Rethinking the Image Fusion: A Fast Unified Image Fusion Network based on Proportional Maintenance of Gradient and Intensity

Hao Zhang,[1] Han Xu,[1] Yang Xiao,[2] Xiaojie Guo,[3] Jiayi Ma[1,*]

[1]Electronic Information School, Wuhan University, Wuhan 430072, China
[2]School of Automation, Huazhong University of Science and Technology, Wuhan 430074, China
[3]College of Intelligence and Computing, Tianjin University, Tianjin 300350, China
{zhpersonalbox, xj.max.guo, jyma2010}@gmail.com, xu_han@whu.edu.cn, Yang_Xiao@hust.edu.cn

Abstract

In this paper, we propose a fast unified image fusion network based on proportional maintenance of gradient and intensity (PMGI), which can end-to-end realize a variety of image fusion tasks, including infrared and visible image fusion, multi-exposure image fusion, medical image fusion, multi-focus image fusion and pan-sharpening. We unify the image fusion problem into the texture and intensity proportional maintenance problem of the source images. On the one hand, the network is divided into gradient path and intensity path for information extraction. We perform feature reuse in the same path to avoid loss of information due to convolution. At the same time, we introduce the pathwise transfer block to exchange information between different paths, which can not only pre-fuse the gradient information and intensity information, but also enhance the information to be processed later. On the other hand, we define a uniform form of loss function based on these two kinds of information, which can adapt to different fusion tasks. Experiments on publicly available datasets demonstrate the superiority of our PMGI over the state-of-the-art in terms of both visual effect and quantitative metric in a variety of fusion tasks. In addition, our method is faster compared with the state-of-the-art.

Introduction

Image fusion is to extract the most meaningful information from images acquired by different sensors, and combine the information to generate a single image, which contains more abundant information and is more conducive to subsequent applications. Common image fusions include infrared and visible image fusion, multi-exposure image fusion, multi-focus image fusion, medical image fusion, and remote sensing image fusion (also called pan-sharpening). They are used in target detection, HDTV, medical diagnosis and other fields (Ma, Ma, and Li 2019; Ma et al. 2017; Xing et al. 2018).

Existing fusion methods can be divided into two categories. i) Traditional methods. These methods usually use the related mathematical transformation and manually designed fusion rules to get fused images. Paul *et al.* (Paul,

---

Sevcenco, and Agathoklis 2016) propose a general algorithm for multi-focus and multi-exposure image fusion, which is based on blending the gradients of the luminance components of source images using the maximum gradient magnitude at each pixel location and then obtaining the fused luminance using a Haar wavelet-based image reconstruction technique. The pyramid methods achieve medical image fusion by the diverse resolutions in the level and the iteration of the images (Patil and Mudengudi 2011). Rahmani *et al.* (Rahmani et al. 2010) come up with an adaptive IHS pan-sharpening Method, which produces images with higher spectral resolution while maintaining the high-quality spatial resolution of the original IHS. ii) Deep learning-based methods. Due to the strong fitting ability of neural networks, these methods can usually achieve better fused results by designing objective functions. In multi-focus image fusion, Liu *et al.* (Liu et al. 2017) used the convolutional neural network to jointly generate activity level measurement and fusion rules. Ma *et al.* (Ma et al. 2019a) proposed an unsupervised network to generate the decision map for fusion. In multi-exposure image fusion, Prabhakar *et al.* (Ram Prabhakar, Sai Srikar, and Venkatesh Babu 2017) proposed an unsupervised deep learning framework that utilizes a no-reference quality metric MEF-SSIM (Ma, Zeng, and Wang 2015) as the loss function and can produce satisfactory fused results. In medical image fusion, PA-PCNN proposed by Yin *et al.* (Yin et al. 2018) can realize the fusion of different types of medical images, which is a novel method in nonsubsampled shearlet transform (NSST) domain. Ma *et al.* (Ma et al. 2019b) proposed an end-to-end model called FusionGAN, which generates a fused image with dominant infrared intensities and additional visible gradients on the basis of GAN. Subsequently, they proposed to use a dual-discriminator to avoid loss of information in source images (Xu et al. 2019). In addition, they also introduced a detail loss and a target edge-enhancement loss based on FusionGAN to further enhance the texture details in the fused results (Ma et al. 2020).

Although existing methods can achieve good results in corresponding fusion tasks, there are still several aspects to be improved. First, existing methods usually need to manually design the activity level measurement and fusion rules.

Considering the diversity of source images, it will become more and more complex. Second, most methods are only applicable for specific fusion tasks and not generalized. It is significant to design a general method from the essence of image fusion. Third, because of computational complexity and large parameter quantities, existing fusion methods are often less competitive in terms of time.

To address these challenges, we propose a fast unified image fusion network based on proportional maintenance of gradient and intensity (PMGI), which can end-to-end implement various types of image fusion tasks with high efficiency. Firstly, PMGI is an end-to-end model with source images as the input and the fused image as the output, without any manual intervention in the middle. Secondly, we transform the fusion problem as the maintenance of gradient and intensity information. The intensity information gives the fused image similar histogram distribution with source images, while the gradient information gives finer texture details. Accordingly, we define the loss function of a uniform form for multiple image fusion tasks. To make the network adapt to different image fusion tasks, we can selective more effective and interesting information to be preserved in the fused result by adjusting the weight of each loss item. Finally, we divide the network into the gradient path and the intensity path to extract the corresponding information in source images respectively. In order to minimize the information loss caused by convolution, features of each layer in the same extraction path are reused. We also introduce the pathwise transfer block between the two paths. On the one hand, it can pre-fuse gradient and intensity information. On the other hand, it can enhance the information to be processed later. It is worth noting that due to the use of 1×1 convolution kernels and the control of the number of feature channels, the quantity of parameters in our network is limited within a certain range. As a result, our method can achieve fusion at a high speed.

The contributions of our work include the following three aspects:

- We propose a new end-to-end image fusion network, which can realize a variety of image fusion tasks uniformly. The proposed PMGI can fuse infrared and visible images, multi-exposure images, medical images, multi-focus images and remote sensing images well.

- We design a specific loss function, which is applicable to almost all image fusion tasks and can achieve expected results by adjusting the weight of each loss item.

- Our method can perform image fusion with higher efficiency in multiple fusion tasks. The code is available at: https://github.com/HaoZhang1018/PMGI_AAAI2020.

Method

Problem Formulation

The essence of image fusion is to combine the most important information in source images, and generate a single image with richer information and better visual effects. In different image fusion tasks, there are great differences in the properties of source images, and it is inappropriate to treat them in the same way. However, in most cases, the two types of source images maintain an underlying correlation, as they characterize the same scene and the source images contain the complementary information. So we try to solve different kinds of fusion tasks in a unified manner through reasonable design of network architecture and loss function.

As the most essential element of the image is the pixel, the intensity of pixels can represent the histogram distribution of the image, and the difference between the pixels constitutes the gradient, which can represent the texture details in the image. Thus, we characterize the whole image with information from these two aspects: the gradient and the pixel intensity. And it is reflected in both the network architecture and the loss function.

We divide the network into two information extraction paths: the gradient path and the intensity path. For the gradient path, it is responsible for extracting texture information, i.e., high-frequency features. Similarly, for the intensity path, it is responsible for the extraction of intensity information. Because both the gradient information and the intensity information are needed to be extracted and preserved from two types of source images at the same time, the input of each information extraction path is composed of different source images concatenated along the channel dimension for remaining potentially relevant. We set the concatenating ratio of these two source images as β. In addition, we also perform feature reuse and information exchange operations. First of all, loss of information is inevitable during the convolution process. Feature reuse can reduce information loss to a certain extent and increase the feature utilization. The exchange of different kinds of information between the information extraction paths can pre-fuse the gradient and the intensity information, and is also an enhancement of the information before the next extraction.

In addition to the general network structure described above, we also design a loss function with a uniform form based on the nature of the image. We transform the image fusion problem into the proportional maintenance problem of the gradient and the intensity information. Our loss function consists of two types of loss terms: the gradient loss and the intensity loss. They are both constructed for two source images. Separately, the intensity constraint can provide a rough pixel distribution, while the gradient constraint can enhance the texture details. And the joint constraint of them can achieve a satisfactory fused result with the reasonable intensity distribution and rich texture details. Because the fused image cannot preserve all the information of source images, we have to trade-off between the intensity distribution and texture details to preserve more important gradient and intensity information. Therefore, we can adjust the weight of each loss term to change the proportion of various types of information, so that it can be adapted to different image fusion tasks.

Network Architecture

The proposed PMGI is a very fast convolutional neural network. As shown in Fig. 1, we divide the network into the gradient and intensity path for corresponding information extraction. Gradient and intensity information is communi-
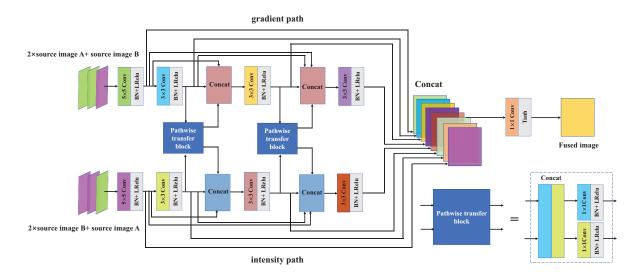
Figure 1: Network architecture of the proposed PMGI.

cated through the pathwise transfer block. It is worth noting that after many trials, the concatenating ratio β of the two source images in the inputs is determined to be $1 : 2$.

In both paths, we use four convolutional layers for feature extraction. Using the idea of DenseNet (Huang et al. 2017) for reference, we carry on dense connection in the same path to realize feature reuse. In addition, the pathwise transfer block is used to communicate information across these two paths, so the input of the third and fourth convolution layers depends not only on the output of all the previous convolution layers, but also on the output of the convolution layer in the other path. The first layer uses 5×5 convolution kernels and the latter three layers use 3×3 convolution kernels combined with batch normalization and leaky ReLU activation function. The structure of the pathwise transfer block is also shown in the lower right corner in Fig. 1. It uses a 1×1 convolution kernel combined with batch normalization and leaky ReLU activation function.

Then, we use the strategies of concatenating and convolution to fuse the features extracted from the two paths. We concatenate the two feature maps along the channel. It is worth noting that the idea of feature reuse is still used here. Eight feature maps involved in the concatenating come from a total of eight convolution layers of two paths. The kernel size of the last convolutional layer is 1×1, and the activation function is tanh. In all convolution layers, the padding is set to $SAME$ and stride is set to 1. As a result, none of these convolutional layers change the size of the feature map.

Loss Function

The loss function determines the type of information extracted and the proportional relationship between various types of information. The loss function of our network consists of two types of loss terms, the intensity loss and the gradient loss. The intensity loss constrains the fused image to maintain a similar intensity distribution as the source images, while the gradient loss forces the fused image to

contain rich texture details. It is important to note that we construct these two types of losses for each source image. Therefore, the loss function contains four terms, which are expressed as:

$$\mathcal{L}_{PMGI} = \lambda_{A_{\text{int}}}\mathcal{L}_{A_{\text{int}}} + \lambda_{A_{\text{grad}}}\mathcal{L}_{A_{\text{grad}}} + \lambda_{B_{\text{int}}}\mathcal{L}_{B_{\text{int}}} + \lambda_{B_{\text{grad}}}\mathcal{L}_{B_{\text{grad}}},$$
(1)

where A and B refer to the two source images, $\mathcal{L}_{(\cdot)_{\text{int}}}$ represents the intensity loss term for one source image and $\mathcal{L}_{(\cdot)_{\text{grad}}}$ means the corresponding gradient constraint. $\lambda_{(\cdot)}$ is the weight of each loss item.

The intensity loss is defined as:

$$\mathcal{L}_{A_{\text{int}}} = \frac{1}{HW}\|I_{\text{fused}} - I_{\text{A}}\|_2^2,$$
$$\mathcal{L}_{B_{\text{int}}} = \frac{1}{HW}\|I_{\text{fused}} - I_{\text{B}}\|_2^2,$$
(2)

where I_{fused} is the fused image generated by PMGI. I_{A} and I_{B} are two source images. H and W are the height and width of the image respectively.

Similarly, with ∇ denotes the gradient operator, definition of gradient loss is as follows:

$$\mathcal{L}_{A_{\text{grad}}} = \frac{1}{HW}\|\nabla I_{\text{fused}} - \nabla I_{\text{A}}\|_2^2,$$
$$\mathcal{L}_{B_{\text{grad}}} = \frac{1}{HW}\|\nabla I_{\text{fused}} - \nabla I_{\text{B}}\|_2^2.$$
(3)

Note that $\lambda_{(\cdot)}$ in Eq. (1) can be adjusted to change the proportion of various types of information in the fused image to adapt to different tasks. The parameter setting rules corresponding to specific tasks are summarized below.

For infrared and visible image fusion, we expect that the gradient information of visible images and the intensity information of infrared images are mainly preserved in the fused results, while the intensity information of visible images and the gradient information of infrared images are of secondary importance. Then the parameter $\lambda_{(\cdot)}$ should meet

the following setting rules:

$$\lambda_{ir_{\text{int}}} > \lambda_{vis_{\text{int}}}, \ \lambda_{ir_{\text{grad}}} < \lambda_{vis_{\text{grad}}}. \quad (4)$$

For multi-exposure image fusion, both overexposed and underexposed images contain equal texture details, but their intensities are too strong or too weak. So we set the same weights to balance them to get the appropriate intensity and rich texture details, which can be formalized as:

$$\lambda_{over_{\text{int}}} = \lambda_{under_{\text{int}}}, \ \lambda_{over_{\text{grad}}} = \lambda_{under_{\text{grad}}}. \quad (5)$$

For multi-focus image fusion, the two types of information (gradient and intensity) of the two source images are equally important. This is because we want to maintain the intensity and texture information of the two source images at the same time, and the defocused (blurred) regions can be supplemented by the focused (clear) regions in the other source image. Therefore, it is also necessary to set the corresponding parameters to remain the same:

$$\lambda_{focus1_{\text{int}}} = \lambda_{focus2_{\text{int}}}, \ \lambda_{focus1_{\text{grad}}} = \lambda_{focus2_{\text{grad}}}. \quad (6)$$

Similarly, for medical image fusion, the structural medical image reflects the texture information of the organ, and the functional medical image indicates the functional information, such as the metabolic intensity. We use MRI and PET images as the example of the structural image and the functional image respectively. We obtain the main texture information from the MRI image and the main intensity information from the PET image. However, considering that the pixel intensity of the I component of the PET image is much larger than the MRI, if the pixel intensity of the PET image is mainly constrained, the excessive intensity of the fused image masks the texture. So in order to balance the texture and intensity, we make the pixel intensity of PET and MRI equally constrained. So $\lambda_{(\cdot)}$ should be set as:

$$\lambda_{PET_{\text{int}}} = \lambda_{MRI_{\text{int}}}, \ \lambda_{PET_{\text{grad}}} < \lambda_{MRI_{\text{grad}}}. \quad (7)$$

Finally, for pan-sharpening, the panchromatic image has a high spatial resolution (rich texture details), and the multi-spectral image contains rich color information. The purpose is to improve clarity while keeping the spectrum undistorted. Therefore, we only constrain the texture information of the panchromatic image without constraining the intensity to avoid spectral distortion, which can be formalized as:

$$\lambda_{PAN_{\text{int}}} = 0, \ \lambda_{PAN_{\text{grad}}} > \lambda_{MS_{\text{grad}}}. \quad (8)$$

Experimental Results

Experimental Settings

Data We perform PMGI on five fusion tasks: 1) visible and infrared image fusion; 2) multi-exposure image fusion; 3) medical image fusion; 4) multi-focus image fusion; 5) pan-sharpening. The training and test sets are from publicly available datasets: *TNO*[1] dataset for task1; *MEF* dataset[2] and

---

[1]http://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029.

[2]https://drive.google.com/drive/folders/0BzXT0LnoyRqlVjhtOEhiUzU5a2M

the dataset provided by (Cai, Gu, and Zhang 2018) for task2; MRI and PET images from Harvard medical school website[3] for task3; the dataset provided by (Nejati, Samavi, and Shirani 2015) and sametaymaz[4] for task4; and the panchromatic and multi-spectral images generated by the Quickbird satellite for task5.

For these five fusion tasks, we enter the entire image as test data into the network. The number of image pairs used for testing is 17, 19, 20, 18 and 50, respectively. In order to obtain more training data, we adopt the expansion strategy of tailoring and decomposition. In particular, we crop each image into 120×120 image patches in all tasks except multi-focus image fusion. For the multi-focus image fusion task, we expect that there is only one attribute (focused or defocused) in each image patch. So we crop each image into image patches with smaller size 60×60. The number of patch pairs for training is 20036, 33961, 4698, 139001, and 210000, respectively.

Training Details Among these image fusion tasks, source images and fused images in infrared and visible image fusion and multi-focus image fusion tasks are grayscale images, while those in the other fusion tasks are color images. In medical image fusion, the multi-spectral PET image with RGB channels are firstly transformed to IHS channels. We fuse the I component and the MRI image. Then the fused result is concatenated with H and S components and transformed to get the final RGB fused result. For multi-exposure image fusion, we transform the images from RGB to YCbCr color space. Because the Y channel (luminance channel) can represent structural details and the brightness variation, we just devote to fusing the Y channel values. And for Cb and Cr channels (chrominance channels), we fuse them in a traditional way. Then, the fused components of these channels are transferred to RGB color space to obtain the final result. In the pan-sharpening task, the multi-spectral image contains four channels and the panchromatic image is one-channel. The resolution ratio of them is $1 : 4$. First, we use the bicubic method to upsample the multi-spectral image to the same size as the panchromatic image. Then, each channel of the multi-spectral image is fused with the panchromatic image separately to obtain four one-channel fused results. Finally, they are concatenated to obtain the final fused image.

It is worth noting that all deep learning based methods run on the same GPU RTX 2080Ti, while other methods run on the same CPU i7-8750H.

Results on Infrared and Visible Image Fusion

In infrared and visible image fusion, we compare our PMGI with five existing methods: LPP (Toet 1989), GTF (Ma et al. 2016), DDLatLRR (Li and Wu 2018a), LatLRR (Li and Wu 2018b) and FusionGAN (Ma et al. 2019b). Qualitative and quantitative experimental results are shown below.

---

[3]http://www.med.harvard.edu/AANLIB/home.html

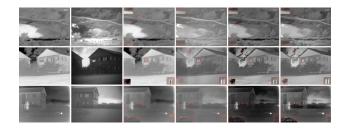[4]https://github.com/sametaymaz/Multi-focus-Image-Fusion-Dataset

Figure 2: Qualitative Results of infrared and visible image fusion. From left to right: infrared image, visible image, and fused results of GTF, LatLRR, FusionGAN and our PMGI.
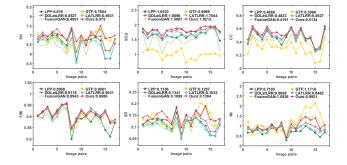

Figure 3: Quantitative results of infrared and visible image fusion on six metrics.

Qualitative Comparisons Comparison results on three typical image pairs qualitatively demonstrate the superiority of our method, as shown in Fig. 2. According to the characteristics of fused results, these methods can be divided into two categories. In the first category, fused results are biased towards visible images, which contain rich textures but lose thermal radiation information, such as LatLRR. Results of the other type of methods are closer to infrared images in the respect of contrast, but the texture details are not rich enough, such as GTF and FusionGAN. By comparison, our method is more like a combination of these two categories. First, our results can maintain thermal radiation information by effectively highlighting the target, such as the lake in the first row and the human in the third row. At the same time, our results contain rich texture details, such as the grass in the first row and the car in the third row, which are more in line with human observation characteristics.

Quantitative Comparisons In order to validate more comprehensively, we conduct quantitative comparisons on seventeen image pairs. We used six objective metrics to evaluate the results: entropy (EN), sum of the correlations of differences (SCD) (Aslantas and Bendes 2015), correlation coefficient (CC), Feature mutual information (FMI), standard deviation (SD) and mutual information (MI).

The quantitative results are shown in Fig. 3. Our method achieves the largest average values on four metrics including EN, SCD, CC and SD. From these results, we can conclude that our results contain the largest amount of information, strong correlation with source images and significant contrast. In addition, our results contains comparable amount of

feature information transferred from source images, as our method only ranks second behind DDLatLRR on FMI. Last but not least, although GTF and FusionGAN have larger values on MI, our method achieves a better balance between the information of two source images.

The average run time of LPP, GTF, DDLatLRR, LatLRR, FusionGAN and our PMGI on the test data is about 0.1156, 4.5426, 49.4816, 133.2983, 0.2027 and 0.0831 (unit:second). Obviously, our PMGI is the fastest.

Results on Multi-exposure Image Fusion

To verify the effectiveness in multi-exposure image fusion, we compare it with three existing methods: DSIFT (Hayat and Imran 2019), EF (Mertens, Kautz, and Van Reeth 2007) and AWPIGG (Lee, Park, and Cho 2018).

Qualitative Experiment We firstly perform qualitative experiments on three typical image sequences. The fused results are shown in Fig. 4. We analyze them from two aspects: the whole image and local details. As for the whole image, compared with other methods, our results are more suitable for human eye perception with more vivid colors and more appropriate exposure. And in our results, there are no strange shadows like those in the results of the comparative methods, such as dark shadows in the clouds. As for local details, our results have clearer and more realistic textures such as grasses and rocks highlighted in images. In general, our method has great advantages over the comparative methods.

Quantitative Experiment We still choose EN, SCD, CC, FMI, SD and MI described above to evaluate fused results. The comparison results on nineteen image pairs are shown in Fig. 5. As for EN, SCD, CC, SD and MI, our results can achieve the highest average values. From these results, we can see that our fused results have the most information, the strongest correlation with the source images, the most information from the source images, and the best contrast. In addition, although our method follows behind DSIFT, EF and AWPIGG on FMI, we can also see that our results get comparable feature information from source images, because the values of all the methods are very close.

The average run time of DSIFT, EF, AWPIGG and PMGI is 1.3307, 0.0053, 0.4887 and 0.0908 (unit: second). PMGI is second only to EF, but still stay within 100 milliseconds.

Results on Medical Image Fusion

We employ three existing medical image fusion methods to compare with our PMGI: DCHWT (Kumar 2013), ASR (Liu and Wang 2014) and PCA (Naidu and Raol 2008).

Qualitative Experiment Three typical and intuitive results on three different transaxial sections of the brain-hemispheric are exhibited in Fig. 6. Among them, the fused results of PCA can retain the functional information (presented as color) of PET images, but lose much texture structure information (presented as texture details) of MRI images. On the contrary, the fused results of DCHWT have significant structural information, but reduce the color intensity in PET images. Results of ASR, DCHWT and our

Figure 4: Qualitative results of multi-exposure image fusion. From left to right: underexposed image, overexposed image, and the results of DSIFT, EF, AWPIGG and our PMGI.
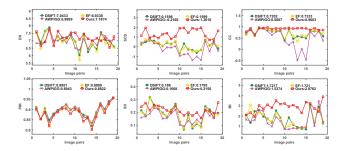

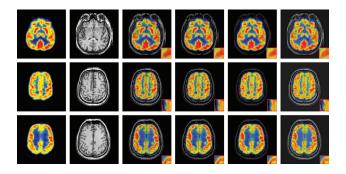Figure 5: Quantitative results of multi-exposure image fusion.

Figure 6: Qualitative results of medical image fusion. From left to right: PET image, MRI image, and fused results of DCHWT, ASR, PCA and our PMGI.
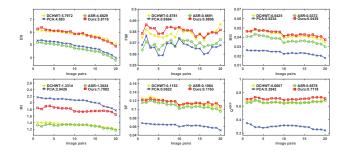

Figure 7: Quantitative results of medical image fusion.

method not only have vivid color intensity, but also contain rich texture structure. But by contrast, the texture details of our results are more clear, as shown in the highlighted parts.

Quantitative Experiment Because of the particularity of medical images, we need to strictly control the preservation of edge information, which represents the shape and area of the organ. So we add the metric $Q^{AB/F}$ to measure the edge information that is transferred from source images to the fused image. In addition, spatial frequency (SF) and mean gradient (MG) are added to evaluate the gradient information in order to measure the degree of structural information (texture details) retention of MRI images. Therefore, the six objective metrics we select to evaluate medical fused images are EN, FMI, MG, MI, SF and $Q^{AB/F}$.

The quantitative results on twenty image pairs are shown in Fig. 7. As can be seen from the statistical results, our PMGI can generate the largest average values on the five metrics: EN, FMI, MG, SF and $Q^{AB/F}$. It demonstrates that our results have the most information, the most abundant texture details, and the most feature and edge information from source images. As for MI, our method ranks second only to PCA. It is due to the reason that our results are more balanced between structural information and functional information compared with PCA.

The average run time of DCHWT, ASR, PCA and our PMGI is about 0.8636, 23.6055, 0.0046, and 0.0150 (unit: second). Our method is second to PCA, but is much faster than the other two methods.

Results on Multi-focus Image Fusion

Five multi-focus image fusion methods are selected to compare with our PMGI: CNN (Liu et al. 2017), Dct-Var (Haghighat, Aghagolzadeh, and Seyedarabi 2011), DWTDE (Liu and Wang 2013), GD (Paul, Sevcenco, and Agathoklis 2016) and QBMF (Bai et al. 2015).

Qualitative Experiment To give intuitive results, we select three typical image pairs for qualitative evaluation, as shown in Fig. 8. It can be seen from the results that there are some intensity distortions and unnatural black shadows in the fused results of GD. The results of DWTDE have a good visual effect, which can restore the details of far and near-focused regions clearly. Although there is a slight of loss of details of the far-focused regions in the results of QBMF and our method, they can also achieve relatively satisfactory results. In addition, it should be noted that our method is a general end-to-end fusion method. Considering the universality, we do not perform multi-focus image fusion by detecting the focused areas but extracting and fusing the texture and intensity information of source images. Nevertheless, the qualitative results still prove the effectiveness of our method.

Quantitative Experiment The essence of multi-focus image fusion is the fusion of focused (clear) and defocused (blurred) parts. Focused and defocused areas have similar intensity distributions but large differences in sharp appearance, so the key is the preservation of edge and texture information. Therefore, we use the structural similarity index measure (SSIM) and $Q^{AB/F}$ to evaluate the maintained
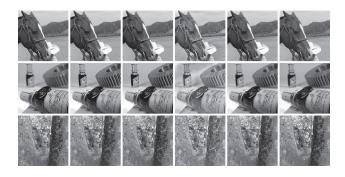
Figure 8: Qualitative results of multi-focus image fusion. The first two columns are source images with different focus. The last four columns are fused results of DWTDE, GD, QBMF and our PMGI.
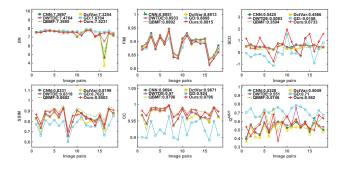

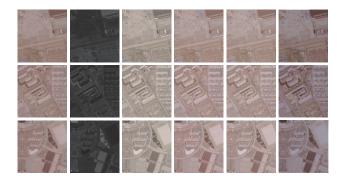Figure 9: Quantitative results of multi-focus image fusion.

Figure 10: Qualitative results of Pan-Sharpening. From left to right: multi-spectral image, panchromatic image, and fused results of PCA, FISH, Brovey and our PMGI.

structure and the transferred edge information in the fused results. To maintain consistency with other tasks, EN, FMI, SCD and CC are also used. The experimental results on eighteen image pairs are shown in Fig. 9.

Our PMGI can obtain the largest average values on SCD and CC, and the second largest average values on EN and SSIM. For QABF, the average value of our method is ranked third behind GD and QBMF and ranked fifth on the FMI. It can be seen that our results have the strongest correlation with source images, a lot of information, and strong structural similarity with source images. However, our results are not good enough in maintaining of the edge information and the acquisition of the feature information, which also corresponds to the fact that the details of our results are not rich enough in far-focused regions.

The average run time of CNN, DctVar, DWTDE, GD, QBMF and our PMGI on the testing data is 105.4371, 0.7388, 10.3615, 0.7905, 0.4313 and 0.0691(unit: second). Obviously, our PMGI has a significant advantage in running time, which is about one order of magnitude faster compared with other five methods.

Results on Pan-sharpening

To demonstrate the versatility of our approach, we use PMGI to perform pan-sharpening in remote sensing image fusion tasks. We compare our method with classical methods PCA (Chavez et al. 1991), FISH (Tu et al. 2004) and Brovey (Gillespie, Kahle, and Walker 1986). Some typical qualitative results are shown in Fig. 10. It is obvious that PMGI can accomplish the pan-sharpening task well, and the fused results are color images with high resolution with a good visual effect. First, because we only constrain the texture of the panchromatic image without constraining its intensity, the fused results only maintain the color information of the multi-spectral image without introducing large spectral distortion like PCA. Second, the fused results have the richest and high-quality texture details, which are almost as good as those in the panchromatic image.

Conclusion

In this paper, we propose a fast unified image fusion network based on proportional maintenance of gradient and intensity, termed as PMGI. It is an end-to-end model which can perform a variety of different image fusion tasks well, including infrared and visible image fusion, multi-exposure image fusion, medical image fusion, multi-focus image fusion and pan-sharpening. We divide the network into the gradient path and the intensity path, which carry out feature reuse in the same path and information exchange in different paths. At the same time, we unify multiple image fusion tasks into the maintaining problem of intensity and gradient information, and define a specific loss function, which is composed of the gradient loss and the intensity loss. The PMGI can adapt to different image fusion tasks by adjusting the weight of each loss item. Qualitative and quantitative experiments show that our PMGI has advantages in various image fusion tasks compared to existing methods, and it is also faster than the competitors.

Acknowledgments

References

Aslantas, V., and Bendes, E. 2015. A new image quality metric for image fusion: the sum of the correlations of differences.

Aeu-International Journal of Electronics and Communications 69(12):1890–1896.

Bai, X.; Zhang, Y.; Zhou, F.; and Xue, B. 2015. Quadtree-based multi-focus image fusion using a weighted focus-measure. *Information Fusion* 22:105–118.

Cai, J.; Gu, S.; and Zhang, L. 2018. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing* 27(4):2049–2062.

Chavez, P.; Sides, S. C.; Anderson, J. A.; et al. 1991. Comparison of three different methods to merge multiresolution and multi-spectral data- landsat tm and spot panchromatic. *Photogrammetric Engineering and Remote Sensing* 57(3):295–303.

Gillespie, A. R.; Kahle, A. B.; and Walker, R. E. 1986. Color enhancement of highly correlated images. i. decorrelation and hsi contrast stretches. *Remote Sensing of Environment* 20(3):209–235.

Haghighat, M. B. A.; Aghagolzadeh, A.; and Seyedarabi, H. 2011. Multi-focus image fusion for visual sensor networks in dct domain. *Computers & Electrical Engineering* 37(5):789–797.

Hayat, N., and Imran, M. 2019. Ghost-free multi exposure image fusion technique using dense sift descriptor and guided filter. *Journal of Visual Communication and Image Representation* 62:295–308.

Huang, G.; Liu, Z.; Van Der Maaten, L.; and Weinberger, K. Q. 2017. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 4700–4708.

Kumar, B. S. 2013. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform. *Signal, Image and Video Processing* 7(6):1125–1143.

Lee, S.-h.; Park, J. S.; and Cho, N. I. 2018. A multi-exposure image fusion based on the adaptive weights reflecting the relative pixel intensity and global gradient. In *Proceedings of the IEEE International Conference on Image Processing*, 1737–1741.

Li, H., and Wu, X.-J. 2018a. Infrared and visible image fusion using a novel deep decomposition method. *arXiv preprint arXiv:1811.02291*.

Li, H., and Wu, X.-J. 2018b. Infrared and visible image fusion using latent low-rank representation. *arXiv preprint arXiv:1804.08992*.

Liu, Y., and Wang, Z. 2013. Multi-focus image fusion based on wavelet transform and adaptive block. *Journal of Image and Graphics* 18(11):1435–1444.

Liu, Y., and Wang, Z. 2014. Simultaneous image fusion and denoising with adaptive sparse representation. *IET Image Processing* 9(5):347–357.

Liu, Y.; Chen, X.; Peng, H.; and Wang, Z. 2017. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion* 36:191–207.

Ma, J.; Chen, C.; Li, C.; and Huang, J. 2016. Infrared and visible image fusion via gradient transfer and total variation minimization. *Information Fusion* 31:100–109.

Ma, K.; Li, H.; Yong, H.; Wang, Z.; Meng, D.; and Zhang, L. 2017. Robust multi-exposure image fusion: a structural patch decomposition approach. *IEEE Transactions on Image Processing* 26(5):2519–2532.

Ma, B.; Ban, X.; Huang, H.; and Zhu, Y. 2019a. Sesf-fuse: An unsupervised deep model for multi-focus image fusion. *arXiv preprint arXiv:1908.01703*.

Ma, J.; Yu, W.; Liang, P.; Li, C.; and Jiang, J. 2019b. Fusiongan: A generative adversarial network for infrared and visible image fusion. *Information Fusion* 48:11–26.

Ma, J.; Liang, P.; Yu, W.; Chen, C.; Guo, X.; Wu, J.; and Jiang, J. 2020. Information fusion of passive sensors for detection of moving targets in dynamic environments. *Information Fusion* 54:85–98.

Ma, J.; Ma, Y.; and Li, C. 2019. Infrared and visible image fusion methods and applications: A survey. *Information Fusion* 45:153–178.

Ma, K.; Zeng, K.; and Wang, Z. 2015. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing* 24(11):3345–3356.

Mertens, T.; Kautz, J.; and Van Reeth, F. 2007. Exposure fusion. In *Proceedings of the Pacific Conference on Computer Graphics and Applications*, 382–390.

Naidu, V., and Raol, J. R. 2008. Pixel-level image fusion using wavelets and principal component analysis. *Defence Science Journal* 58(3):338–352.

Nejati, M.; Samavi, S.; and Shirani, S. 2015. Multi-focus image fusion using dictionary-based sparse representation. *Information Fusion* 25:72–84.

Patil, U., and Mudengudi, U. 2011. Image fusion using hierarchical pca. In *Proceedings of the International Conference on Image Information Processing*, 1–6.

Paul, S.; Sevcenco, I. S.; and Agathoklis, P. 2016. Multi-exposure and multi-focus image fusion in gradient domain. *Journal of Circuits, Systems and Computers* 25(10):1650123.

Rahmani, S.; Strait, M.; Merkurjev, D.; Moeller, M.; and Wittman, T. 2010. An adaptive ihs pan-sharpening method. *IEEE Geosci. Remote Sens. Lett.* 7(4):746–750.

Ram Prabhakar, K.; Sai Srikar, V.; and Venkatesh Babu, R. 2017. Deepfuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs. In *Proceedings of the IEEE International Conference on Computer Vision*, 4724–4732.

Toet, A. 1989. Image fusion by a ratio of low-pass pyramid. *Pattern Recognition Letters* 9(4):245–253.

Tu, T.-M.; Huang, P. S.; Hung, C.-L.; and Chang, C.-P. 2004. A fast intensity-hue-saturation fusion technique with spectral adjustment for ikonos imagery. *IEEE Geoscience and Remote Sensing Letters* 1(4):309–312.

Xing, L.; Cai, L.; Zeng, H.; Chen, J.; Zhu, J.; and Hou, J. 2018. A multi-scale contrast-based image quality assessment model for multi-exposure image fusion. *Signal Processing* 145:233–240.

Xu, H.; Liang, P.; Yu, W.; Jiang, J.; and Ma, J. 2019. Learning a generative model for fusing infrared and visible images via conditional generative adversarial network with dual discriminators. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3954–3960.

Yin, M.; Liu, X.; Liu, Y.; and Chen, X. 2018. Medical image fusion with parameter-adaptive pulse coupled neural network in nonsubsampled shearlet transform domain. *IEEE Transactions on Instrumentation and Measurement* 68(1):49–64.