334–1346.

[49] Nencini, F.; Garzelli, A.; Baronti, S.; Alparone, L. Remote sensing image fusion using the curvelet transform. Inf. Fusion 2007, 8, 143–156.

[50] Yu Liu, Xun Chen, Rabab Ward, Z.Jane Wang "Image fusion with convolutional sparse representation", IEEE SIGNAL PROCESSING LETTERS, vol. 23, no. 12, pp. 1882-1886, 2016.

[51] Lewis, J.J.; O'Callaghan, R.J.; Nikolov, S.G.; Bull, D.R.; Canagarajah, N. Pixel-and region-based image fusion with complex wavelets. Inf. Fusion 2007, 8, 119–130.

[52] B. K. Shreyamsha Kumar,Image fusion based on pixel significance using cross bilateral filter Signal, Image and Video Processing, pp. 1-12, 2013. (doi:10.1007/s11760-013-0556-9)

[53] Li H , Wu X J , Kittler J . Infrared and image fusion using a novel deep decomposition method. 2018.

[54] Ma J, Zhou Z, Wang B, et al. Infrared and image fusion based on visual saliency map and weighted least square optimization. Infrared Physics & Technology, 2017, 82:8-17

[55] Yu Liu, Xun Chen, Juan Cheng, Hu Peng, Zengfu Wang, "Infrared and image fusion with convolutional neural networks", International Journal of Wavelets, Multiresolution and Information Processing, vol. 16, no. 3, pp. 1850018: 1-20, 2018.