Neural Computing and Applications (2021) 33:6133–6145
https://doi.org/10.1007/s00521-020-05387-4

ORIGINAL ARTICLE

# GANFuse: a novel multi-exposure image fusion method based on generative adversarial networks

Zhiguang Yang[1] · Youping Chen[1] · Zhuliang Le[2] · Yong Ma[2]

Received: 14 May 2020 / Accepted: 24 September 2020 / Published online: 4 November 2020

## Abstract

In this paper, a novel multi-exposure image fusion method based on generative adversarial networks (termed as GANFuse) is presented. Conventional multi-exposure image fusion methods improve their fusion performance by designing sophisticated activity-level measurement and fusion rules. However, these methods have a limited success in complex fusion tasks. Inspired by the recent FusionGAN which firstly utilizes generative adversarial networks (GAN) to fuse infrared and visible images and achieves promising performance, we improve its architecture and customize it in the task of extreme exposure image fusion. To be specific, in order to keep content of extreme exposure image pairs in the fused image, we increase the number of discriminators differentiating between fused image and extreme exposure image pairs. While, a generator network is trained to generate fused images. Through the adversarial relationship between generator and discriminators, the fused image will contain more information from extreme exposure image pairs. Thus, this relationship can realize better performance of fusion. In addition, the method we proposed is an end-to-end and unsupervised learning model, which can avoid designing hand-crafted features and does not require a number of ground truth images for training. We conduct qualitative and quantitative experiments on a public dataset, and the experimental result shows that the proposed model demonstrates better fusion ability than existing multi-exposure image fusion methods in both visual effect and evaluation metrics.

**Keywords** Image fusion · Multi-exposure image · Generative adversarial network

## 1 Introduction

Powered by advanced digital image technology, the effect of image vision is more demanding than ever before. High dynamic range (HDR) technology, the one of the ways to improve image quality, has aroused extensive attention. It is widely applied in the fields of digital electronic products, remote sensing, security monitoring and so on. The dynamic range of image is the ratio of maximum brightness to minimum brightness. The dynamic range of real-world scenes is very wide [1]. However, ordinary image sensors have fixed exposure settings and can only get images with low dynamic range (LDR). Thus, due to the limitation of ordinary image sensors, it is difficult for ordinary image sensors to fully present the visual information in the real scene. The HDR technology can improve the dynamic range of the image. Through this technology, the visual information of the extreme exposed area of real-world scenes can be preserved [2]. Multi-exposure image fusion (MEF) is the most common technique in HDR technology, which merges LDR images with different exposures into a well-expose image of HDR.

In 1984, MEF was firstly proposed in [3]. After that, MEF has become a hot field, and many related methods have been proposed. Existing MEF methods could be generally divided into three categories: pixel- [4–7], region- [8–12], and deep learning-based methods [13–18]. The first two categories have developed for many years and widely used in all kinds of scenarios. Consequently, these methods are known as traditional fusion methods.

✉ Youping Chen
ypchen@hust.edu.cn

[1] The State Key Laboratory of Digital Manufacturing Equipment and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

[2] Electronic Information School, Wuhan University, Wuhan 430072, China

Generally, traditional methods contain three major steps, including image transformation, activity-level measurement and fusion rule designing [19]. However, these steps are limited by implementation difficulty and high computational costs.

Deep learning-based methods can avoid these problems. Because the trained network can generate the complex relationship between source images and fused image, it can automatically extract feature information from the images and fuse these features without manual participation in transformation and activity-level measurement. The fusion process is simpler and more applicable. Through constraint of loss function, the fused image has obvious targets, rich details and good visual effects. Existing deep learning-based methods have made some progress. But there is definitely room for improvement. More effective loss function and structure of the network will lead to better fusion results. Specifically, SSIM is one of the quality evaluations for image fusion, which measures the correlation loss, brightness loss and contrast loss between source images and fused image. DeepFuse sets SSIM as loss function in their model [13]. But it will lose other key information, such as contrast and texture information and so on. IFCNN regards pre-trained CNN as a tool to extract features from the source image [15]. However, the fusion rule is still designed manually. FusionGAN formulates the image fusion as an adversarial game between keeping the infrared thermal radiation information and preserving the visible appearance texture information [16]. Whereas, FusionGAN pays too much attention on information from the visible image and neglects information of the infrared image, which may cause the loss of information from the infrared image.

To overcome the above-mentioned problems, we propose a novel unsupervised MEF method based on GAN, named as GANFuse. GANFuse consists of three components: a generator and two discriminators. The generator attempts to obtain a fused image which contains valid information of the source images. Whereas, the discriminators are conducted to distinguish between fused image and source images. This adversarial process will force the generator to have better performance. As for loss functions, the pixel intensities loss and gradient loss are applied in our network that can help fused image to preserve luminance information and texture information from the source images. As shown in Fig. 1, the result of our GANFuse shows the better visual effect, including luminance and texture. Furthermore, in order to improve the robustness of the algorithm, we establish the training dataset from extreme exposure image pairs in different environments (indoor/outdoor, day/night, side-lighting/back-lighting and so on).

The contributions of this work are as follows:

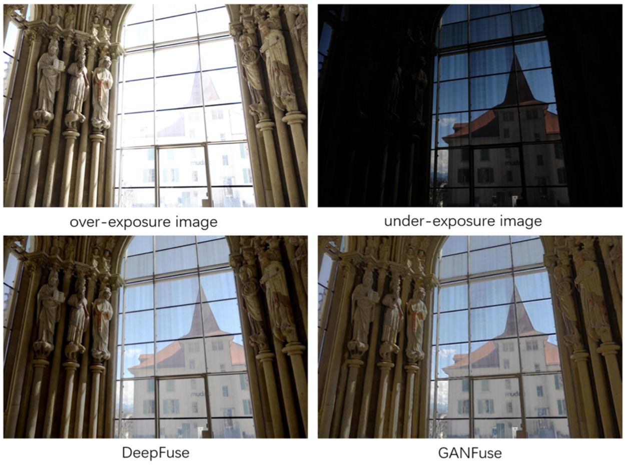

**Fig. 1** Schematic illustration of image fusion

- A GAN-based unsupervised image fusion algorithm for fusing extreme-exposure images is proposed. The adversarial relationship enables fused image to have more details from source images.
- Different from FusionGAN, a novel structure of GAN is developed, which is more suitable for the task of MEF.
- We design a new loss function for MEF which can help fused image to preserve more information from source images.
- We construct a new training dataset which contains all kinds of conditions. This dataset could enhance the robustness of our method.

The rest parts of this paper are listed as follows. In Sect. 2, we briefly review related works from the literatures. In Sect. 3, we introduce our proposed GANFuse, including the architecture, loss function, training and testing processes. The results of comparison experiments are presented in Sect. 4. And the conclusion is described in Sect. 5.

## 2 Related works

This section provides a brief summary of existing image fusion methods based on deep learning. Furthermore, in consideration of our fusion method is based on GAN, we will discuss the basic theory of GAN, representative variants of GAN and their applications.

### 2.1 Fusion methods based on deep learning

In recent years, since the deep learning has aroused extensive concern, deep learning has been applied in image fusion, due to its outstanding ability of feature extraction and universality. The main research theories are divided into the following three categories. (1) Methods combine traditional methods with deep learning. In these methods, deep learning framework functions as a tool to extract image features. Representatively, Liu et al. [20] decomposes the source images into detail layer and base layer, and then utilizes convolutional sparse representation (CSR) to merge these layers. Finally, the fused image is reconstructed by the fused base layer and detail layer. In IFCNN [15], Liu et al. proposed a universal network for image fusion and designed different fusion rule according to different type of source images. (2) These methods regard the convolutional network function as a way to generate weight map which shows the importance of each pixel from source images. For instance, Li et al. [21] uses the VGGNet to extract image features and construct a robust weight map for fusion. (3) Other methods present an end-to-end learning framework for image fusion. Prabhakar et al. [13] proposes an unsupervised deep learning framework for multi-exposure fusion. They utilize a novel CNN architecture and designed a no-reference quality metric as the loss function. In FusionGAN and its variants [16, 22, 23], a generative adversarial network is applied to fuse infrared and visible images. The fused image generated by the generator is forced to have more details existing in the visible image by applying the discriminator to distinguish differences between them.

Although these works have achieved promising progress, there are still some drawbacks. (a) Many existing methods use neural network to extract features and reconstruct these features. However, fusion rules are designed manually. Thus, these methods still have limitations of traditional fusion methods. (b) Unsupervised deep learning methods are implemented by designing a suitable loss function. However, finding an effective loss function is still a challenge. (c) Existing GAN-based fusion methods applied discriminator to force fused image to contain more details in one of the source image, leading to the loss of information from the other source image.

To address these drawbacks, we research an approach to MEF that can preserve more effective information from source images under the framework of GAN. Motivated by the success of the FusionGAN on infrared and visible image fusion, we aim to develop the structure of FusionGAN and make our structure suitable for MEF. In general, there are three improvements in our method.

1. FusionGAN feeds visible image and fused image into discriminator. Actually, fused image contains not only visible information, but also infrared information. Therefore, it is easy for discriminator to distinguish between the visible image and fused image. According to the principle of GAN, the stronger distinguishing ability the discriminator has, the better the fused image generated by the generator performs. To improve discriminator's ability, we want to find a way to get the contribution of source images in the fused image and set these contributions and fused image as input of discriminators. By coincidence, this idea is included in the SCD loss function [24]. The main idea of SCD is that the difference between the fused image ($F$) and the source image ($S_1$) represents the contribution of the source image ($S_2$) and vice versa. Therefore, we think that the difference image between the fused image ($F$) and the input image ($S_2$) almost contains the information transmitted from another input image ($S_1$). Consequently, in our networks, by feeding $|F - S_2|$ and $S_1$ into discriminator 1 and $|F - S_1|$ and $S_2$ into discriminator 2, we make it difficult for the discriminators to distinguish the input data and makes the adversarial relationship between the two discriminators and the generator more fierce. The proposed network overall architecture is shown in Fig. 2.
2. In the generator, instead of using concat operation to fuse feature maps which is applied in FusionGAN, we choose tensor addition as the fusion rule. It is due to the fact that purpose of MEF is to get the well-exposure image whose exposure value is the average of the source image. According to this theory, IFCNN uses the elementwise-mean fusion rule [15] to fuse multi-exposure images. But simply using elementwise-mean fusion rule may cause the loss of information from source images. Therefore, we choose tensor addition to merge feature maps. Average operation is done by the follow-up networks.
3. As mentioned in Sect. 1, DeepFuse sets SSIM as loss function. Due to the fact that DeepFuse is the first work that uses deep CNN architecture for MEF, following with DeepFuse, many existing deep learning-based image fusion methods employ the metric SSIM as the loss function [15, 20, 25]. However, simply depending on SSIM to constrain whole network leads to loss of other information. As we know, the most important

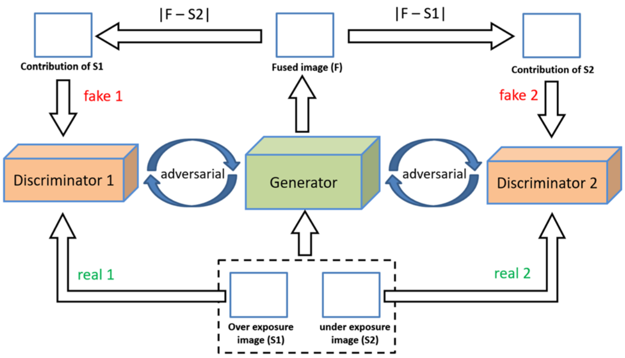

**Fig. 2** Overall architecture of GANFuse

information in an image is texture information and luminance information. Consequently, to preserve these key information in fused image, we include the gradient loss and pixel intensities loss as the loss function. Moreover, in the experimental result section, we use SSIM as one of metrics to evaluate the fused image, and our result shows the highest value among the comparison methods.

## 2.2 The basic theory of GAN

Generative adversarial net was initially proposed by IanJ Goodfellow et al. in 2014 [26]. Different from conventional neural networks, network training requires a generator ($G$) and a discriminator ($D$) to work simultaneously. This framework corresponds to a minimax two-player game. Game players are generator and discriminator network. During training step, the ability of $G$ and $D$ are gradually improved until the two sides to achieve equilibrium. Given the input variable ($x$), the generator $G$ is used to generate output $y = G(x)$. Through training process, $G$ can learn a training distribution $P_G(x)$ which approximate to real data distribution $P_{\mathrm{Data}}(x)$. Then, the discriminator $D$ is trained to determine whether the input is from $P_{\mathrm{Data}}(x)$ or $P_G(x)$. The purpose of $G$ is to generate a fake data which can fool the $D$. However, $D$ aims at differentiating between real data and fake data. Through this adversarial relationship, the distribution generated by $G$ will gradually approximate the real data.

The optimization formulation of $G$ is formulated as:

$$G = \arg\min_{G} \mathrm{Div}(P_G(x), P_{\mathrm{Data}}(x)), \tag{1}$$

where Div denotes the divergence between $P_{\mathrm{Data}}(x)$ and $P_G(x)$. The function of $D$ can be expressed as:

$$D = \arg\max_{D} V(G, D), \tag{2}$$

where $V(G, D)$ is defined as follows:

$$\begin{aligned} V(G,D) = {} & E_{x\sim P_{\mathrm{data}}}[\log P(x)] \\ & + E_{x\sim P_G}[\log(1 - P(x))]. \end{aligned} \tag{3}$$

Thus, the optimization formulation of generative adversarial network can be expressed as:

$$G = \arg\min_{G} \max_{D} V(G, D). \tag{4}$$

$G$ and $D$ are alternately trained. With the advance of the adversarial process, the data generated by $G$ will be gradually similar to the real data.

## 2.3 Variants of GAN and their applications

GAN is a novel network which can generate more real-like data. However, GAN suffers from unstable training. Since the year of 2014, several works have attempted to solve this problem. For example, deep convolutional GAN [27] defines a set of constraints on the architecture of GAN that makes their model stable to train. For optimizing the unreasonable divergence measurement in original GAN, WGAN [28] introduces the Wasserstein distance to improve the stability of training. To overcome the vanishing gradients problem caused by loss function, least squares GAN (LSGAN) [29] adopts the least squares loss function for the discriminator. StyleGAN [30] embeds the input latent code into an intermediate latent space and proposes two new distribution quality metrics for generator architecture that makes their model more linear, less entangled representation of variation. Conditional GAN (cGAN) extends GAN to a conditional model by feeding auxiliary information such as class labels or data from other modalities into the discriminator and generator [31]. For translating clothing images between two specific clothing categories, Liu et al. [32] proposes category-attribute GAN (CA-GAN) framework, including three discriminators. Overall, in the future, GANs have the potential to apply in many fields.

# 3 Proposed method

The color conversion of the proposed fusion model is presented in Fig. 3. We decomposed source images into three channels, Y, Cb and Cr. The model we proposed is used for fusing Y channel of source images since the texture details of image are mainly presented by luminance channel (Y) of image. The fusion rule for chrominance channels (Cb and Cr) will be introduced in Sect. 3.5. The architecture of networks, loss function, training and testing processes will be described in the remainder of this section.

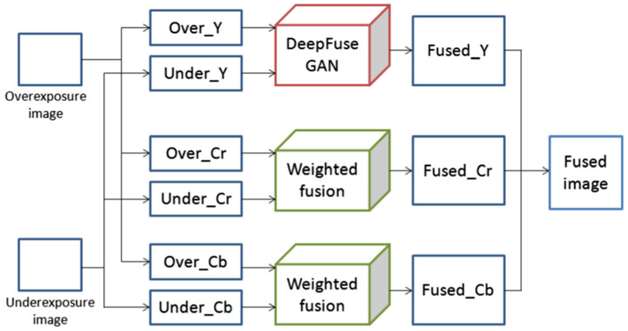

**Fig. 3** The whole procedure of the proposed fusion model

### 3.1 GANFuse

Learning ability of GAN is depending on the structure of network and the loss function. There are three differences between the GANFuse and FusionGAN. Firstly, we train two discriminators to optimize the generator that makes the fused image to contain more details in extreme exposure image pairs. Secondly, we design a new input mode to improve the discretion ability of discriminators. Thirdly, according to the purpose of MEF, we set pixel intensity loss and gradient loss of source images as generator's loss function. The proposed network overall architecture is shown in Fig. 2. The Y channel of under-exposure image and the Y channel of over-exposure image are fed to generator ($G$), and the output of the generator is the Y channel of the fused image.

As mentioned in Sect. 2.1, we input $|F - S_2|$ and $S_1$ into discriminator 1 $(D_1)$ to distinguish between contribution of $S_1$ in fused image and $S_1$. In the meantime, we input $|F - S_1|$ and $S_2$ into discriminator 2 $(D_2)$ to distinguish between contribution of $S_2$ in fused image and $S_2$. This input model enhances the differentiating capacity of discriminators, which force generator to generate real-like fused image. In the training phase, the two discriminators are trained simultaneously by testing them against the $G$. After training procedure, $D_1$ cannot differentiate between the contribution of $S_1$ in fused image and $S_1$, and $D_2$ cannot differentiate between the contribution of $S_2$ in fused image and $S_2$.

### 3.2 Loss function

The loss function contains two parts, losses of generator and discriminators. The details are presented as follows. The loss function of generator ($G$) consists of over-exposure image's loss $L_{I_\text{o}}$ and under-exposure image's loss $L_{I_\text{u}}$.

$$L_G = L_{I_\text{o}} + \gamma L_{I_\text{u}}, \tag{5}$$

where $\gamma$ is used to control the trade-off between over-exposure image and under-exposure image loss. $L_{I_\text{o}}$ is defined as follows:

$$L_{I_\text{o}} = L_{I_\text{o}}^{\text{adv}} + \alpha L_{I_\text{o}}^{\text{con}}, \tag{6}$$

where $\alpha$ is a weight controlling the trade-off between two terms. $L_{I_\text{o}}^{\text{con}}$ denotes the content loss of over-exposure image, which aims to save the over-exposure image information in the fused image. As mentioned in Sect. 2, we aim to reserve the gradient information and pixel intensities information in fused image. Therefore, $L_{I_\text{o}}^{\text{con}}$ is defined as follows:

$$\begin{aligned} L_{I_\text{o}}^{\text{con}} = \frac{1}{h \cdot w} \big[ & \text{sum}(\|I_\text{f} - I_\text{o}\|_F) \\ & + \sigma \cdot \text{sum}(\|\Gamma I_\text{f} - \Gamma I_\text{o}\|_F) \big], \end{aligned} \tag{7}$$

where the weight $\sigma$ is used to control the trade-off. $h$ and $w$ presents the height and width of the source image. sum represents element summation of the input. $\|\cdot\|_F$ is the matrix Frobenius norm, $\Gamma$ denotes the gradient operation. $L_{I_\text{o}}^{\text{adv}}$ conveys the adversarial loss between $G$ and $D_1$, which is defined as follows:

$$L_{I_\text{o}}^{\text{adv}} = \frac{1}{h \cdot w} \cdot \text{sum}[-D_1(|I_\text{f} - I_\text{u}|)]. \tag{8}$$

In order to establish an adversarial relationship between discriminators and generator, we set a negative sign in front of $D_1$.

The second term of $L_G$ presents the loss of under-exposure image, which is defined as follows:

$$L_{I_\text{u}} = L_{I_\text{u}}^{\text{adv}} + \beta L_{I_\text{u}}^{\text{con}}, \tag{9}$$

where the weight $\beta$ is used to control the trade-off. Similarly, $L_{I_\text{u}}^{\text{con}}$ is the content loss of $I_\text{u}$ and $I_\text{f}$, which is defined as follows:

$$\begin{aligned} L_{I_\text{u}}^{\text{con}} = \frac{1}{h \cdot w} \big[ & \text{sum}(\|I_\text{f} - I_\text{u}\|_F) \\ & + \sigma \cdot \text{sum}(\|\Gamma I_\text{f} - \Gamma I_\text{u}\|_F) \big]. \end{aligned} \tag{10}$$

$L_{I_\text{u}}^{\text{adv}}$ conveys the adversarial loss between $G$ and $D_2$, which is defined as follows:

$$L_{I_\text{u}}^{\text{adv}} = \frac{1}{h \cdot w} \cdot \text{sum}[-D_2(|I_\text{f} - I_\text{o}|)]. \tag{11}$$

Discriminator shortens the difference between fused image and source images. The adversarial loss of $D_1$ and $D_2$ judge the similarity of source images and fused image. The loss function of $D_1$ and $D_2$ are shown as follows:

$$\begin{aligned} L_{D_1} = & \frac{1}{h \cdot w} \cdot \text{sum}[D_1(|I_\text{f} - I_\text{u}|)] \\ & - \frac{1}{h \cdot w} \cdot \text{sum}[D_1(I_\text{o})], \end{aligned} \tag{12}$$

$$\begin{aligned} L_{D_2} = & \frac{1}{h \cdot w} \cdot \text{sum}[D_2(|I_\text{f} - I_\text{o}|)] \\ & - \frac{1}{h \cdot w} \cdot \text{sum}[D_2(I_\text{u})]. \end{aligned} \tag{13}$$

We regard $D_1(|I_\text{f} - I_\text{u}|)$ and $D_2(|I_\text{f} - I_\text{o}|)$ as fake data which is decreased by discriminator, and regard $D_1(I_\text{o})$ and $D_2(I_\text{u})$ as real data which is increased by discriminator. Thus, there was a negative sign in front of real data.

## 3.3 Network structure

From Fig. 2, we can see that the whole network structure consists of two discriminators $(D_1$ and $D_2)$ and one generator $(G)$. In this section, the structure of $(D_1,\ D_2)$ and $G$ will be introduced.

### 3.3.1 Generator

The structure of $G$ consists of three parts, namely, feature extraction layers, fusion operation and reconstruction layers, as illustrated in Fig. 4. The function of feature extraction layer is to get features from source images. We use the same feature extraction layer to get features of under-exposure image and over-exposure image. Therefore, we can add these extracted information, and then fed them into the reconstruction layer. The output of the reconstruction model is the fused image.

Owing to the random initialized kernels, training the end-to-end model is unstable and difficult. An effective way to handle this issue is using a well-trained feature extraction model [33, 34]. Thus, we choose pre-trained Resnet V1 [35] as the feature extraction layers. It learns residual representations between inputs and outputs by using multiple parametric layers, which can avoid vanishing gradient. As is shown in Fig. 4, our feature extraction layers has five bottlenecks. And n48 on bottleneck 1 denotes that the depth of bottleneck 1 is 48. The architecture of each bottleneck is illustrated in Fig. 5. For avoiding loss information in extreme exposure image pairs, we set the stride of all kernels to 1. Reconstruction layers comprise five CNN layers. Batch normalization and ReLU are applied to alleviate gradient exploding and accelerate the training.

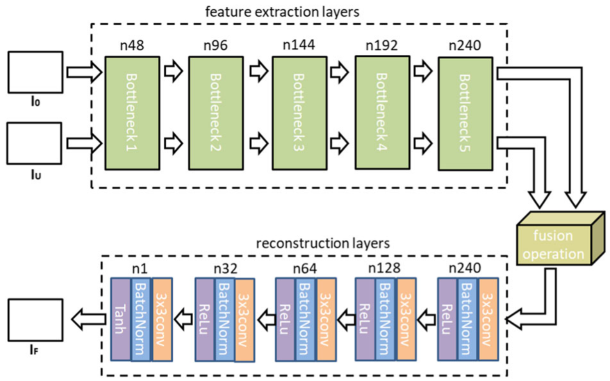

**Fig. 4** The structure of generator

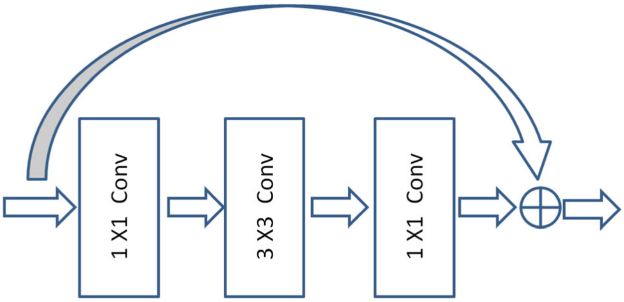

**Fig. 5** The network architecture of the bottleneck

### 3.3.2 Discriminator

By designing discriminators, the details of fused image are more similar to under-exposure image and over-exposure image. The networks of these two discriminators have the same network, which is shown in Fig. 6. And the stride of all layers is set to 2 without padding.

## 3.4 Training

As for the training data set, we collect 30 pairs of exposure stacks which are available publicly from the Internet [36]. It contains all kinds of conditions. Due to the huge amount of source data, we down-sample the source images and crop them into 7552 patch pairs with the size of 84 × 84. We set the learning rate to $10^{-4}$ and train the network for 5 epochs with all the training patches being processed in each epoch.

## 3.5 Testing

After training phase, we can get the fused image in the Y channel. The chromaticity channels of fused image are got by weighting sum of input chromaticity channel values. The main information is presented in the Y channel. Thus, different fusion strategies are applied in literature for Y, Cb and Cr fusion [13, 37]. We can choose different methods to merge RGB channels. However, there is usually a substantial correlation between the RGB channels. Therefore, fusing source image in RGB channels will ignore this correlation and cause obvious color difference. We merge the chromaticity channels of the source image by following the strategy of Prabhakar [37], which is shown as follows:

$$x = \frac{x_1|x_1-\tau| + x_2|x_2-\tau|}{|x_1-\tau| + |x_2-\tau|}, \tag{14}$$

where the $x_1$ and $x_2$ denote the pixel intensities of image pairs. The fused chrominance value is obtained by weighing two chrominance values with $\tau$ subtracted value from itself. In our work, the value of $\tau$ is set to 128. The

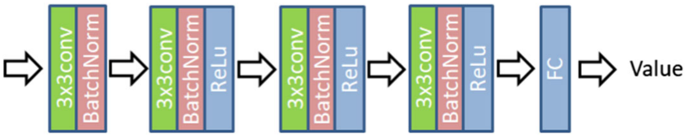

**Fig. 6** The network architecture of the discriminator

final step is converting $\mathrm{Y}_{\mathrm{fuse}}$, $\mathrm{Cb}_{\mathrm{fuse}}$ and $\mathrm{Cr}_{\mathrm{fuse}}$ channels into RGB image.

## 4 Experiments

We have conducted extensive evaluation and comparison study against state-of-the-art algorithms. For verifying the effect of the experiment, we select the images pairs as the test set with different conditions, including indoor and outdoor, day and night, natural and artificial lighting. To evaluate the performance of algorithms objectively, we adopt five types of metrics. All the experiments are conducted on a desktop with 2.4 GHz Intel Xeon CPU E5-2673 v3, GeForce GTX 2080Ti, and 64 GB memory.

### 4.1 Comparison methods

The method we proposed is compared against with five representative methods, including GFF [38], DSIFT [39], FLER [40], the gradient-based method (GBM) [36], DeepFuse [13]. GFF is a novel guided filtering-based fusion method for creating a highly informative fused image [38]. In DSIFT, the dense SIFT descriptor is applied as the activity level measurement to extract information from source images [39]. FLER proposed a strategy which brightens the high-light regions in the dark image and darkens the darkest regions in the bright image and finally generates virtual image via intensity mapping functions [40]. In the GBM, two different fusion strategies are applied for chrominance and luminance channels separately [36]. DeepFuse is the landmark multi-exposure fusion method based on deep learning [13].

### 4.2 Qualitative comparisons

We firstly perform qualitative comparison experiments on three typical image sequences. Fused results of our method and five comparison methods are shown in Figs. 7, 8 and 9. In this paper, we evaluate the effect of image fusion from two aspects, the overall image visual effect and the detail effect of the image. From the aspect of the overall visual effect, the method we proposed is well proportioned in light distribution and closer to the actual scene. There are local dark areas in the image of compared methods, which will lead to the loss of detail features. As for the detail effect, our method can provide additional texture information in some regions.

From Figs. 7, 8 and 9, we can see the results of the three methods of GFF, DSIFT and FLER, these methods have obvious black regions in the fused image. The fusion results of GBM and DeepFuse are more consistent with human visual perception. However, as we have shown in the red box in the ground truth, they also have some loss of detail textures. To be specific, as can be seen in Fig. 7, the details of the tree in our method are of abundant texture information. The same phenomenon can be found in Fig. 8. In the red box, our method shows that the texture on the wall is more clear and the outline of texture is closer to ground truth. As for the window we marked in Fig. 9, there is a bird pattern in the center part of the mark land. The bird pattern of ours shows the most colorful and clear result.

### 4.3 Quantitative comparisons

In the multi-exposure image fusion community, MEF-SSIM [36] is a commonly used metric for quantitative evaluation. In addition, we select SD, PSNR, CC and SCD as metrics. These methods are commonly used in MEF evaluation. We apply these metric to evaluate the source images and fused results of five comparisons methods. These five metrics are introduced as follows.

#### 4.3.1 Standard deviation (SD)

SD is a metric reflecting contrast and distribution of images. Due to the fact that human pays more attention to the region with high contrast. Thus, the larger value of SD means the higher contrast of the fused image. SD is defined as follows:

$$\mathrm{SD} = \sqrt{\frac{1}{w \cdot h}\sum_{i=1}^{h}\sum_{j=1}^{w}(f_{i,j} - \mu_{\mathrm{f}})^2}, \qquad (15)$$

where $h$ and $w$ denotes the height and width of image. $\mu_{\mathrm{f}}$ denotes the mean value of image $f$.

#### 4.3.2 Peak signal-to-noise ratio (PSNR)

PSNR is a metric reflecting the distortion by the ratio of peak value power and noise power.

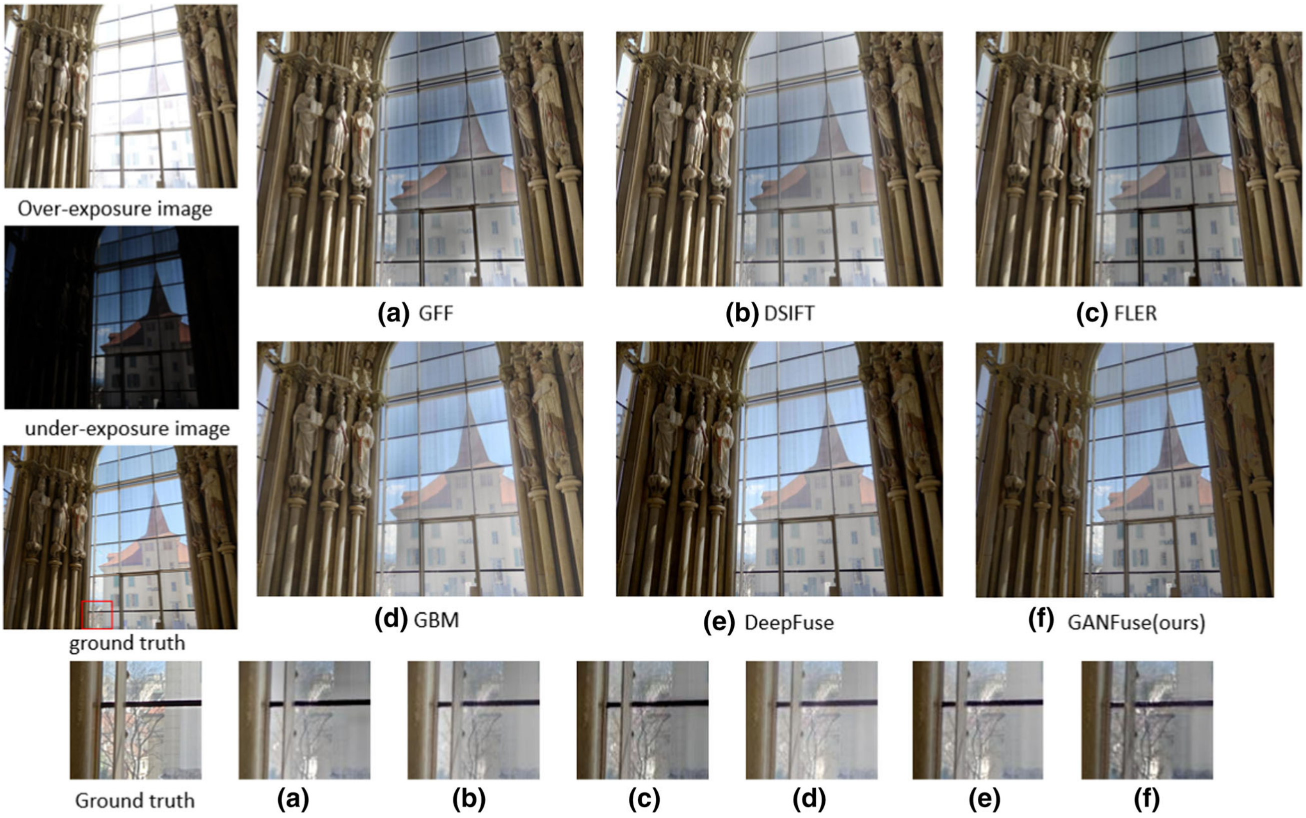

**Fig. 7** Qualitative comparison results on image sequence 1

$$\mathrm{PSNR} = 10 \cdot \log_{10} \frac{r^2}{\mathrm{MSN}}, \tag{16}$$

where $r$ is the max value of the fused image. $r$ is set as 256 in this paper. MSE is the mean square error that measures the dissimilarity between the source images and the fused image that is defined as follows:

$$\mathrm{MSE} = \frac{1}{w \cdot h} \sum_{i=1}^{h} \sum_{j=1}^{w} (f_{i,j} - g_{i,j})^2. \tag{17}$$

A larger PSNR indicates the less distortion between source images and fused image.

#### 4.3.3 Correlation coefficient (CC)

CC measures the correlation between the source images and the fused image. It is mathematically defines as follows:

$$\mathrm{CC} = \frac{\sum_{i=1}^{h} \sum_{j=1}^{w} (f_{i,j} - g_{i,j})(f_{i,j} - \mu_{i,j})}{\sqrt{\left(\sum_{i=1}^{h} \sum_{j=1}^{w} (f_{i,j} - g_{i,j})^2\right)\left(\sum_{i=1}^{h} \sum_{j=1}^{w} (f_{i,j} - \mu_{i,j})^2\right)}}. \tag{18}$$

A large CC indicates that there is a strong correlation between the fused image and the source images.

#### 4.3.4 Mean structural similarity (MSSIM)

MSSIM measures the average of the individual SSIM values for each sliding window. SSIM is a metric used to model image loss and distortion. It is defined as follows:

$$\mathrm{SSIM}_{x,f} = \sum_{x,f} \frac{2\mu_x \mu_\mathrm{f} + c_1}{\mu_x^2 + \mu_\mathrm{f}^2 + c_1} \cdot \frac{2\sigma_x \sigma_\mathrm{f} + c_2}{\sigma_x^2 + \sigma_\mathrm{f}^2 + c_2} \cdot \frac{\sigma_{xf} + c_3}{\sigma_x \sigma_\mathrm{f} + c_3}, \tag{19}$$

where $x$ and $f$ are the image patches of the source image $X$ and the fused image $F$, respectively, $\sigma$ denotes the covariance or the standard deviation. $\mu$ denotes the mean values. $C1$, $C2$ and $C3$ are the parameters for stability.

#### 4.3.5 The sum of the correlations of differences (SCD)

In the SCD loss function, the difference image between one of the input images ($S_2$) and the fused image ($F$) almost discloses the information transferred from the other input image ($S_1$). These differences ($D_1$ and $D_2$) can then be formulated as:

$$D_1 = F - S_2, \tag{20}$$

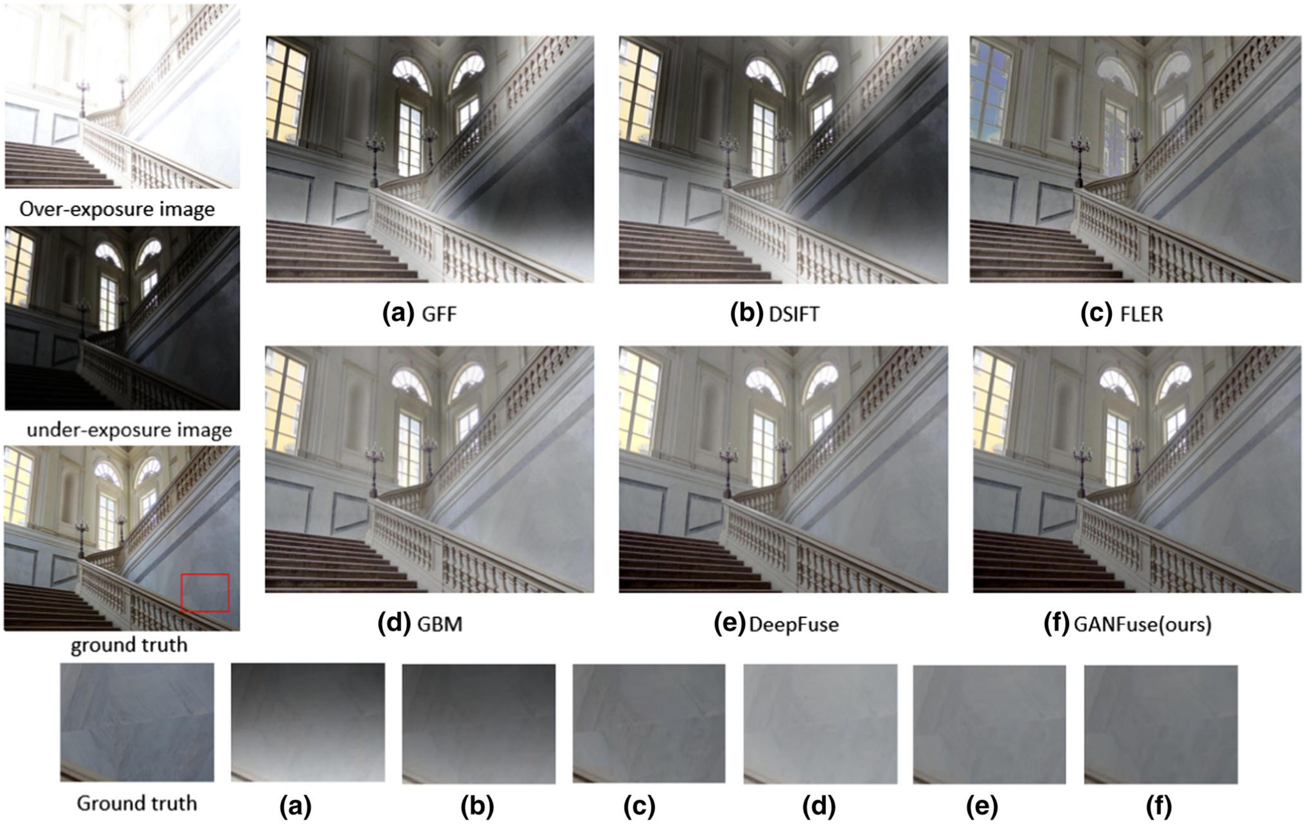

**Fig. 8** Qualitative comparison results on image sequence 2

$$D_2 = F - S_1. \tag{21}$$

The $D_1$ and $D_2$ indicates the amount of transferred information from each of the input images into the fused image. SCD loss function is formulated as the following:

$$\mathrm{SCD} = r(D_1, S_1) + r(D_2, S_2). \tag{22}$$

$r(D_k, S_k)$ is to calculate the similarity between $D_k$ and $S_k$, which is defined as the following:

$$r(D_k, S_k) = \frac{\sum_i \sum_j (D_k(i,j) - \bar{D_k})(S_k(i,j) - \bar{S_k})}{\sqrt{(\sum_i \sum_j (D_k(i,j) - \bar{D_k})^2)(\sum_i \sum_j (S_k(i,j) - \bar{S_k})^2)}}, \tag{23}$$

where $\bar{D_k}$ and $\bar{S_k}$ are the average of the pixel values of $D_k$ and $S_k$.

In addition, these metrics can only handle single-channel images. Thus, we perform these metrics on $Y$ channel. We test these five metrics on 30 multi-exposure image pairs, and the results are presented in Fig. 10. And the results of mean value are shown in Table 1. The red and blue index we marked in Table 1 is the largest value and second largest value, respectively.

Table 1 presents that our method can perform a good result. Our method gets the largest mean value in CC, PSNR, SSIM and SCD. SD of GFF and DSIFT own the first and second place, respectively. However, these methods have the phenomenon of inhomogeneous illumination which will result in a high value of SD. Our method achieved the largest mean values among the rest methods.

### 4.4 Comparative experiment

To prove the effect of creations in our framework, we perform an ablation study on the components of GANFuse. The comparative experiment 1 shows the result of GANFuse without discriminators. In the comparative experiment 2, removing the theory of SCD, we directly feed $F$ and $S_1$ into discriminator 1 and $F$ and $S_2$ into discriminator 2. The qualitative result is displayed in Fig. 11. It is obvious that the result of Fig. 11b is most similar to ground truth, including color and textural detail. Particularly, as presented in red box, tableware of Fig. 11b has better visual effect. And the quantity result is given in Table 2 which also testify to the effect of creations of our method.

Moreover, we perform a parametric study on those important parameters in our GANFuse. In our GANFuse, we mainly set the weight $\sigma$ to trade off the gradient loss and pixel intensities loss. In the following part, we will show results of different $\sigma$. In our model, we set $\sigma$ as 0.1.

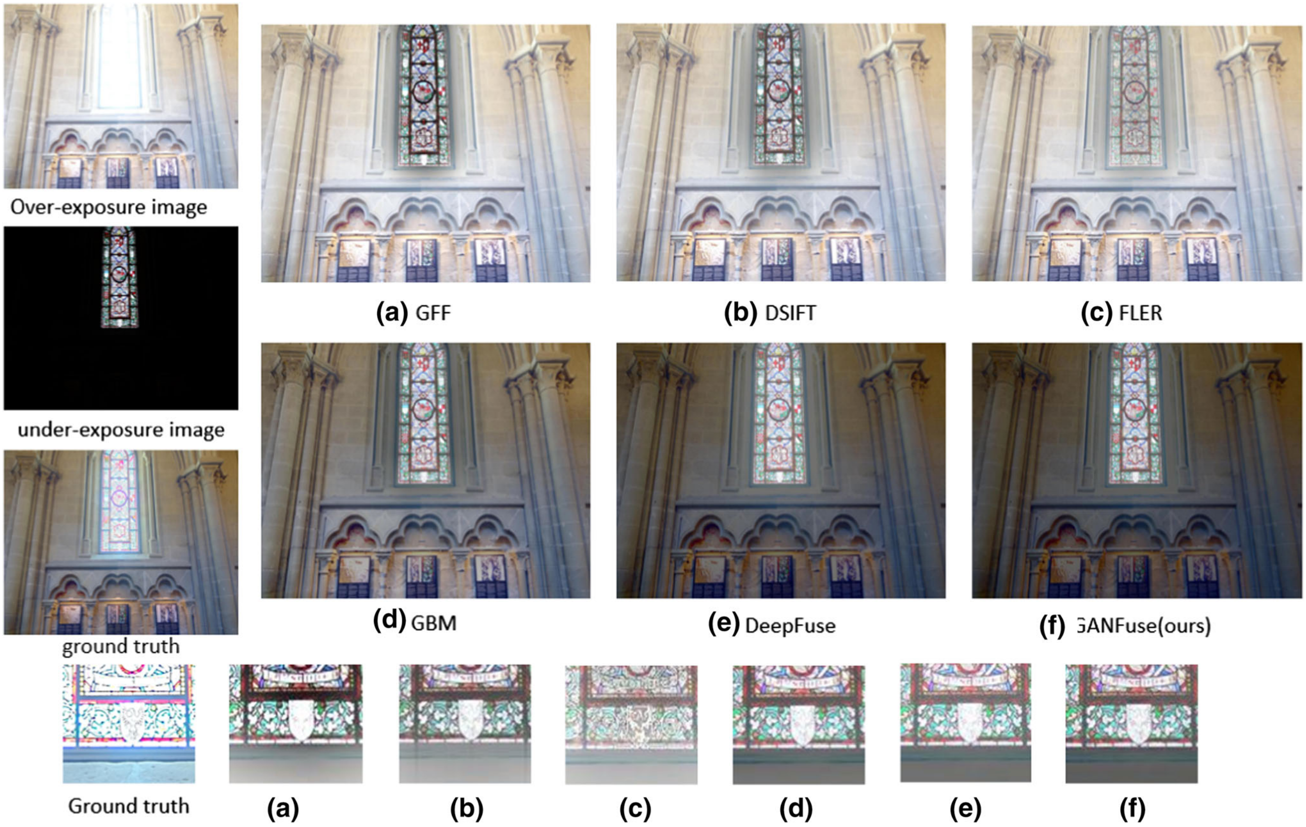

**Fig. 9** Qualitative comparison results on image sequence 3

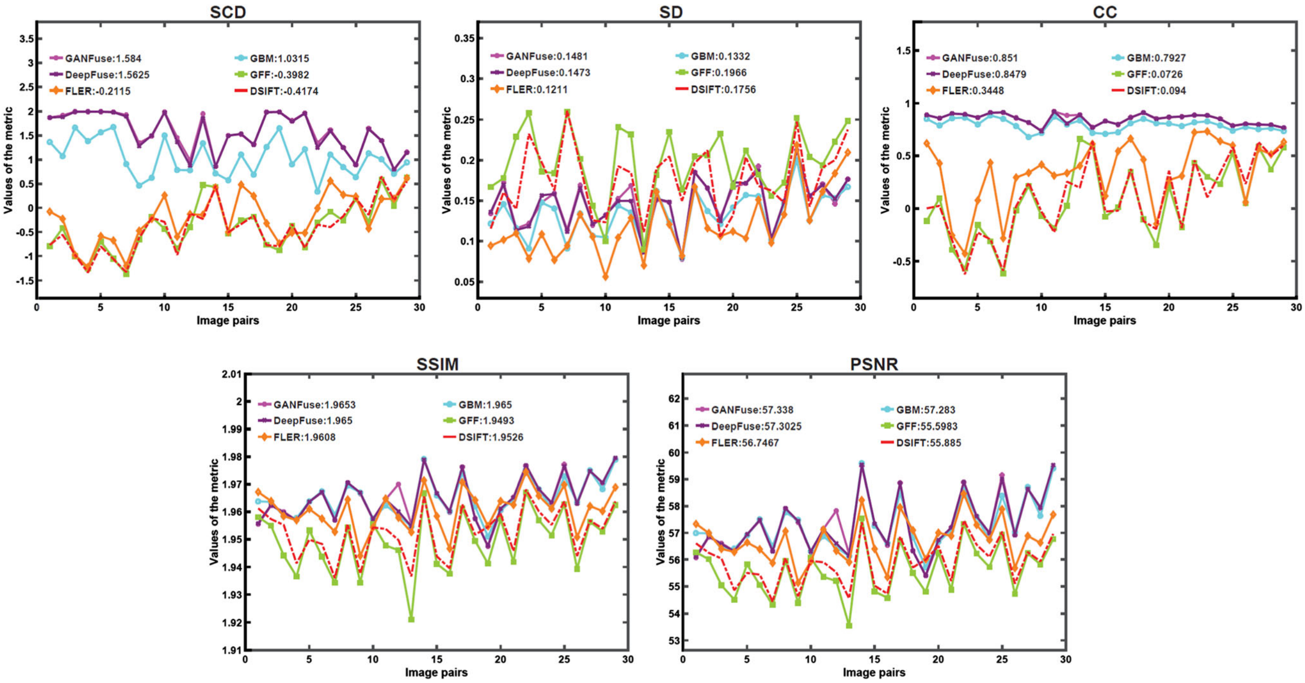

**Fig. 10** Qualitative comparison of our GANFuse with 5 state-of-the-art methods on five metrics

**Table 1** Quantitative comparisons of the five metrics

| Metrics | SD | PSNR | SSIM | CC | SCD |
|---|---|---|---|---|---|
| GANFuse | 0.148115426 | 57.33800149 | 1.965263834 | 0.850954307 | 1.584032931 |
| GBM | 0.13316501 | 57.28301727 | 1.96499124 | 0.792732864 | 1.031510881 |
| DeepFuse | 0.147293475 | 57.30245913 | 1.964967841 | 0.847856712 | 1.56252029 |
| GFF | 0.196556506 | 55.59830182 | 1.949333007 | 0.072599073 | − 0.398211353 |
| FLER | 0.121066135 | 56.74671019 | 1.960781543 | 0.344774143 | − 0.211503979 |
| DSIFT | 0.175588204 | 55.88504831 | 1.952625491 | 0.094026786 | − 0.417428915 |

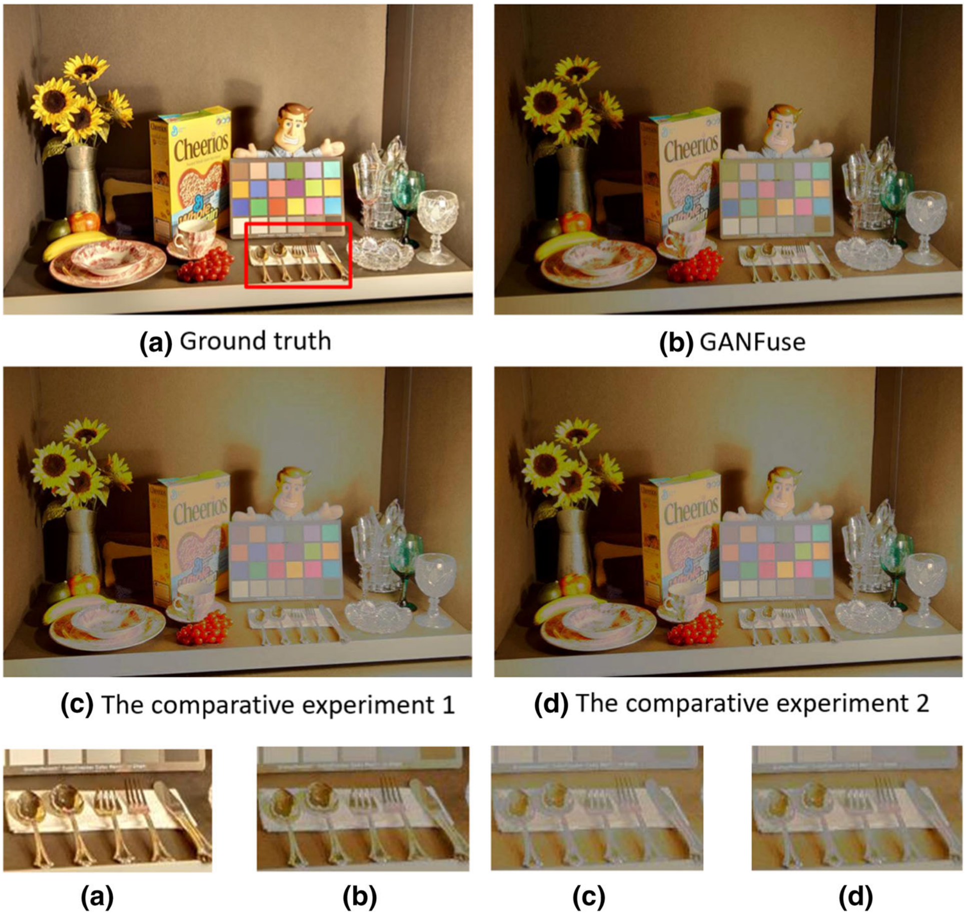
(a) Ground truth
(b) GANFuse
(c) The comparative experiment 1
(d) The comparative experiment 2
(a)
(b)
(c)
(d)

**Fig. 11** The qualitative results of the comparative experiments

**Table 2** The quantity result in comparative experiments

| Metrics | SD | PSNR | SSIM | CC | SCD |
|---|---|---|---|---|---|
| GANFuse | 0.148115426 | 57.33800149 | 1.965263834 | 0.850954307 | 1.584032931 |
| The comparative experiment 1 | 0.148030308 | 57.30291199 | 1.964888646 | 0.850155029 | 1.574653232 |
| The comparative experiment 2 | 0.139084259 | 57.30443939 | 1.964849778 | 0.846627754 | 1.449303726 |

And the weight $\sigma$ is 0.3 and 0.5 in the comparative experiment 3 and the comparative experiment 4, respectively. The quantity result is presented in Table 3. In Table 3, we can find that GANFuse owns the best quantity result when the value of $\sigma$ is 0.1. Due to the fact that the values of the result is approximate, the quality results are almost same. Therefore, the quality results will not show anymore.

**Table 3** The quantity result in comparative experiments

| Metrics | SD | PSNR | SSIM | CC | SCD |
|---|---|---|---|---|---|
| GANFuse | 0.148115426 | 57.33800149 | 1.965263834 | 0.850954307 | 1.584032931 |
| The comparative experiment 3 | 0.147009365 | 57.33418144 | 1.965185013 | 0.850091644 | 1.565150411 |
| The comparative experiment 4 | 0.14853375 | 57.33491155 | 1.965202621 | 0.849739465 | 1.579227282 |

## 5 Conclusion and future work

In this paper, we propose a novel GAN-based multi-exposure image fusion method, termed as GANFuse. On the basis of FusionGAN, we increase the number of discriminator and propose a novel way to change the input of discriminators. By doing so, we can preserve more information of source images in the fused image. Furthermore, we train and test our networks with all kinds of dataset. Thus, our method can achieve better robust with different conditions. Compared with other five state-of-the-art fusion methods, our method can achieve advanced performance both qualitatively and quantitatively. In our current work, GANFuse is trained to fuse static multi-exposure images. However, for moving objects in image, our method does not possess a good visual effect. Moving objects may lead the ghost phenomenon in fused image. For future research, we aim to handle the ghost phenomenon and generalize GANs or their variants to fuse multi-modal images.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China under Grant 51775215.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Wandell B (1995) Foundations of vision, vol 42, p 01
2. Ma K, Zeng K, Wang Z (2015) Perceptual quality assessment for multi-exposure image fusion. IEEE Trans Image Process 24(11):3345–3356
3. Burt PJ (1984) The pyramid as a structure for efficient computation. Springer Series in Information Sciences, vol 12, pp 6–35
4. Vonikakis V, Bouzos O, Andreadis I (2011) Multi-exposure image fusion based on illumination estimation. Springer Series in Information Sciences, pp 135–142
5. Song M, Tao D, Chen C, Bu J, Luo J, Zhang C (2012) Probabilistic exposure fusion. IEEE Trans Image Process A Publ IEEE Signal Process Soc 21(1):341
6. Li Z, Zheng J, Zhu Z, Wu S (2014) Selectively detail-enhanced fusion of differently exposed images with moving objects. IEEE Trans Image Process 23(10):4372–4382
7. Lee S, Sung PJ, Ik CN (2018) A multi-exposure image fusion based on the adaptive weights reflecting the relative pixel intensity and global gradient, pp 1737–1741
8. Wang J, Xu G, Lou H (2015) Exposure fusion based on sparse coding in pyramid transform domain. In: ACM Press the 7th international conference, pp 1–4
9. Xu J, Huang Y (2013) Multi-exposure images of wavelet transform fusion. In: Proceedings of the SPIE, p 8878
10. Abdelkader A, Eldin MH, Ebrahim RS (2011) Performance measures for image fusion based on wavelet transform and curvelet transform, pp 1–7
11. Goshtasby A (2005) Fusion of multi-exposure images. Image Vis Comput 23(6):611–618
12. Ma K (2015) Multi-exposure image fusion: a patch-wise approach. In: IEEE International conference on image processing (ICIP), pp 1717–1721
13. Ram Prabhakar K, Sai Srikar V, Venkatesh Babu R (2017) Deepfuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs. In: IEEE international conference on computer vision, vol 505, pp 4724–4732
14. Liu Y, Chen X, Peng H, Wang Z (2017) Multi-focus image fusion with a deep convolutional neural network. Inf Fusion 36:191–207
15. Zhang Y, Liu Y, Sun P, Yan H, Zhao X, Zhang L (2020) IFCNN: a general image fusion framework based on convolutional neural network. Inf Fusion 54:99–118
16. Ma J, Yu W, Liang P, Li C, Jiang J (2019) Fusiongan: a generative adversarial network for infrared and visible image fusion. Inf Fusion 48:11–26
17. Xu H, Ma J, Zhang X-P (2020) MEF-GAN: multi-exposure image fusion via generative adversarial networks. IEEE Trans Image Process 29:7203–7216
18. Han X, Ma J, Jiang J, Guo X, Ling H (2020) U2fusion: a unified unsupervised image fusion network. IEEE Trans Pattern Anal Mach Intell
19. Li S, Kang X, Fang L, Hu J, Yin H (2017) Pixel-level image fusion: a survey of the state of the art. Inf Fusion 33:100–112
20. Liu Y, Chen X, Ward RK, Wang ZJ (2016) Image fusion with convolutional sparse representation. IEEE Signal Process Lett 23(12):1882–1886

21. Li H, Wu X, Kittler J (2018) Infrared and visible image fusion using a deep learning framework. In: 2018 24th international conference on pattern recognition (ICPR), pp 2705–2710
22. Ma J, Liang P, Yu W, Chen C, Guo X, Wu J, Jiang J (2020) Infrared and visible image fusion via detail preserving adversarial learning. Inf Fusion 54:85–98
23. Ma J, Han X, Jiang J, Mei X, Zhang X (2020) DDCGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion. IEEE Trans Image Process 29:4980–4995
24. Aslantas V, Bendes E (2015) A new image quality metric for image fusion: the sum of the correlations of differences. AEU-Int J Electron Commun 69(12):1890–1896
25. Li H, Wu X (2019) Densefuse: a fusion approach to infrared and visible images. IEEE Trans Image Process 28(5):2614–2623
26. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680
27. Radford A, Metz L, Chintala S (2015) Unsupervised representation learning with deep convolutional generative adversarial networks. In: International conference on learning representation (ICLR), pp 1-16
28. Arjovsky M, Chintala S, Bottou L (2017) Wasserstein generative adversarial networks. In: Proceedings of the international conference on machine learning, pp 214–223
29. Mao X, Li Q, Xie H, Lau RYK, Wang Z, Paul Smolley S (2017) Least squares generative adversarial networks, pp 2813–2821
30. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: 2019 IEEE/CVF conference on computer vision and pattern recognition (CVPR)
31. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv:Learning
32. Liu L, Zhang H, Xu X, Zhang Z (2019) Collocating clothes with generative adversarial networks cosupervised by categories and attributes: a multidiscriminator framework. IEEE Trans Neural Netw Learn Syst 31(9):3540–3554
33. Ahmed KT, Irtaza A, Iqbal MA (2017) Fusion of local and global features for effective image extraction. Appl Intell 47(2):526–543
34. Hermessi H, Mourali O, Zagrouba E (2018) Convolutional neural network-based multimodal image fusion via similarity learning in the shearlet domain. Neural Comput Appl 30(7):2029–2045
35. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE conference on computer vision and pattern recognition, pp 770–778
36. Paul S, Sevcenco IS, Agathoklis P (2016) Multi-exposure and multi-focus image fusion in gradient domain. J Circuits Syst Comput 25(10):1–18
37. Ram Prabhakar K, Venkatesh Babu R (2016) Ghosting-free multi-exposure image fusion in gradient domain. In: IEEE International conference on acoustics, speech and signal processing, pp 1766–1770
38. Li S, Kang X, Jianwen H (2013) Image fusion with guided filtering. IEEE Trans Image Process 22(7):2864–2875
39. Liu Y, Wang Z (2015) Dense sift for ghost-free multi-exposure fusion. J Vis Commun Image Represent 31:208–224
40. Yang Y, Cao W, Shiqian W, Li Z (2018) Multi-scale fusion of two large-exposure-ratio images. IEEE Signal Process Lett 25(12):1885–1889

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.