Article

LPGAN: A LBP-based Proportional Input Generative Adversarial Network for Image Fusion

Dongxu Yang[1], Yongbin Zheng [1,*], Wanying Xu [1], Peng Sun [1], Di Zhu [1]

[1] College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, China; dxyang2021@outlook.com(D.X. Yang), zybnudt@nudt.edu.cn (Y.Z.); wy.xu@163.com (W.X.); sunpeng@nudt.edu.cn (P.S.); zhudi@nudt.edu.cn(D. Zhu)

* Correspondence: zybnudt@nudt.edu.cn;

Abstract: Image fusion is the process of combining multiple input images from single or multiple imaging modalities into a fused image, which is expected to be more informative for human or machine perception as compared to any of the input images. In this paper, we propose a novel method based on deep learning for fusing infrared images and visible images, named the LBP-based proportional input generative adversarial network (LPGAN). In the image fusion task, the preservation of structural similarity and image gradient information is contradictory, and it is difficult for both to achieve good performance at the same time. To solve this problem, we innovatively introduce Local Binary Patterns (LBP) into Generative Adversarial Networks (GANs), which effectively utilize the texture features of the source images, so that the network has stronger feature extraction ability and anti-interference ability. In the feature extraction stage, we introduce a pseudo-siamese network for the generator to extract the detailed features and the contrast features. At the same time, considering the characteristic distribution of different modal images, we propose a 1:4 scale input mode. Extensive experiments on the publicly available TNO dataset and CVC14 dataset show that the proposed method achieves the state-of-the-art performance. We also test the universality of LPGAN through the fusion of RGB and infrared images on the RoadScene dataset. In addition, LPGAN is applied to multi-spectral remote sensing image fusion. Both qualitative and quantitative experiments demonstrate that our LPGAN can not only achieve good structural similarity, but also retain rich detailed information.

Keywords: Image fusion; generative adversarial network (GAN); local binary patterns (LBP); multi-modal images.

1. Introduction

Image fusion aims to merge or combine images captured with different sensors or camera settings to generate a greater quality composite image [1]. It is crucial for many applications in image processing[2,3], computer vision [4,5], remote sensing [6], and medical image analysis [7].

In image processing field, visible images and infrared images are generated by sensors with different sensitivities to light in different wavelength bands. The images of different bands contain different information. Each kind of image can only focus on a given operating range and environmental condition, and it is difficult to receive all the necessary information for object detection or the object or scene classification [8]. Due to the strong complementarity between them, it is a feasible way to improve visual understanding by fusing them. The fused image can combine the characteristics of different modal images to generate an image with rich details and significant contrast, an example is shown in Fig.1. Therefore, for the full exploitation of multi-modal data, advanced image fusion has been developed rapidly in the last few years.

The key to multi-modal image fusion is effective image information extraction and appropriate fusion principles [9]. Traditional research has focused on multi-scale

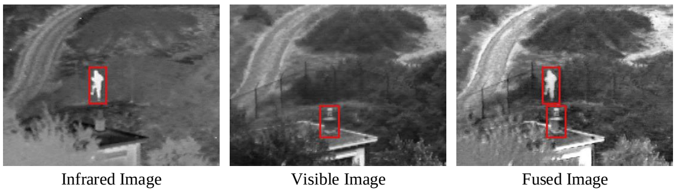


Figure 1. An example of image fusion.

transform-based methods [10], [11], sparse representation-based methods [12], [13], subspace-based methods [14], saliency-based methods [15], hybrid methods [16,17] and other fusion methods [18], [19]. However, the performance improvement of artificially designed feature extraction and fusion rules is limited.

With the widespread application of deep learning, deep learning-based fusion methods have achieved rapid progress, showing advantages over conventional methods and leading to state-of-the-art results [20,21]. Although these algorithms have achieved positive results under most conditions, there are still some shortcomings that need to be improved:

- The extraction of source image feature information is incomplete. Most image fusion algorithms cannot achieve good structural similarity and retain rich detailed features at the same time due to the incomplete extraction of feature information [22], [23], [24].
- The mission objectives and network structure do not match. The same network is employed to extract features while ignoring the feature distribution characteristics of different modal images, resulting in the loss of meaningful information [25], [24].
- Improper loss function leads to missing features. In the previous methods [22], [23], only the gradient is used as a loss to supervise the extraction of detailed features and texture features, which makes the network feature extraction incomplete.

To overcome the above challenges, a completely deep learning-based method, LPGAN, is proposed to achieve image fusion. Firstly, we innovatively introduce local binary patterns (LBP) into the network to effectively extract the texture features of the source image. The LBP [26] can accurately describe the local texture features of the image, and because of its strong robustness to illumination, the introduction of LBP not only enables the network to have stronger feature extraction capabilities but also strong anti-interference capabilities. Sencondly, we introduce a pseudo-siamese network for the generator to extract the detailed features and contrast features in the feature extraction stage. It is worth mentioning that, the improved generator inputs infrared and visible images at different ratios. In some works [27], [28], the source image is concatenated in a 1:2 ratio as the input of the network, but we found through experiments that a ratio of 1:4 can achieve better results. Extensive experiments on the publicly available TNO dataset and CVC14 dataset [29] show that the proposed method achieves the state-of-the-art performance. We also test the universality of LPGAN through the fusion of RGB and infrared images on the RoadScene dataset [25]. Finally, the proposed method can also be applied to multi-spectral remote sensing image fusion, and the expansion experiment reveals the advantages of our LPGAN compared to other methods.

The contributions of our work are as three-fold:

1) We introduce LBP into the network and design a new content loss for the generator, which enables the model to extract features completely and reduce image distortion.
2) Taking into account the characteristics of the information distribution of the source images, we propose a 1:4 scale input in the feature extraction stage.
3) We design a pseudo-siamese network to extract feature information from source images. It fully considers the differences in the imaging mechanism and image features of the source images, encouraging the generator to preserve more features in source images.

We organized the remainder of this paper as follows. We first briefly review some work related to our method in Section II. Then, we provide the overall framework, network architecture and loss function of the proposed LPGAN in Section III. In Section IV, we show and analyse the experimental results of the proposed method and the competitors. In Section V, we give a discussion about our method. Finally, we make a conclusion of this paper in Section VI.

2. Related Work

Image fusion tasks can be divided into five types: infrared-visible image fusion, multi-focus image fusion, multi-exposure image fusion, remote sensing image fusion and medical image fusion. In this section, we briefly introduce several existing deep learning-based image fusion methods and some basic theories of cGAN and LBP.

2.1. Deep Learning-Based Image Fusion

Many deep learning based image fusion methods have been proposed in the last five years and have achieved promising performances. In some methods, the framework of deep learning is combined with traditional methods to solve image fusion tasks. Representatively, Liu *et al.* [30] proposed a fusion method based on convolutional sparse representation (CSR). The method employs CSR to extract multi-layer features and uses the features to reconstruct the image. Later they proposed a convolutional neural network (CNN)-based method for multi-focus image fusion tasks [31]. They use image patches containing different features as input to train the network and obtain a decision map, and then directly use the map to guide the image fusion. Deep learning is also used by some algorithms to diversify the extraction of image features. In [32], a model based on the multi-layer fusion strategy of the VGG-19 model was proposed. The method decomposes the source image into two parts: one part contains the low-frequency information of the image and the other part contains the high-frequency information with detailed features. This strategy can retain the deep features of the detailed information. In addition, to make the generated images more realistic, PSGAN [33] handles the remote sensing image fusion task by using a GAN to fit the distribution of high-resolution multi-spectral images. However, it still requires artificially constructing the ground-truth to train the model.

The above methods only apply the deep learning framework in some parts of the fusion process. In other methods, the entire image fusion process uses a deep learning framework. For instance, Prabhakar *et al.* [34] proposed an unsupervised model for multi-exposure image fusion named DeepFuse in 2017. The model consists of an encoder, a fusion layer and a decoder. The parameter sharing strategy is adopted to ensure that the feature types extracted from the source images are the same, which facilitates subsequent fusion operations and reduces the parameters of the model. Based on DeepFuse, Li *et al.* [24] improved the method by applying dense blocks and proposed a new image fusion method called DenseFuse. They utilize no-reference metrics as the loss function to train the network and achieve a high-quality performance. Since image fusion tasks usually lack ground-truth and are generative tasks, Ma *et al.* [22] proposed a GAN-based method to fuse infrared and visual images. The network uses a generator to fuse images and a discriminator to distinguish the generated image from the visible image, achieving a state-of-the-art performance. However, it is easy to lose infrared image information. To avoid the above problem, Ma *et al.* [23] used dual discriminators to encourage the generator, and the method achieved a better performance. Analogously, Ma *et al.* [35] applied a dual-discriminator architecture in remote sensing image fusion and proposed an unsupervised method based on GAN, termed PanGAN. The method establishes adversarial games to preserve the rich spectral information of multi-spectral images and the spatial information of panchromatic images. By considering the different characteristics of different image fusion tasks, Xu *et al.* [25] performed continual learning to solve multiple fusion tasks for the first time and proposed a unified unsupervised image fusion network named U2Fusion, which could be applied to a variety of image fusion tasks, including multi-modal, multi-exposure, and multi-focus cases.

However, the abovementioned works still have three drawbacks:

- Due to the lack of ground-truth, the existing methods usually supervise the work of the model by adopting no-reference metrics as the loss function. However, only the gradient is used as the loss to supervise the extraction of the detailed features, and the texture information is always ignored.
- They ignore the information distribution of the source images, *i.e.*, the visible image has more detailed information and the infrared image has more contrast information.
- These methods all use only one network to extract features from infrared images and visible images, ignoring the difference in imaging mechanisms between these two kinds of images.

To address these problems, a new content loss function is designed using LBP to effectively utilize the texture information of the source images, which also improves the anti-interference ability of our method. Then, according to the distribution characteristics of the feature information of the source images, we concatenate the source images in fixed proportions in the feature extraction stage. Finally, considering the characteristics of infrared image and visible image, we design a pseudo-siamese network to extract detailed features and contrast features respectively.

2.2. Generative Adversarial Networks

GAN is a framework for unsupervised distribution estimation via an adversarial process, proposed by Goodfellow [36] in 2014. The GAN simultaneously trains two models: a generative model G that captures the data distribution and a discriminative model D that estimates the probability that a sample comes from the training data rather than G. The GAN establishes an adversarial game between a discriminator and a generator, the generator tries to continuously generate new samples to fool the discriminator, and the discriminator aims to judge whether a sample is real or fake. Finally, the discriminator can no longer distinguish the generated sample. Assuming that the real data obey the specific distribution P_{data}, the generator is dedicated to estimating the distribution of real data and producing the fake distribution P_G that approaches the real distribution P_{data}. D and G play the following two-player minimax game with value function $V(G, D)$:

$$\begin{aligned}\min_G \max_D V_{GAN}(G,D) = {} & \mathbb{E}_{x\sim P_{data}}[\log D(x)] \\ & + \mathbb{E}_{x\sim P_G}[\log(1-D(x))],\end{aligned} \tag{1}$$

$\mathbb{E}$ is the average operation. Due to the adversarial relationship, the generator and the discriminator promote each other in continuous iterative training, and the capabilities of the two are continuously improved. The sample distribution generated by the generator approaches the distribution of the real data. When the similarity between the two is high enough, the discriminator cannot distinguish between the real data and the fake data, and the training of the generator is successful.

GANs can be extended to a cGAN if we add some extra information that could be any kind of auxiliary information as a part of the input. We can perform conditioning by feeding the extra information as an additional input layer, and this model is defined as a cGAN [37]. The formulation between G and D of cGAN is as follows:

$$\begin{aligned}\min_G \max_D V_{GAN}(G,D) = {} & \mathbb{E}_{x\sim P_{data}}[\log D(x|y)] \\ & + \mathbb{E}_{x\sim P_G}[\log(1-D(x|y))].\end{aligned} \tag{2}$$

Standard GANs consist of a single generator and only one discriminator. In order to generate higher quality samples in fewer iterations, Durugkar *et al.* [38] proposed the

Generative Multi-Adversarial Network (GMAN), a framework that extends GANs to multiple discriminators. Inspired by GMAN, the structure of multi-adversarial network is applied for dealing with different tasks, such as PS2MAN [39] and FakeGAN [40], *etc.* PS2MAN considered the photo-sketch synthesis task as an image-to-image translation problem and explored the multi-adversarial network (MAN) to generate high-quality realistic images. Novelly, FakeGAN first adopted GAN for a text classification task. The network provided the generator with two discriminators, which avoided the mod collapse issue and provided the network with high stability. One discriminator is trained to guide the generator to produce samples similar to deceptive views, and the other one aims to distinguish deceptive views from data.

Image fusion is a generating task that integrates two images of different characteristics. GAN is a network suitable for unsupervised generative tasks. Therefore, we adopt GAN as the framework of our method. To preserve detailed information and contrast information of two source images completely, we employ dual discriminators to improve the quality of our fusion results.

2.3. Local Binary Patterns

LBP is an operator used to describe the local texture features of images, which is gray scale invariant and can be easily calculated by comparing the center value with its 3×3 neighbors [26]. Although the original LBP can effectively extract the texture features of the image and has strong robustness to illumination, it cannot cope with the scaling and rotation of the image. To address this problem, Ojala *et al.* [41] in 2002 proposed an improved LBP with scale invariance and rotation invariance. The improved LBP compares the center pixel with pixels on a fixed radius, which changes as the image is scaled, thus achieving scale invariance of LBP features. At the same time, the minimum value of the encoded binary number is taken to achieve the rotation invariance of the LBP feature. LBP is used in many fields of machine learning. In [42], Zhao *et al.* applied it to recognize the dynamic textures and extend their approach to deal with specific dynamic events such as facial expression recognition. Maturana *et al.* [43] also proposed a LBP-based face recognition algorithm. LBP has also been applied in the field of gender recognition and was once the most effective method in this field. Tapia *et al.* [44] extract the iris features of the human eyes through the effective texture feature extraction capability of LBP, and perform gender recognition based on the extracted features, achieving the state-of-the-art performance of gender recognition at that time.

Although LBP was widely used in the past, it has rarely been mentioned in recent years. Since image fusion has high requirements on the detailed features and structural similarity of the fused image, the existing algorithms cannot achieve the above two points simultaneously. We believe that LBP can help the network to extract lower-level detailed features while keeping the fused image with a high structural similarity to the source images. Because the source images for image fusion are highly registered, there is no need to consider the rotation and scaling of the image, so we use the original LBP to extract the texture features of the source images.

3. Proposed Method

In this section,we introduce the proposed LPGAN in detail. Firstly, we describe the overall framework of LPGAN, and then we provide the network architectures of the generator and the discriminators. Finally, the loss function is designed.

3.1. Overall Framework

The overall framework of the proposed LPGAN is sketched in Fig. 2. It is a dual-discriminator cGAN. Visible images have rich detailed information that is saved through gradient and texture, and infrared images save significant contrast information through pixel intensity. The goal of the infrared-visible image fusion task is to generate a new image with rich detailed information and significant contrast information, which is es-

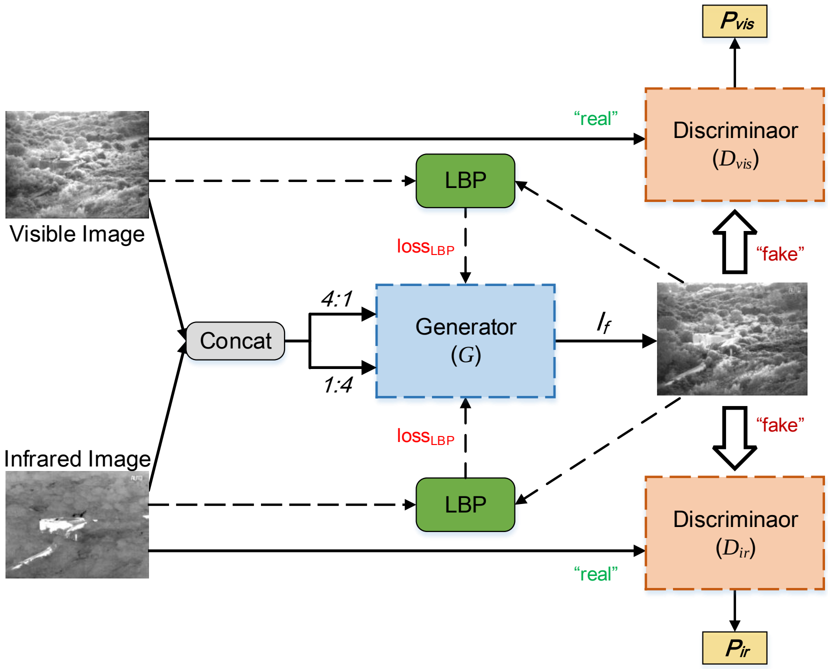


Figure 2. Overall fusion framework of our LPGAN.

sentially an unsupervised generation task. GANs have significant advantages in dealing with such problems, so this research chooses GAN network as the basic framework. Given a visible image I_{vis} and an infrared image I_{ir}, the goal of GAN applied to image fusion is to train a generator G to produce a fused image I_f , and then I_f is realistic enough to fool the discriminator. Due to the differences between infrared and visible images, we introduce a GAN with two discriminators to ensure that the feature extraction of infrared images and visible images is complete.

We design a pseudo-siamese network to extract the detailed features and contrast features. Considering that visible images still have some contrast information, infrared images also have some detailed information. We design different input ratios for different encoders to concatenate the visible image I_{vis} and infrared image I_{ir} in the channel dimension. Specifically, the ratio of the detailed feature extraction path and the contrast feature extraction path is set to 4:1 ($I_{vis} : I_{ir}$) and 1:4 ($I_{vis} : I_{ir}$), respectively. Then, the concatenated images are fed into the generator G, and the output of G is a fused image I_f. After that, LBP distributions of source images and I_f are calculated and the loss of LBP used to supervise G to extract texture features is obtained. Simultaneously, we design two adversarial discriminators, D_{vis} and D_{ir}. D_{vis} and D_{ir} generate scalars based on the input image to distinguish the generated image and real data. D_{vis} is trained to generate the probability that the image is a real visible image, while D_{ir} is trained to estimate the probability of the image belonging to the real infrared images.

3.2. Network Architecture

3.2.1. Generator Architecture

The generator consists of a feature extraction network and a feature reconstruction network, as shown in Fig. 3. The feature extraction network takes the form of a pseudo-siamese network, which is divided into gradient path and intensity path for information extraction. The process of feature reconstruction is performed in a decoder, and the output is the fused image, which has the same resolution as the source images.

In the feature extraction stage, we propose a pseudo-siamese network with two encoders. Inspired by DenseNet [45], to mitigate the vanish of gradient, remedy feature loss and reuse previously computed features, both encoders are densely connected and have the same network structure, but the parameters of them are different. In each

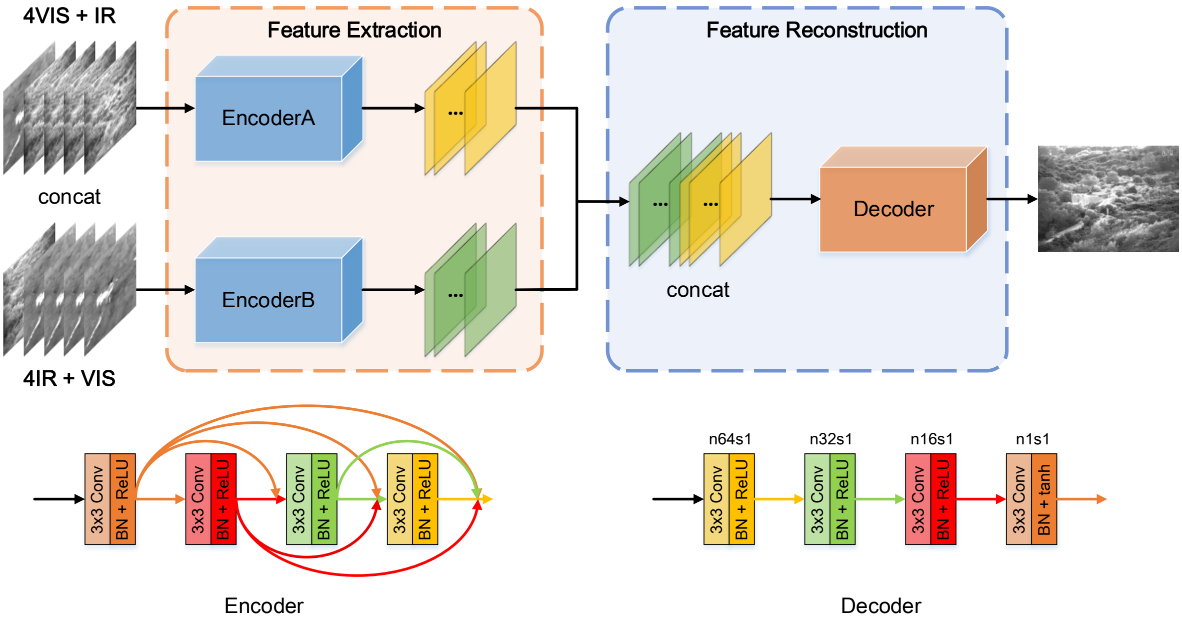


Figure 3. Network architecture of the generator.

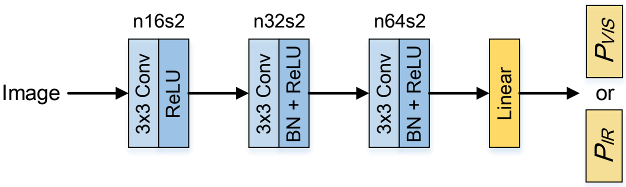


Figure 4. Network architecture of the discriminator.

path of feature extraction, the encoder consists of 4 convolutional layers. The 3 × 3 convolutional kernel is adopted in each layer, and all strides are set to 1 with a batch normalization (BN) and a ReLU activation function to speed up the convergence and avoid gradient sparsity [46]. To fully extract the information, we concatenate 4 visible images and 1 infrared image as input in the gradient path, as well as 4 infrared images and 1 visible image as input in the intensity path. After that, the outputs of the two paths are concatenated in the channel dimension. The final fusion result is generated by a decoder. The decoder is a 4-layer CNN, and the parameter settings of each layer are shown in the bottom right sub-figure of Fig. 3.

3.2.2. Discriminator Architecture

The architecture of discriminator D_{vis} and discriminator D_{ir} adopt the same structure. The discriminator is a simple 4-layer convolution neural network, which is shown in Fig. 4. In the first 3 layers, the 3 × 3 filter is adopted in each convolution layer, and the stride is set to 2. BN and ReLU activation function are followed in each convolution layer. In the last layer, the full connection and the tanh activation function are employed to generate the probability of the input image belonging to the real data.

3.3. Loss Function

We adopt two types of loss, loss $\mathcal{L}_G$ and loss $\mathcal{L}_D$, to guide the parameter optimization of G and D.

3.3.1. Loss Function of Generator

The loss function of G consists of two parts, *i.e.*, the content loss $\mathcal{L}_{con}$ and the adversarial loss $\mathcal{L}_{adv}$:

$$\mathcal{L}_G = \mu\mathcal{L}_{con} + \mathcal{L}_{adv}, \tag{3}$$

where $\mathcal{L}_G$ is the total loss and μ is a parameter to strike a balance between $\mathcal{L}_{con}$ and $\mathcal{L}_{adv}$. As the thermal radiation and texture details are mainly characterized by pixel intensities and gradient variation [18], we design four loss functions to guide G to preserve the gradient and texture information contained in the visible image and contrast information of the infrared image and reduce image distortion. We employ the L1 norm to constrain the fused image to retain similar gradient variation with the visible image. The calculation of gradient loss is as follows:

$$\mathcal{L}_{gradient} = \left\| \nabla I_f - \nabla I_{vis} \right\|_1, \tag{4}$$

where ∇ is the unification of the horizontal and vertical gradients of the image. It is calculated as:

$$\nabla = \sqrt{[I(i+1,j) - I(i-1,j)]^2 + [I(i,j+1) - I(i,j-1)]^2}\ , \tag{5}$$

where $I(i,j)$ represents the pixel value of the image at (i,j). Contrast information is mainly saved by the pixel intensities of the image. Therefore, the *Frobenius* norm is applied to encourage the fused image to exhibit pixel intensities similar to those of the infrared image, and the contrast loss is calculated as:

$$\mathcal{L}_{intensity} = \frac{1}{WH}\left\| I_f - I_{ir} \right\|_F^2, \tag{6}$$

where W and H are the width and height of the image.

To prevent image distortion, we use a structural similarity loss $\mathcal{L}_{SSIM}$ to constrain the fusion of the generator. The loss $\mathcal{L}_{SSIM}$ is obtained by this equation:

$$\mathcal{L}_{SSIM} = \tfrac{1}{2}[(1 - SSIM(I_f, I_{vis})) + (1 - SSIM(I_f, I_{ir}))]\ , \tag{7}$$

where $SSIM(\cdot)$ reflects the structural similarity of two images [47]. Visible images have rich detailed and texture information, which greatly improves the efficiency of tasks, such as target detection. In the previous algorithms, only the gradient is used as a loss to supervise the extraction of texture features. In this paper, we innovatively introduce LBP into the loss function to improve the extraction of texture features.

The formulation of $\mathcal{L}_{LBP}$ is shown as follows:

$$\mathcal{L}_{LBP} = \frac{1}{L}(\left\| LBP(I_f) - LBP(I_{vis}) \right\|_1, \tag{8}$$

where $LBP(\cdot)$ represents the operation of calculating the LBP features of the image, and L is the feature vector length of $LBP(\cdot)$. The calculation of $LBP(\cdot)$ is defined as follows:

$$LBP(I) = concat[lbp(cell_1), ..., lbp(cell_{36})], \tag{9}$$

where $cell_i$ is a 21×21 image patch, and there are 16 cells in image I, as shown in Fig. 5. $lbp(cell_i)$ aims to calculate the LBP feature of $cell_i$. To obtain $lbp(cell_i)$, we first calculate the LBP value of each pixel in $cell_i$ according to the method proposed in [26]. After calculating the LBP value of each pixel, a 256-dimensional vector is used to represent the LBP feature of the cell, which is the final result of $lbp(\cdot)$. As shown in Fig. 5, the result of $LBP(\cdot)$ is the concatenation of $lbp(cell_i)$, $(i = 1, ..., 16)$, and L in Eq. 9 is $4,096$ here.

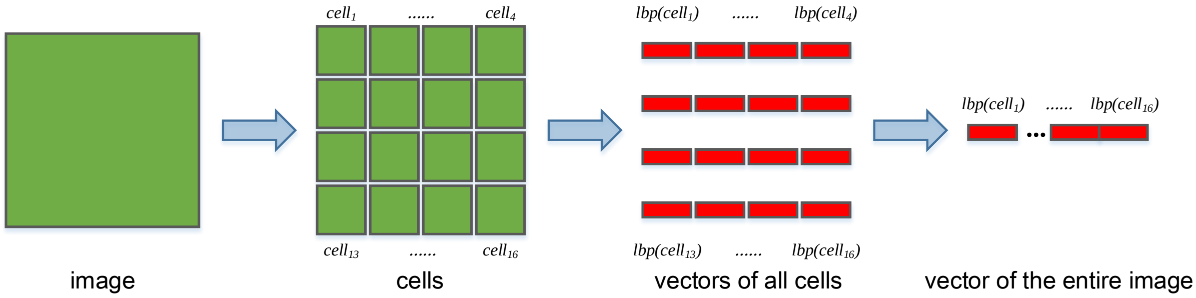

cell1 ...... cell4
cell13 ...... cell16
lbp(cell1) ...... lbp(cell4)
lbp(cell13) ...... lbp(cell16)
lbp(cell1) ...... lbp(cell16)
image
cells
vectors of all cells
vector of the entire image

Figure 5. The calculation process of the LBP distribution of the image. The image is divided into 16 cells, and the LBP distribution is calculated for each cell separately; sixteen 256-dimensional vectors are obtained. Finally, all of the vectors are concatenated to obtain a 4,096-dimensional vector, which is the LBP distribution of the entire image.

To summarize, the proposed $\mathcal{L}_{con}$ consists of 4 parts as shown in Eq. 10:

$$\mathcal{L}_{con} = \alpha\mathcal{L}_{gradient} + \beta\mathcal{L}_{intensity} + \gamma\mathcal{L}_{SSIM} + \lambda\mathcal{L}_{LBP}, \tag{10}$$

where α, β, γ and λ are parameters used to control the trade-off between four terms.

The adversarial loss $\mathcal{L}_{adv}$ in Eq.3 denotes the sum of two adversarial losses between the generator G and two discriminators, which can be formulated as:

$$\mathcal{L}_{adv} = \mathbb{E}\left[-\log\left(1 - D_{vis}\left(I_f\right)\right)\right] + \mathbb{E}\left[-\log\left(1 - D_{ir}\left(I_f\right)\right)\right], \tag{11}$$

where I_f denotes the fused image, $D_{vis}\left(I_f\right)$ denotes the probability that I_f belongs to the real visible image and $D_{ir}\left(I_f\right)$ denotes the probability that I_f is an infrared image.

3.3.2. Loss Function of Discriminators

The visible and infrared images contain rich texture details and contrast information. We establish an adversarial game between a generator and two discriminators for the result of the generator to match more with the distribution of the real data. Formally, the loss functions of discriminators are defined as follows:

$$\mathcal{L}_{D_{vis}} = \mathbb{E}[-\log D_{vis}(I_{vis})] + \mathbb{E}\left[-\log\left(1 - D_{vis}\left(I_f\right)\right)\right], \tag{12}$$

$$\mathcal{L}_{D_{ir}} = \mathbb{E}[-\log D_{ir}(I_{ir})] + \mathbb{E}\left[-\log\left(1 - D_{ir}\left(I_f\right)\right)\right]. \tag{13}$$

4. Experiments

In this section, we evaluate our method on publicly available datasets: the TNO dataset, the CVC14 dataset, the RoadScene dataset and some multi-spectral images. First, we provide the detailed experimental configurations. Then, we compare the results of our methods with four state-of-the-art methods: FusionGAN [22], DenseFuse [24], DDcGAN [23] and U2Fusion [25] on the first two public datasets. We also verify the improvement of the network performance by LBP and 1:4 ratio input through ablation experiment. Third, we test our method on the RoadScene dataset by fusing the infrared image and RGB image. Finally, we apply our method to multi-spectral remote sensing images and compare it with above four algorithms.

4.1. Implementation

4.1.1. Dataset

We validate our method on three classical fusion image datasets: the TNO dataset, the CVC14 dataset and the RoadScene dataset. Some multispectral images are also selected to test it.

The TNO dataset is the most commonly used infrared visible image *Human Factors* dataset. It contains multi-spectral images of different scenarios registered with different multi-band camera systems [48].

The CVC14 dataset is committed to promoting the development of autonomous driving technologies [49], [50]. It consists of two sets of sequences: the day set and night set. The day set includes $8,821$ images, the night set includes $9,589$ images, and all images have a 640×471 resolution.

The RoadScene dataset is a new image fusion dataset that has 221 infrared and visible image pairs. The images of the dataset are all collected from naturalistic driving videos, including roads, pedestrians, vehicles and other road scenes.

The multi-spectral remote sensing images used in this paper are recorded under the USA Airborne Multisensor Pod System (AMPS) program and include a large number of industrial, urban and natural scenes from a number of geographical locations captured by two hyper-spectral airborne scanners [51].

In the training stage, we adopt the overlapping cropping strategy to expand the dataset. Thirty-six infrared and visible image pairs of TNO are cropped into $22,912$ patch pairs with 84×84 pixels. The 84×84 visible and infrared image patches are used as source images to train the generator G and as labels to encourage the discriminators. For testing, we select 14, 26, 5 and 29 image pairs from the TNO, CVC14, RoadScene datasets and multi-spectral images, respectively.

4.1.2. Training Details

As mentioned in Section 3, parameters μ, α, β, γ and λ are used to control the balance of loss functions. We set $\mu = 0.6$, $\alpha = 0.2$, $\beta = 0.03$, $\gamma = 500$ and $\lambda = 0.5$. The initial learning rate and decay rate are set to 0.0002 and 0.75 to train the model, and RMSprop and SGD are adopted as the optimizers of the generator and discriminators, respectively. All experiments are conducted on a desktop with 2.30 GHz Intel Xeon CPU E5-2697 v4, NVIDIA Titan Xp, and 12 GB memory.

4.1.3. Metrics

In the test stage, we evaluate our method under 8 evaluation metrics to comprehensively compare it with the current methods.

An image fusion qualitative assessment mainly starts from the human visual system and judges the fusion effect according to the task goal. The goal of infrared-visible image fusion is to preserve the detailed and texture features of visible images and the contrast features of infrared images as much as possible. Conversely, quantitative evaluation comprehensively reflects the effect of image fusion through a variety of evaluation metrics.

In this paper, we select 8 metrics to use to evaluate our LPGAN and 4 other state-of-the-art methods. The metrics are standard deviation (SD) [52], average gradient (AG) [53], spatial frequency (SF) [54], mutual information (MI) [55], entropy (EN) [56], peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM) [57], and visual information fidelity (VIF) [58]. SD reflects the distribution of pixel values and contrast information. The larger the SD is, the higher the contrast and the better the visual effect. AG quantifies the gradient information of an image and reflects the amount of image details and textures. The larger AG is, the more detailed information the image contains and the better the fusion effect. SF is a gradient-based metric that can measure the gradient distribution effectively and reveal the details and texture of an image. The larger SF is, the richer edges and texture details are preserved. MI is a quality index that measures the amount of information that is transferred from source images to the fused image [55]. A larger MI represents more information that is transferred from source images to the fused image, which means better fusion performance. EN is a metric to measure the amount of information contained in the image, and the larger the EN value is, the more informative. $PSNR$ is a metric reflecting the distortion and anti-interference ability by the ratio of peak value power and noise power [23]. A large $PSNR$ indicates that little distortion occurred and there is a strong anti-interference ability. $SSIM$ is used to measure the structural similarity between two images and consists of three components: loss of correlation, loss of luminance and contrast distortion. The product of the three components is the assessment result of the fused image [9]. We calculate the average $SSIM$ between the fused image and two source images as the final result. A larger value of $SSIM$ indicates that more structural information is maintained. VIF is consistent with the human visual system and is applied to measure information fidelity. The larger VIF is, the better the visual effect and the less distortion there is between the fused image and source images.

4.2. Results on the TNO Dataset

4.2.1. Qualitative Comparison

We provide 6 image pairs to report some intuitive results on the fusion performance, as shown in the Fig. 6. Compared with other methods, our LPGAN has 3 advantages. First, the proposed method maintains the high-contrast characteristics of infrared images, as shown in the third and fourth examples, which proves to be effective for automatic target detection and location. Taking the third result as an example, only our LPGAN and FusionGAN can clearly distinguish the front and top of the pipeline and clearly see the edges of it, but FusionGAN poorly preserves the details of other areas. Second, our LPGAN can preserve rich texture details of visible images, which is beneficial for accurate target recognition. As shown in the first example, only the result of LPGAN can distinguish the detailed and texture features of the ground. Finally, our results all have a rich pixel intensity distribution, which means our results have more information and are consistent with the human visual system.

As shown in Fig. 6, DenseFuse and U2Fusion cannot preserve the thermal information well in the infrared image. The third group of results shows the brightness of the buildings and pipelines is considerably lower than that of the infrared image. Conversely, FusionGAN and DDcGAN cannot retain the detailed and texture information of visible images. In the second set of results, FusionGAN cannot retain the detailed features of the chimney well, while DDcGAN hardly retains the detailed features of the human. In contrast, although LPGAN tends to retain the detailed features of visible images, it still preserves important contrast information from the infrared images. As shown in the first column of Fig. 6, in our result, the details of the ground can be clearly distinguished, while the thermal information of the infrared image is also well preserved.

4.2.2. Quantitative Comparison

Fig. 7 shows the results of all of the examined methods on 14 tested image pairs of the TNO dataset. Our LPGAN can generate the largest average values on SD, AG, SF, and MI and the second largest values on EN. In particular, our LPGAN achieves the best values of SD, SF, MI and EN on 6, 6, 7 and 6 image pairs, respectively. For $PSNR$, $SSIM$ and VIF, our method ranks third, but the gaps with the top-ranked methods are very small. These results demonstrate that our method is able to preserve the best edges and texture details and contains the highest contrast information. Our results also contain rich information. The value on EN is only less than DDcGAN, but our performance on MI is better, meaning that DDcGAN has more false information than our method. In addition, our method can reduce noise interference very well and has a strong correlation with the source images. Finally, although DenseFuse and U2Fusion have larger values of $SSIM$, our method achieves a better balance between feature information and the visual

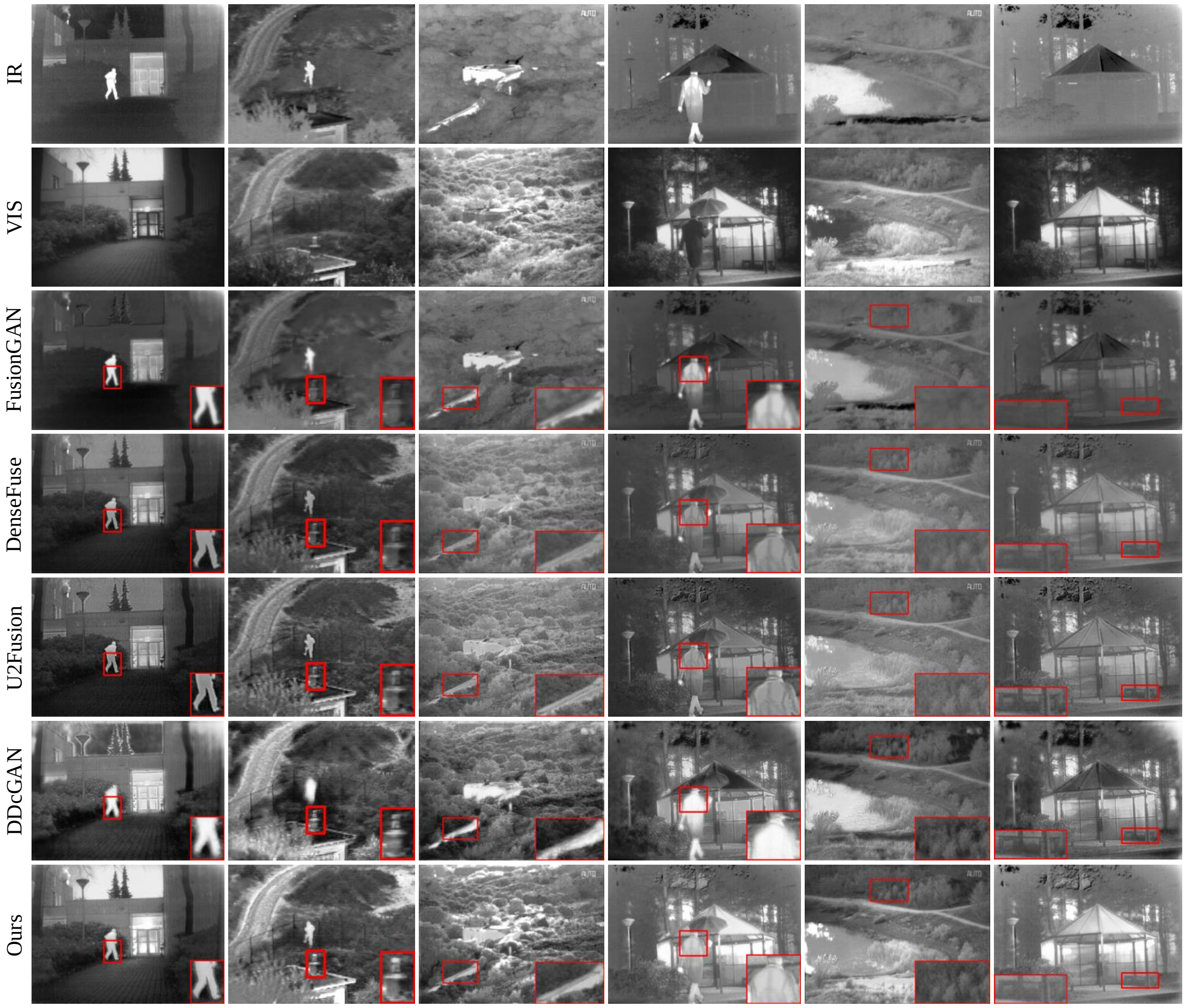


Figure 6. Qualitative results on the TNO dataset. From top to bottom: infrared image, visible image, fusion results of FusionGAN, DenseFuse, U2Fusion, DDcGAN and our LPGAN.

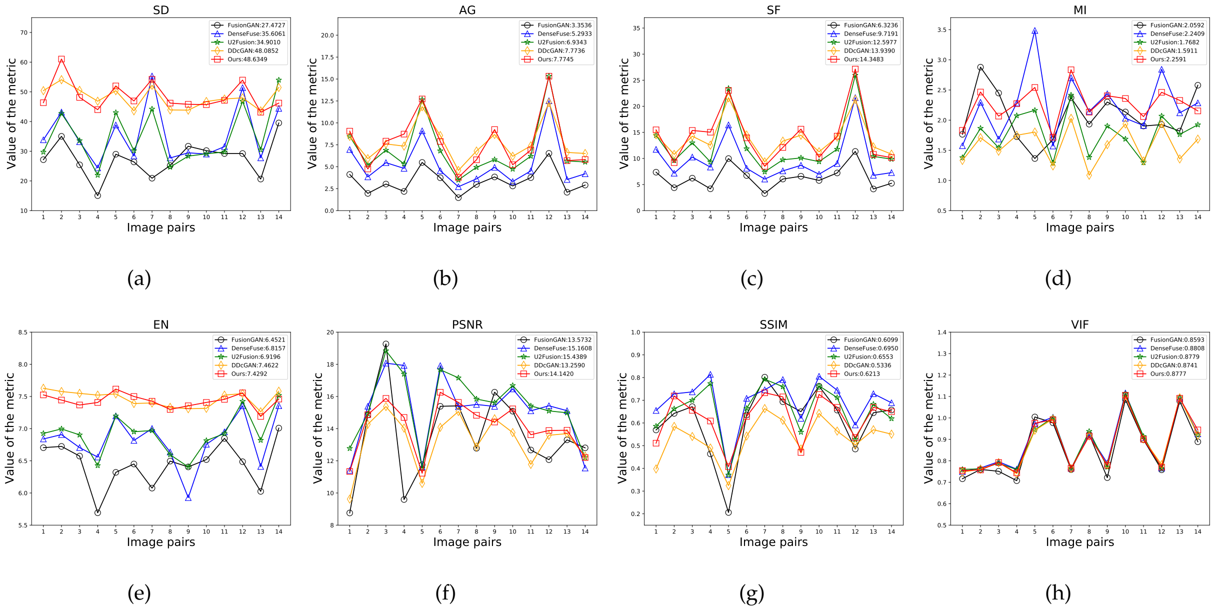
effect. From the VIF, it can be seen that the results of our method have a good visual effect.

4.3. Results on the CVC14 Dataset

To evaluate the effectiveness of our proposed method, we conduct an experiment on the CVC14 dataset. Twenty-six image pairs are selected from different scenes for evaluation.

4.3.1. Qualitative Comparison

We perform a qualitative comparison on 5 typical image pairs, as shown in the first two rows of Fig. 8, to demonstrate the characteristics of our method. The different light and dark changes in the results indicate that only DDcGAN and our LPGAN have high contrast information. For example, in the first and second images, the pixel intensity distributions of the cars, buildings, roads and people in the results of these two methods are abundant, but the results of other methods are not as obvious. Nevertheless, our results preserve more light information, as shown in the last group of results, meaning that LPGAN can extract more features from the source images. In terms of detail information preservation and visual effects, DDcGAN has more artifacts in the image due to the pursuit of high contrast, while the results of DenseFuse and U2Fusion have



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 7. Quantitative comparison with 4 state-of-the-art methods on the TNO dataset. Means of metrics for different methods are shown in the legends.

poor image visual effects due to less contrast. In contrast, our method can retain rich edges and texture details, while avoiding blurring and recognition difficulties due to darker colors, as shown in the third group.

4.3.2. Quantitative Comparison

As shown in Fig. 9, 26 testing image pairs of the CVC14 dataset are selected to further display quantitative comparisons of our LPGAN and the other examined methods. Our LPGAN still achieves the largest mean values on AG, SF, and MI. In particular, our LPGAN achieves the largest values of SF, AG, and MI on 21, 15 and 15 image pairs. On SF and MI, our results are 8.3% and 7.0% higher than the second place, respectively. For the metric SD and EN, our LPGAN can also achieve comparable results and only follows behind DDcGAN. However, the lower values of MI and $PSNR$ indicate that there is more noise and fake information in the DDcGAN results. For $PSNR$, $SSIM$ and VIF, all algorithms performed very well, with small gaps. The results demonstrate that LPGAN can not only extract rich detailed information from visible images but also retain important contrast information from infrared images.

4.4. Ablation Study

Because images of different modalities have different information distributions, we adopt a 1:4 ratio input to improve the ability of network feature extraction. At the same time, in order to ensure that the fused images have rich details and high structural similarity, we introduce LBP into the loss function to guide the optimization of the network. In this work, to evaluate the effect of a 1:4 ratio input and LBP, we train six models with exactly the same parameter settings on the TNO dataset according to the ratio of the input and whether LBP is used. In Fig. 10, we select a set of experimental results to show the differences between the six models.

4.4.1. The effect of LBP

We use an ablative comparison by removing the LBP. As shown in Fig. 10, given the same ratio of inputs, the fusion results of LPGAN trained with $\mathcal{L}_{LBP}$ contain more detailed information and are more in line with human-visible systems. In the second row of the figure, the images generated by LPGAN trained with $\mathcal{L}_{LBP}$ obviously have sharper

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 April 2023 doi:10.20944/preprints202304.0086.v1

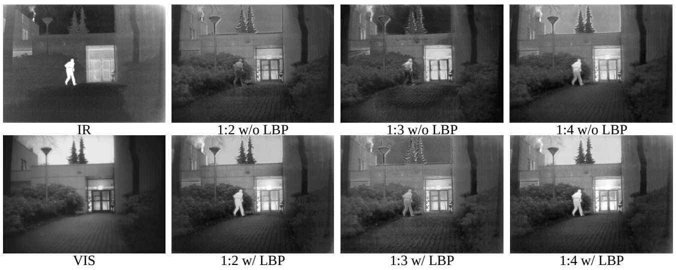


Figure 10. Ablation experiment on TNO dataset. From left to right: source images and fusion results of LPGAN with different settings.

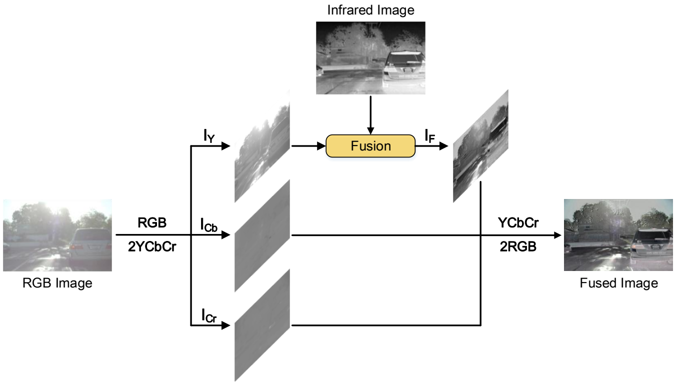


Figure 11. The fusion framework for RGB-infrared image fusion.

outlines, and there is more detailed background information. In addition, without using $\mathcal{L}_{LBP}$, the results have less contrast and more distortion. Through Table. 1, we can find that after adding LBP, the performance of all models are improved in almost all evaluation metrics, especially in SD, AG and SF. This shows that after adding LBP, the ability of the model to extract detailed features are indeed enhanceed.

4.4.2. The effect of proportional input

The proportional input method can enhance the ability of the network to extract the feature information of different modal images, which is conducive to the effect of fusion. We explore its effectiveness of different ratios of input by setting the ratio as 1:2, 1:3 and 1:4. Since the models using LBP have better performance, we compare the models using LBP. As shown in Table. 1, the last model (1:4 w/LBP) reaches the best in three of the eight evaluation metrics, and three reaches the second best value. The other two also has a small gap with the top value, especially in the key metrics such as SD, MI and SSIM, which are better than the first model (1:2 w/LBP) and the second model (1:3 w/LBP). The second model (1:3 w/LBP) achieves the best in the four evaluation metrics, but its performance in MI and SSIM is poor, indicating that its output results contain more false information, and as shown in Fig. 10, the model with 1:4 scale has better visual effect compared with the other two, so we finally choose 1:4 as the input ratio of the network.

4.5. Additional Results for RGB Images and Infrared Images

Apart from grayscale image fusion, LPGAN can also be used in RGB-infrared image fusion task. As shown in Fig. 11, we first convert the RGB image into the YCbCr color space. Then, we use the proposed LPGAN to fuse the luminance channel of the RGB image and infrared image. This is because the structural information is usually saved in the luminance channel. After that, the fused image is combined with chroma (Cb and Cr) channels and then converted into the RGB color space. In Fig. 12, we select 5 sets of experimental results to show the effect of LPGAN. It can be seen from the results that LPGAN can fully extract feature information from infrared images and visible images

Table 1: Ablation Experimrnt Results on The TNO Dataset(Red: Optimal, Blue: Suboptimal).

Algorithms	SD	AG	SF	MI	EN	PSNR	SSIM	VIF
1:2 w/o LBP	34.6916	7.2434	13.0421	1.6990	7.0477	14.4241	0.6287	0.8831
1:2 w/ LBP	47.8106	7.7136	14.1130	2.1066	7.3885	14.0895	0.6167	0.8744
1:3 w/o LBP	36.5414	7.7551	13.9280	1.7276	7.0850	13.8040	0.5578	0.8758
1:3 w/ LBP	37.1527	8.4237	15.2863	1.7129	7.1475	14.4447	0.6134	0.8834
1:4 w/o LBP	44.6710	7.5059	13.7169	1.9475	7.3457	14.4303	0.6035	0.8785
1:4 w/ LBP	48.6349	7.7745	14.3483	2.2591	7.4292	14.1420	0.6213	0.8777

Table 2: Evaluation of Fusion Results of Multi-Spectral Remotr Sensing Images(Red: Optimal, Blue: Suboptimal).

Algorithms	SD	AG	SF	EN	MI	PSNR	SSIM	CC
FusionGAN	30.2032	5.3697	11.0368	6.4712	2.2562	15.3458	0.6251	0.6553
DenseFuse	40.3449	8.2281	16.5689	6.8515	2.5893	16.3808	0.6899	0.7651
U2Fusion	43.3423	10.7603	21.3030	6.9693	2.3393	16.5555	0.6664	0.7496
DDcGAN	52.1831	10.8181	21.0603	7.4602	2.1890	14.2005	0.5887	0.6688
Ours	43.3589	11.4636	22.6095	7.1701	2.4498	15.8229	0.6741	0.7499

and fuse them well. Taking the first image as an example, the license plate number in the visible light image is very fuzzy, but the model accurately fuses the two based on the feature information in the infrared image. Only DenseFusion, U2Fusion and the output results of our method can clearly see the license plate number of the vehicle. However, in the second group of experimental results, only the fusion result of our method can better preserve the color of the sky in the visible image, and only the result of our method can judge that it is daytime.

4.6. Multi-spectral Image Fusion Expansion Experiment

In this work, we apply our method to multi-spectral remote sensing image fusion and compare it with 4 state-of-the-art fusion algorithms.

We report 6 typical image pairs, as seen in Fig. 13. The first two rows are multi-spectral images of two different bands, and both images are taken from the same scene and have the same resolution. The images in the first row have the same high contrast characteristics as the infrared images, and the images in the second row have the same feature of rich detailed information as the visible images. Thus, we follow the idea of infrared and visible image fusion to fuse these two remote sensing images.

Detailed features are the most important information in remote sensing images. DenseFuse, U2Fusion and our LPGAN can preserve it well, but our method performs better. In the third and fifth groups of results, only LPGAN exhibits subtle changes in ground details without producing artifacts, which is very important in small target detection tasks. In addition, it is obvious that only DDcGAN and our method can achieve a high contrast. For example, in the last set of results, DDcGAN and our method can clearly distinguish roads from background information, while for the other 3 methods it is more difficult. However, DDcGAN produces many artifacts in the process of image fusion, as shown in the fourth and fifth experimental results. In contrast, the results of our method all have clear images, no distortion or artifacts and are very consistent with the human visual system.

To evaluate the capability of LPGAN more objectively, we also conduct a quantitative assessment. Twenty-nine pairs of images are selected for testing and eight performance metrics are performed. For the characteristics of remote sensing images, we replace VIF with the correlation coefficient (CC) [59]. The CC expresses the degree to which the source image and fused image are related, and Pearson's correlation is mostly used to measure the abovementioned correlation [60]. Table. 2 shows the results of the quantitative comparisons. Our LPGAN achieves the best performances on AG and SF and achieves the second largest values in other metrics except $PSNR$. For $PSNR$, our method also shows comparable result and generates the third largest average value. DenseFuse is slightly better than our method in terms of $SSIM$, MI, and CC, but there is a large gap with our method in terms of SD, AG, and SF, which indicates that DenseFuse is not sufficient for detailed retention. U2Fusion and our method have a small gap in all metrics, but it only achieves better result than ours on $PSNR$. The largest values of EN and SD are achieved by DDcGAN, and LPGAN all ranks second. However, low values on $SSIM$, MI and CC indicate that there is considerable fake information in the results of DDcGAN, and DDcGAN cannot retain structural information from source images well.

5. Discussion

In this section, we discuss a key issue of image fusion and the effectiveness of our solution. We also introduce the limitations of our method and our future work.

For different characteristics and needs, many evaluation metrics for image fusion have been proposed. An excellent image fusion algorithm must not only have high-quality visual effects but also achieve good results on these metrics. During our experiments, we found that good visual effects conflict with some metrics, such as SD, SF and AG. Specifically, an algorithm with a high $SSIM$ index generally has better visual effects, but its performance on SD, SF and AG will be poor, such as for FusionGAN, DenseFuse and U2Fusion. The reverse is also the same, such as for DDcGAN. Our LPGAN successfully avoids this problem. It not only performs well on SD, SF and AG but also performs well on $SSIM$. We use DDcGAN as an example to analyze the causes of the previous problem. In the process of image fusion, too much emphasis is placed on gradient intensity changes, and the gradient direction and the texture features of the source image are ignored, leading to poor visual effects. Based on this idea, we creatively introduce LBP into the loss function, and to preserve the spatial information of the image, we convert the calculated LBP into a 4096-dimensional vector. As shown in the results of the ablation experiment, the introduction of LBP can improve LPGAN in almost all aspects and successfully solve the abovementioned key problem.

Although LPGAN can achieve good performance on infrared-visible image fusion and multi-spectral image fusion tasks, avoiding the above problem, its visual effects and performance on $SSIM$ still need to be improved. In the future, we will try to consider image fusion from the perspective of decision-making and focus on the use of an attention mechanism to enable the network to perform fusion operations based on the information distribution of the source images. This is because in actual image fusion tasks, one source image is often of low quality. Therefore, we hope to adjust the fusion parameters adaptively by introducing an attention mechanism. Furthermore, we plan to apply our method in other image fusion tasks, such as medical image fusion, multi-focus image fusion and multi-exposure image fusion.

6. Conclusion

In this paper, we propose a novel GAN-based visible-infrared image fusion method, termed as LPGAN. It is an unsupervised end-to-end model. We adopt a cGAN as the framework and employ two discriminators, avoiding the mod collapse issue and providing the network with high stability. Simultaneously, considering the differences of imaging mechanisms and characteristics between visible images and infrared images, a pseudo-siamese network is used for generator to extract the detailed features and contrast features. We also set a 1:4 ratio input method according to the characteristics of different modal images to further improve the feature extraction capability of the network. In response to the existing problem, we innovatively introduce LBP into the loss function, which greatly improves the texture description ability and anti-interference ability of LPGAN. Compared with other four state-of-the-art methods on the publicly available TNO dataset, CVC14 dataset and ROADScene dataset, our method can achieve advanced performance both qualitatively and quantitatively. The experiment on a multi-spectral image fusion task also demonstrates that our LPGAN can achieve state-of-the-art performance.

Author Contributions: The first two authors have equally contributed to the work.

Funding: .

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to their large size.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, S.; Kang, X.; Fang, L.; Hu, J.; Yin, H. Pixel-level image fusion: A survey of the state of the art. *information Fusion* **2017**, *33*, 100–112.
2. Li, S.; Kang, X.; Hu, J. Image fusion with guided filtering. *IEEE Transactions on Image processing* **2013**, *22*, 2864–2875.
3. Yang, J.; Zhao, Y.; Chan, J.C.W. Hyperspectral and Multispectral Image Fusion via Deep Two-Branches Convolutional Neural Network. *Remote Sensing* **2018**, *10*, 800.
4. Eslami, M.; Mohammadzadeh, A. Developing a Spectral-Based Strategy for Urban Object Detection From Airborne Hyperspectral TIR and Visible Data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2016**, *9*, 1808–1816. doi:10.1109/JSTARS.2015.2489838.
5. Wang, J.; Li, L.; Liu, Y.; Hu, J.; Xiao, X.; Liu, B. AI-TFNet: Active Inference Transfer Convolutional Fusion Network for Hyperspectral Image Classification. *Remote Sensing* **2023**, *15*, 1292.
6. Wang, Z.; Ziou, D.; Armenakis, C.; Li, D.; Li, Q. A comparative analysis of image fusion methods. *IEEE transactions on geoscience and remote sensing* **2005**, *43*, 1391–1402.
7. James, A.P.; Dasarathy, B.V. Medical image fusion: A survey of the state of the art. *Information fusion* **2014**, *19*, 4–19.
8. Ghassemian, H. A review of remote sensing image fusion methods. *Information Fusion* **2016**, *32*, 75–89.
9. Ma, J.; Ma, Y.; Li, C. Infrared and visible image fusion methods and applications: A survey. *Information Fusion* **2019**, *45*, 153–178.
10. Hu, H.M.; Wu, J.; Li, B.; Guo, Q.; Zheng, J. An adaptive fusion algorithm for visible and infrared videos based on entropy and the cumulative distribution of gray levels. *IEEE Transactions on Multimedia* **2017**, *19*, 2706–2719.
11. He, K.; Zhou, D.; Zhang, X.; Nie, R.; Wang, Q.; Jin, X. Infrared and visible image fusion based on target extraction in the nonsubsampled contourlet transform domain. *Journal of Applied Remote Sensing* **2017**, *11*, 015011.
12. Bin, Y.; Chao, Y.; Guoyu, H. Efficient image fusion with approximate sparse representation. *International Journal of Wavelets, Multiresolution and Information Processing* **2016**, *14*, 1650024.
13. Zhang, Q.; Liu, Y.; Blum, R.S.; Han, J.; Tao, D. Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review. *Information Fusion* **2018**, *40*, 57–75.
14. Naidu, V. Hybrid DDCT-PCA based multi sensor image fusion. *Journal of Optics* **2014**, *43*, 48–61.
15. Ma, J.; Zhou, Z.; Wang, B.; Zong, H. Infrared and visible image fusion based on visual saliency map and weighted least square optimization. *Infrared Physics & Technology* **2017**, *82*, 8–17.
16. Yin, M.; Duan, P.; Liu, W.; Liang, X. A novel infrared and visible image fusion algorithm based on shift-invariant dual-tree complex shearlet transform and sparse representation. *Neurocomputing* **2017**, *226*, 182–191.
17. Fu, D.; Chen, B.; Wang, J.; Zhu, X.; Hilker, T. An Improved Image Fusion Approach Based on Enhanced Spatial and Temporal the Adaptive Reflectance Fusion Model. *Remote Sensing* **2013**, *5*, 6346–6360.
18. Ma, J.; Chen, C.; Li, C.; Huang, J. Infrared and visible image fusion via gradient transfer and total variation minimization. *Information Fusion* **2016**, *31*, 100–109.
19. Ma, Y.; Chen, J.; Chen, C.; Fan, F.; Ma, J. Infrared and visible image fusion using total variation model. *Neurocomputing* **2016**, *202*, 12–19.
20. Liu, Y.; Chen, X.; Wang, Z.; Wang, Z.J.; Ward, R.K.; Wang, X. Deep learning for pixel-level image fusion: Recent advances and future prospects. *Information Fusion* **2018**, *42*, 158–173.
21. Xu, F.; Liu, J.; Song, Y.; Sun, H.; Wang, X. Multi-Exposure Image Fusion Techniques: A Comprehensive Review. *Remote Sensing* **2022**, *14*, 771.
22. Ma, J.; Yu, W.; Liang, P.; Li, C.; Jiang, J. FusionGAN: A generative adversarial network for infrared and visible image fusion. *Information Fusion* **2019**, *48*, 11–26.

23. Ma, J.; Xu, H.; Jiang, J.; Mei, X.; Zhang, X.P. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion. *IEEE Transactions on Image Processing* **2020**, *29*, 4980–4995.
24. Li, H.; Wu, X.J. DenseFuse: A fusion approach to infrared and visible images. *IEEE Transactions on Image Processing* **2018**, *28*, 2614–2623.
25. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2020**.
26. Ojala, T.; Pietikainen, M.; Harwood, D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. Proceedings of 12th International Conference on Pattern Recognition, 1994, Vol. 1, pp. 582–585 vol.1. doi:10.1109/ICPR.1994.576366.
27. Zhang, H.; Xu, H.; Xiao, Y.; Guo, X.; Ma, J. Rethinking the Image Fusion: A Fast Unified Image Fusion Network based on Proportional Maintenance of Gradient and Intensity. *Proceedings of the AAAI Conference on Artificial Intelligence* **2020**, *34*, 12797–12804. doi:10.1609/aaai.v34i07.6975.
28. Ma, J.; Zhang, H.; Shao, Z.; Liang, P.; Xu, H. GANMcC: A Generative Adversarial Network With Multiclassification Constraints for Infrared and Visible Image Fusion. *IEEE Transactions on Instrumentation and Measurement* **2020**, *PP*, 1–1. doi:10.1109/TIM.2020.3038013.
29. González, A.; Fang, Z.; Socarras, Y.; Serrat, J.; Vázquez, D.; Xu, J.; López, A.M. Pedestrian detection at day/night time with visible and FIR cameras: A comparison. *Sensors* **2016**, *16*, 820.
30. Liu, Y.; Chen, X.; Ward, R.K.; Wang, Z.J. Image fusion with convolutional sparse representation. *IEEE signal processing letters* **2016**, *23*, 1882–1886.
31. Liu, Y.; Chen, X.; Peng, H.; Wang, Z. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion* **2017**, *36*, 191–207.
32. Li, H.; Wu, X.J.; Kittler, J. Infrared and visible image fusion using a deep learning framework. 2018 24th international conference on pattern recognition (ICPR). IEEE, 2018, pp. 2705–2710.
33. Liu, Q.; Zhou, H.; Xu, Q.; Liu, X.; Wang, Y. PSGAN: A generative adversarial network for remote sensing image pan-sharpening. *IEEE Transactions on Geoscience and Remote Sensing* **2020**.
34. Ram Prabhakar, K.; Sai Srikar, V.; Venkatesh Babu, R. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. Proceedings of the IEEE international conference on computer vision, 2017, pp. 4714–4722.
35. Ma, J.; Yu, W.; Chen, C.; Liang, P.; Guo, X.; Jiang, J. Pan-GAN: An unsupervised pan-sharpening method for remote sensing image fusion. *Information Fusion* **2020**, *62*, 110–120.
36. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems* **2014**, *27*.
37. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784* **2014**.
38. Durugkar, I.; Gemp, I.; Mahadevan, S. Generative Multi-Adversarial Networks. *Proceedings of the International Conference on Learning Representations* **2017**.
39. Wang, L.; Sindagi, V.; Patel, V. High-quality facial photo-sketch synthesis using multi-adversarial networks. 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018). IEEE, 2018, pp. 83–90.
40. Aghakhani, H.; Machiry, A.; Nilizadeh, S.; Kruegel, C.; Vigna, G. Detecting deceptive reviews using generative adversarial networks. 2018 IEEE Security and Privacy Workshops (SPW). IEEE, 2018, pp. 89–95.
41. Ojala, T.; Pietikäinen, M.; Mäenpää, T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987.
42. Zhao, G.; Pietikainen, M. Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2007**, *29*, 915–928. doi:10.1109/TPAMI.2007.1110.
43. Maturana, D.; Mery, D.; Soto, Á. Face Recognition with Local Binary Patterns, Spatial Pyramid Histograms and Naive Bayes Nearest Neighbor Classification. 2009 International Conference of the Chilean Computer Science Society, 2009, pp. 125–132.
44. Tapia, J.E.; Perez, C.A.; Bowyer, K.W. Gender Classification from Iris Images Using Fusion of Uniform Local Binary Patterns. Computer Vision - ECCV 2014 Workshops, 2015, pp. 751–763.
45. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 4700–4708.
46. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* **2012**, *25*, 1097–1105.
47. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing* **2004**, *13*, 600–612.
48. Li, G.; Lin, Y.; Qu, X. An infrared and visible image fusion method based on multi-scale transformation and norm optimization. *Information Fusion* **2021**, *71*, 109–129.
49. Li, G.; Yang, Y.; Zhang, T.; Qu, X.; Cao, D.; Cheng, B.; Li, K. Risk assessment based collision avoidance decision-making for autonomous vehicles in multi-scenarios. *Transportation research part C: emerging technologies* **2021**, *122*, 102820.
50. Li, G.; Li, S.E.; Cheng, B.; Green, P. Estimation of driving style in naturalistic highway traffic using maneuver transition probabilities. *Transportation Research Part C: Emerging Technologies* **2017**, *74*, 113–125.
51. AMPS Programme **September 1998**.

52. Eskicioglu, A.M.; Fisher, P.S. Image quality measures and their performance. *IEEE Transactions on communications* **1995**, *43*, 2959–2965.
53. Cui, G.; Feng, H.; Xu, Z.; Li, Q.; Chen, Y. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition. *Optics Communications* **2015**, *341*, 199–209.
54. Eskicioglu, A.M.; Fisher, P.S. Image quality measures and their performance. *IEEE Transactions on communications* **1995**, *43*, 2959–2965.
55. Qu, G.; Zhang, D.; Yan, P. Information measure for performance of image fusion. *Electronics letters* **2002**, *38*, 313–315.
56. Roberts, J.W.; Van Aardt, J.A.; Ahmed, F.B. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *Journal of Applied Remote Sensing* **2008**, *2*, 023522.
57. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE signal processing letters* **2002**, *9*, 81–84.
58. Han, Y.; Cai, Y.; Cao, Y.; Xu, X. A new image fusion performance metric based on visual information fidelity. *Information fusion* **2013**, *14*, 127–135.
59. Du, Q.; Xu, H.; Ma, Y.; Huang, J.; Fan, F. Fusing infrared and visible images of different resolutions via total variation model. *Sensors* **2018**, *18*, 3827.
60. Tian, X.; Zhang, M.; Yang, C.; Ma, J. Fusionndvi: A computational fusion approach for high-resolution normalized difference vegetation index. *IEEE Transactions on Geoscience and Remote Sensing* **2020**, *59*, 5258–5271.
61. Lopez-Molina, C.; Montero, J.; Bustince, H.; De Baets, B. Self-adapting weighted operators for multiscale gradient fusion. *Information Fusion* **2018**, *44*, 136–146. doi:https://doi.org/10.1016/j.inffus.2018.03.004.
62. Dogra, A.; Goyal, B.; Agrawal, S. From Multi-Scale Decomposition to Non-Multi-Scale Decomposition Methods: A Comprehensive Survey of Image Fusion Techniques and Its Applications. *IEEE Access* **2017**, *5*, 16040–16067. doi:10.1109/ACCESS.2017.2735865.
63. Yamamoto, N.; Saito, T.; Ogawa, S.; Ishimaru, I. Middle infrared (wavelength range: 8 μm-14 μm) 2-dimensional spectroscopy (total weight with electrical controller: 1.7 kg, total cost: less than 10,000 USD) so-called hyper-spectral camera for unmanned air vehicles like drones. Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery Xxii. International Society for Optics and Photonics, 2016, Vol. 9840, p. 984028.
64. Yang, D.; Zheng, Y.; Xu, W.; Sun, P.; Zhu, D. A Generative Adversarial Network for Image Fusion via Preserving Texture Information. *International Conference on Guidance, Navigation and Control* **2022**.
65. Tian, J.; Leng, Y.; Zhao, Z.; Xia, Y.; Sang, Y.; Hao, P.; Zhan, J.; Li, M.; Liu, H. Carbon quantum dots/hydrogenated TiO2 nanobelt heterostructures and their broad spectrum photocatalytic properties under UV, visible, and near-infrared irradiation. *Nano Energy* **2015**, *11*, 419–427.
66. Jin, X.; Jiang, Q.; Yao, S.; Zhou, D.; Nie, R.; Hai, J.; He, K. A survey of infrared and visual image fusion methods. *Infrared Physics & Technology* **2017**, *85*, 478–501.
67. Das, S.; Zhang, Y. Color night vision for navigation and surveillance. *Transportation research record* **2000**, *1708*, 40–46.
68. Li, H.; Wu, X.J.; Durrani, T. NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. *IEEE Transactions on Instrumentation and Measurement* **2020**, *69*, 9645–9656.
69. Yang, Y.; Zhang, Y.; Huang, S.; Zuo, Y.; Sun, J. Infrared and visible image fusion using visual saliency sparse representation and detail injection model. *IEEE Transactions on Instrumentation and Measurement* **2020**, *70*, 1–15.
70. Xiang, T.; Yan, L.; Gao, R. A fusion algorithm for infrared and visible images based on adaptive dual-channel unit-linking PCNN in NSCT domain. *Infrared Physics & Technology* **2015**, *69*, 53–61.
71. Ma, J.; Liang, P.; Yu, W.; Chen, C.; Guo, X.; Wu, J.; Jiang, J. Infrared and visible image fusion via detail preserving adversarial learning. *Information Fusion* **2020**, *54*, 85–98.
72. Beck, A.; Teboulle, M. Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE transactions on image processing* **2009**, *18*, 2419–2434.
73. Fu, X.; Jia, S.; Xu, M.; Zhou, J.; Li, Q. Fusion of Hyperspectral and Multispectral Images Accounting for Localized Inter-image Changes. *IEEE Transactions on Geoscience and Remote Sensing* **2021**, pp. 1–1. doi:10.1109/TGRS.2021.3124240.
74. Sun, K.; Tian, Y. DBFNet: A Dual-Branch Fusion Network for Underwater Image Enhancement. *Remote Sensing* **2023**, *15*, 1195.

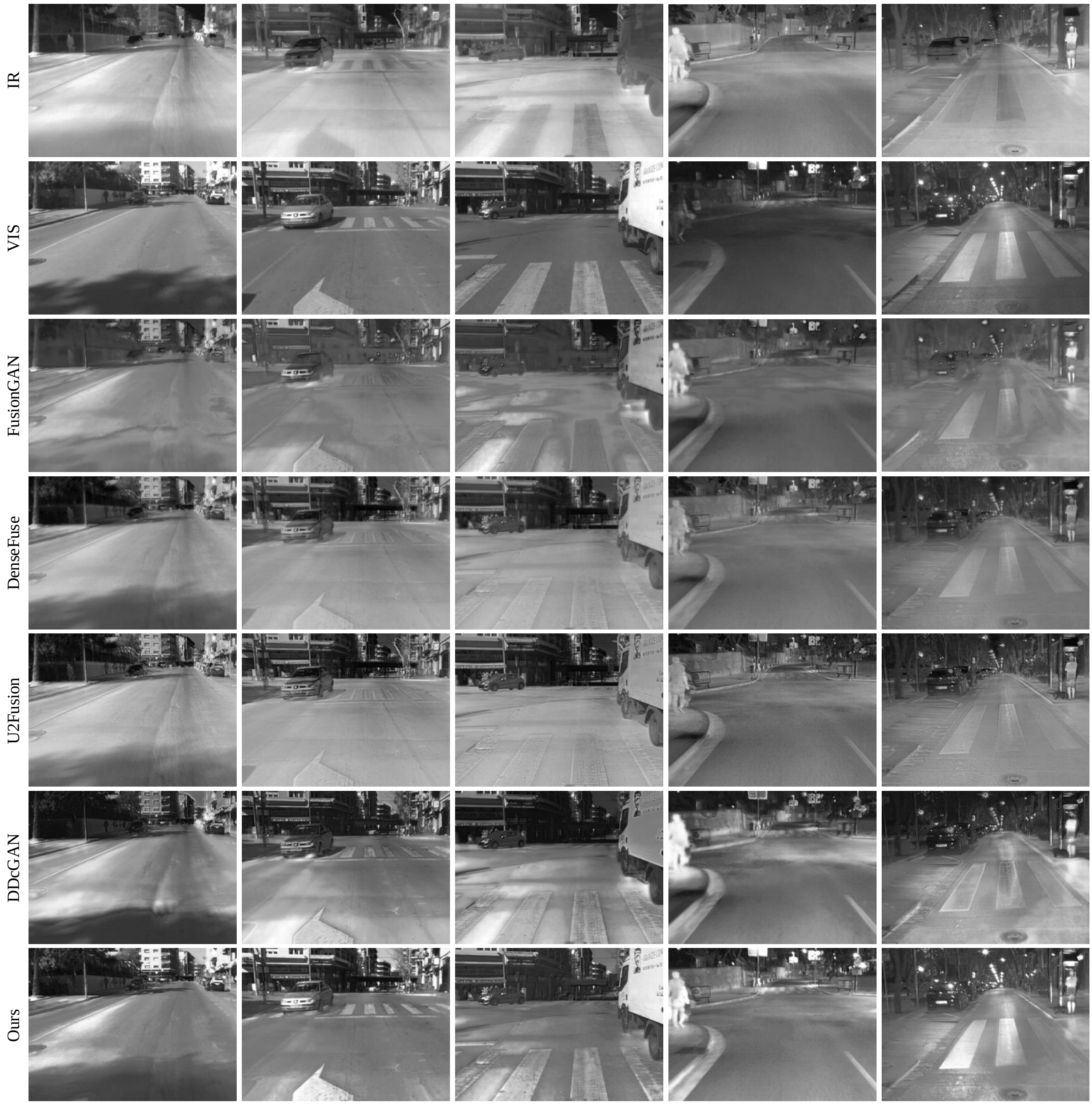
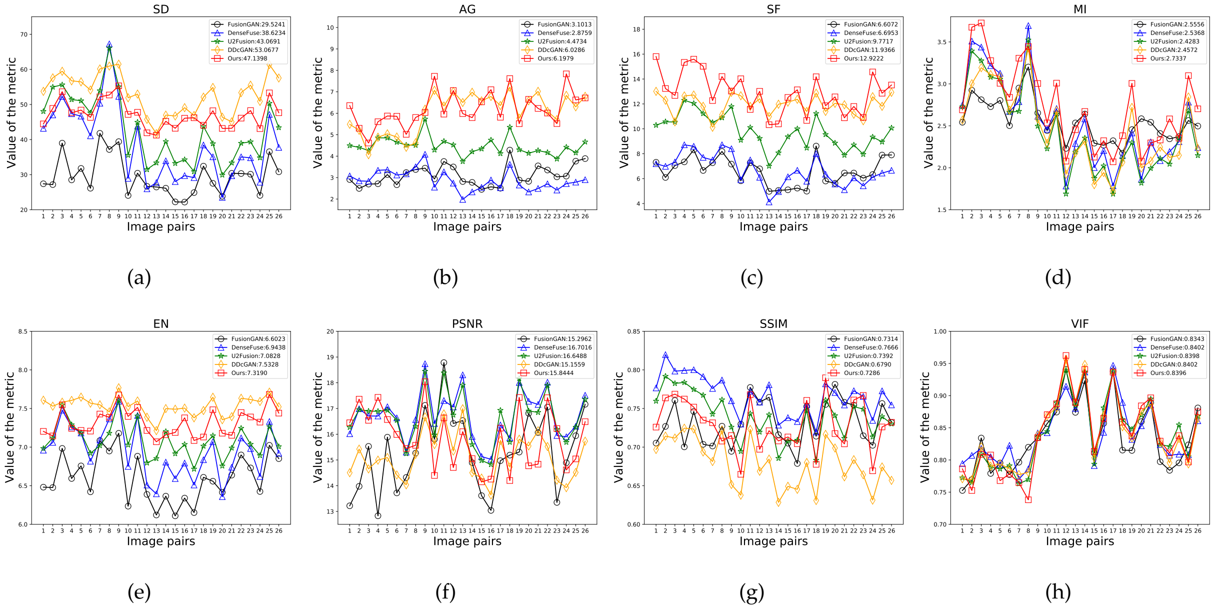


Figure 8. Qualitative results on the CVC14 dataset. From top to bottom: infrared image, visible image, fusion results of FusionGAN, DenseFuse, U2Fusion, DDcGAN and our LPGAN.

doi:10.20944/preprints202304.0086.v1



(a) (b) (c) (d)

(e) (f) (g) (h)

Figure 9. Quantitative comparison with 4 state-of-the-art methods on the CVC14 dataset. Means of metrics for different methods are shown in the legends.

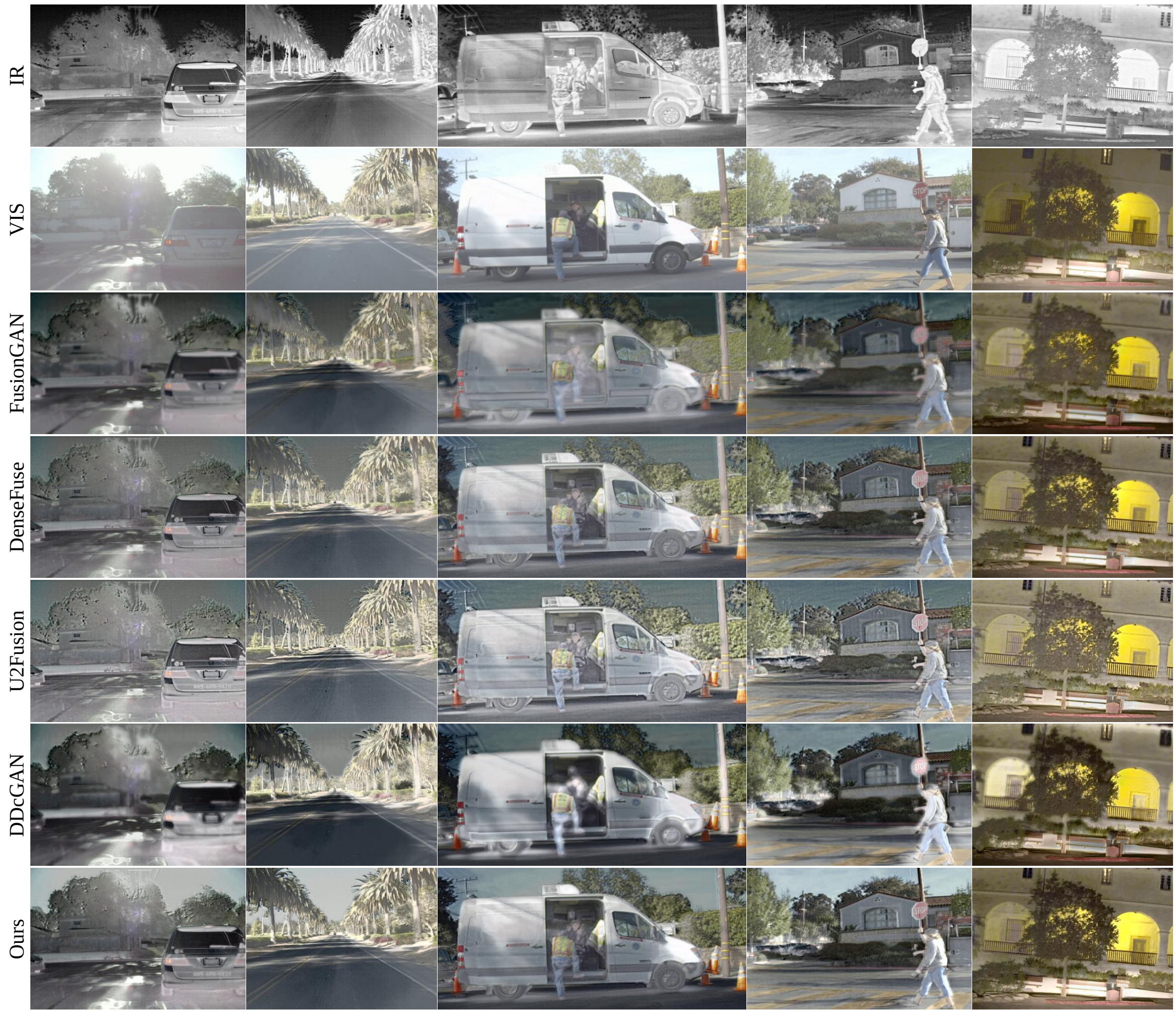


Figure 12. Fused results on the 5 image pairs in the RoadScene dataset. From top to bottom: the infrared images, the RGB images and the fused images.

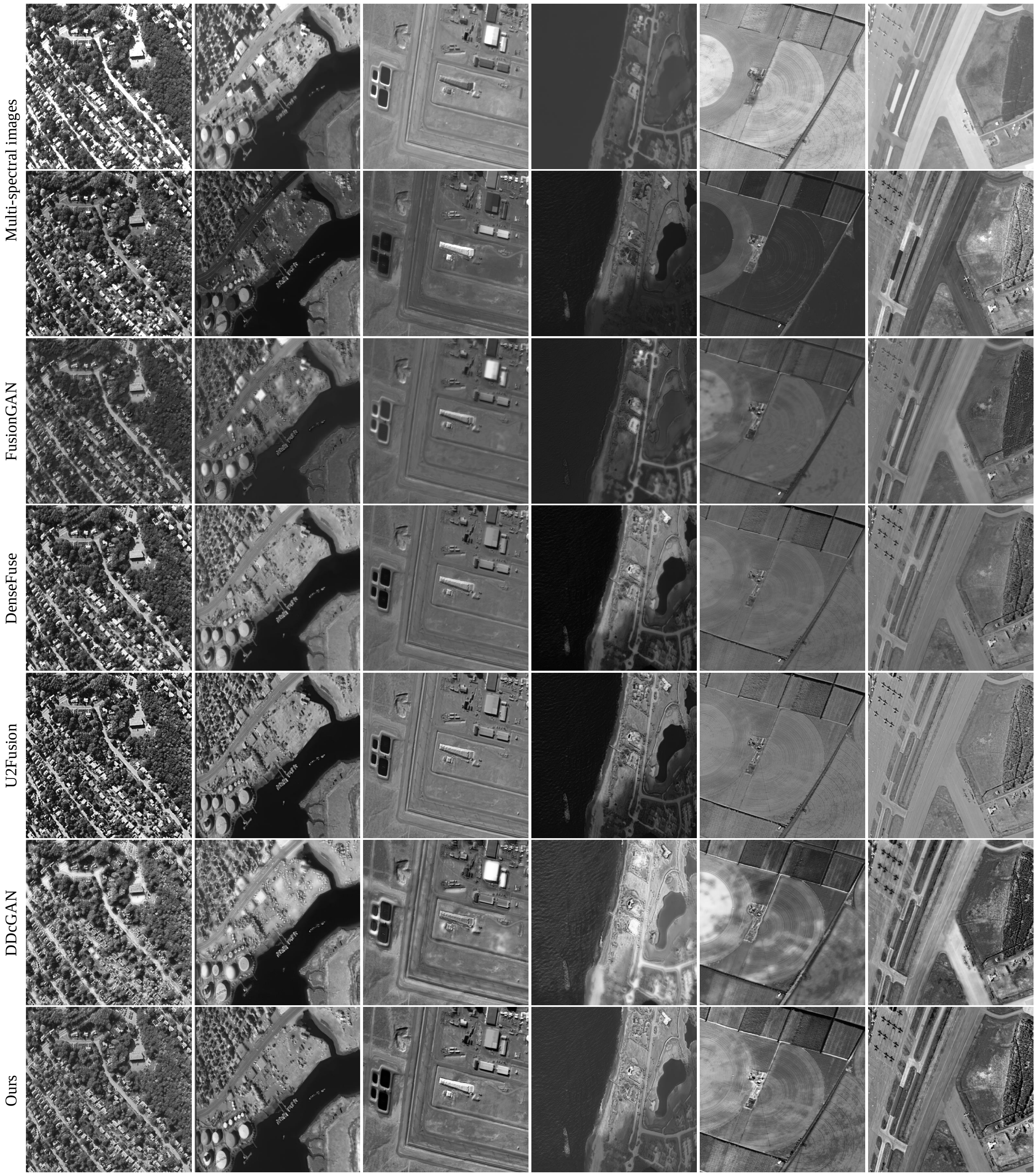


Figure 13. Fusion results of 6 pairs of multi-spectral remote sensing images. From top to bottom: two kinds of remote sensing images, fusion results of FusionGAN, DenseFuse, U2Fusion, DDcGAN and our LPGAN.