Fractional Optimization Model for Infrared and Visible Image Fusion

Kang Zhang
zhang_kang@njust.edu.cn
Shi-Wei Wu
wushiwei98@gmail.com
Zhi-Liang Wu
wu_zhiliang@njust.edeu.cn
Xia Yuan(✉)
yuanxia@njust.edeu.cn
Chun-Xia Zhao
zhaochunxia@126.com

School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract

Infrared and visible image fusion is a fundamental task for image processing to enhance the image quality. To highlight target and retain effective details, different from previous methods using integer gradients, we use the fractional gradient to well represent image features and propose a novel fractional optimization model to fuse infrared and visible images. Specially, a fractional optimization function is designed with global contrast fidelity and fractional gradient constraint to obtain the pre-fused image. Then, the base layer of the pre-fused image obtained by multi-level decomposition latent low-rank representation is taken as the fused base layer, while for the fusion of detail layers, a fractional gradient energy function is designed to evaluate the importance of detail information to generate the fused detail layers. Compared with 10 state-of-the-art image fusion methods qualitatively and quantitatively on two public datasets (TNO and RoadScene), our method generally shows superior performance.

1 Introduction

Generally speaking, infrared images can reflect the thermal radiation of an object and highlight the target, but lack texture information. In contrast, visible images capture reflected light information which contains rich texture and structure information, but are susceptible to illumination and weather conditions. Therefore, using one kind of image alone cannot provide sufficient information for practical applications such as RGB-T target tracking. To solve this problem, infrared and visible image fusion is proposed, which can combine these information to produce a robust fused image containing more detail textures. Figure 1 shows an example of infrared and visible image fusion.

In recent years, researchers have proposed many methods to solve the problem of image fusion, which can be roughly divided into two categories. The first is the traditional image

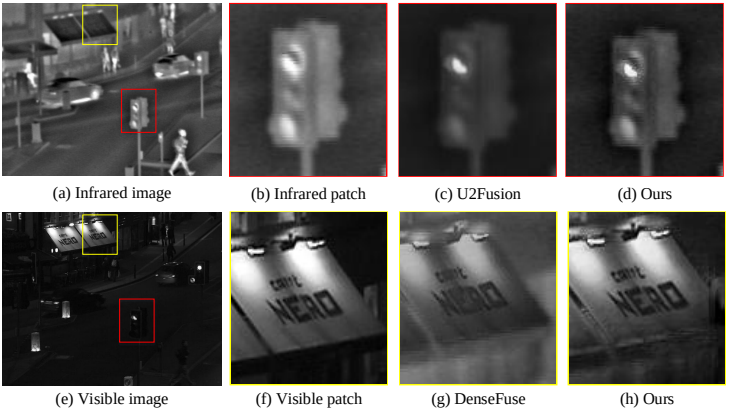

Figure 1: Examples of infrared and visible image fusion. While U2Fusion [23] cannot well preserve the details of the target in infrared images due to the introduction of disturbing visible information, DenseFuse [6] extracts unnecessary infrared information, resulting in reduction of visible texture details. In contrast, the image generated by the proposed method has a good balance of infrared and visible information with richer complementary details.

fusion methods [5, 8, 12, 13], among which the common methods are based on multi-scale transformation. However, the conventional weighted fusion rule for the base layer often makes the fused image ignore the global contrast, which makes it difficult to highlight the target in complex scenes. And these methods usually do not well preserve useful detail information, which is not conductive to subsequent tasks. The second is the fusion methods based on deep learning [6, 7, 9]. These methods usually train the network by constructing the loss function so that the fused image has the required distribution characteristics. Because of the strong nonlinear fitting ability of neural network, these methods achieve good fusion results. However, these methods still have disadvantages. In the case of insufficient training data, network training is difficult, especially in infrared and visible image fusion task.

In this paper, we propose a new infrared and visible image fusion method based on fractional optimization. To alleviate the problem that the general base layer fusion rules easy to ignore the global contrast, we regard the fusion of the base layer as an optimization problem, which uses fractional gradient to better represent image features. And the pre-fused image is generated under the global contrast fidelity and fractional gradient constraint. Then, the fused base layer is obtained by using MDLatLRR [8] to decompose the pre-fused image. To sufficiently extract the useful detail information, a fractional gradient energy function is designed to distribute the weight of detail information and generate the fused detail layers. Finally, the fused image is reconstructed based on the fused base layer and detail layers.

The contributions of this paper are summarized as follows:

- We propose a novel infrared and visible image fusion method based on fractional optimization model. Specifically, a new fractional optimization-based pre-fusion module is designed to fuse the base layer, it introduces the global contrast of infrared images and the fractional gradient information of both infrared and visible images. Moreover, a fractional gradient energy function is innovatively proposed to evaluate the detail layer information, it effectively preserves the rich and effective detail information of infrared and visible images.

- To make a fair and comprehensive comparison with other fusion methods, we conduct comparative experiments on two public datasets. Experimental results show that the proposed method performs better in subjective and objective evaluation.

2 Related Work

2.1 Multi-level decomposition based on LatLRR (MDLatLRR)

Recently, Li *et al.* [8] proposed a multi-level image decomposition method based on latent low-rank representation (LatLRR), called MDLatLRR. In detail, given an input image $I \in \mathbb{R}^{H\times W}$, the base layer $I_b^i \in \mathbb{R}^{H\times W}$ and detail layers $I_d^i \in \mathbb{R}^{H\times W}$ of the input image are obtained by LatLRR-based decomposition at level i. The base layer and detail layers are obtained below.

$$\begin{aligned} F_{dk}^i &= Q \times P(I_{bk}^{i-1}), I_{bk}^i = I_{bk}^{i-1} - I_{dk}^i, \\ I_{dk}^i &= R(F_{dk}^i), I_{bk}^0 = I_k, i = [1,2,\cdots,L], \end{aligned} \tag{1}$$

where k denotes the number of images. i is the decomposition level. '$\times$' denotes dot product. According to MDLatLRR [8], Q denotes the projection matrix learned by LatLRR [10] and the size is 256×256. $P(\cdot)$ represents a two-stage operator composed of sliding window and recombination technology. F_{dk}^i represents the detail features and the size is $256 \times [(H - S)/s + 1] \times [(W - S)/s + 1]$. S and s represent the size and stride of the sliding window, respectively. Here, $S = 16$ and $s = 1$. $R(\cdot)$ represents the reconstruction operator.

2.2 Infrared and visible image fusion

Infrared and visible image fusion is one of the mainstream of image fusion. Over the past few decades, researchers have proposed many methods [13, 22, 24] . Wang *et al.* [22] used a non-negative sparse representation method to extract features from source images. Ma *et al.* [13] proposed a gradient transfer fusion algorithm, which can realize the fusion of images without pre-registration and improve the applicability of the algorithm. Zhang *et al.* [24] used intuitionistic fuzzy sets to extract high-frequency detail features from infrared and visible images. Although these methods have achieved good results, the selection of feature extraction methods and the difficulty of designing fusion rules limit the development of traditional methods.

Due to the powerful feature extraction capability of deep learning, more and more deep learning-based infrared and visible image fusion methods have been proposed in recent years [7, 9, 19, 20]. Li *et al.* [7] established two kinds of attention models to characterize the importance of deep features. Li *et al.* [9] proposed an end-to-end residual fusion network and designed learnable fusion rules. Tang *et al.* [20] proposed a real-time image fusion network based on semantic perception. However, since ground truth is usually absent for this fusion task, utilizing supervised learning is a major hurdle for infrared and visible light image fusion.

Inspired by literature [8], we use MDLatLRR to obtain the multi-layer features of the source images, then combine the advantages of fractional model to design the fusion rules. Thus, we propose a new optimization model for infrared and visible image fusion to improve the fusion performance.

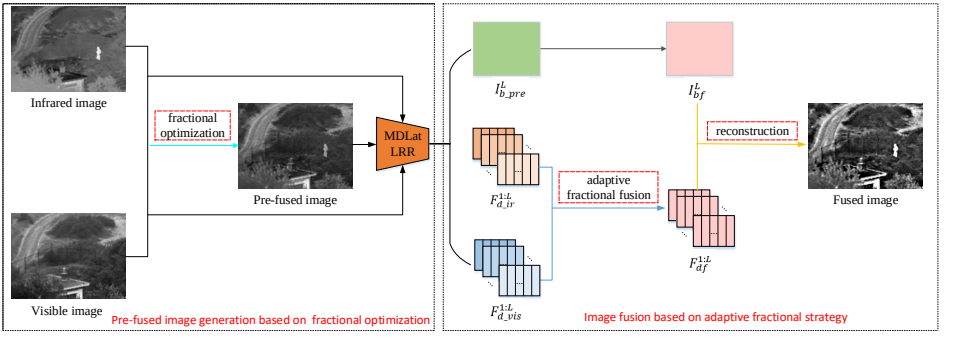

Figure 2: The overview of the proposed method.

3 Proposed Method

To mitigate the problem that common methods tend to ignore global contrast and insufficient extraction of relevant details, we propose a new image fusion method based on fractional optimization. The framework is shown in Figure 2. Specially, a optimization function is designed with global contrast fidelity and fractional gradient constraint, and a Sylvester-based optimization method is used to obtain the pre-fused image. Then, the base layer of the pre-fused image obtained by MDLatLRR is taken as the fused base layer. For the fusion of detail layers, the fractional gradient energy function is introduced to evaluate the importance of detail information to generate the fused detail layers. Finally, the fused image is reconstructed by the fused base layer and detail layers.

3.1 The fusion of base layer

Fractional derivative, as a generalization of integer derivative, has the characteristics that integer derivative does not have. It can represent the high-frequency and low-frequency information in the image better. The base layer of the image contains most of the global features and contrast information of the image. To make the fused image have higher contrast and more global features, this paper uses fractional optimization to generate the pre-fused image, which retains the contrast of the infrared image and extracts the fractional gradient features of the infrared and visible images. The optimization model of pre-fused image f_{pre} generated according to infrared image f_{ir} and visible image f_{vis} is as follows:

$$arg\min_{f} \frac{1}{2}\|f - f_{ir}\|_F^2 + \frac{\mu}{2}\|D^{v_f} f - D^{v_{f_{ir}}} f_{ir} - D^{v_{f_{vis}}} f_{vis}\|_F^2, \tag{2}$$

where $\|\cdot\|_F$ represents the Frobenius norm. $\mu = 100$. $D^v s$ denotes the v-order discrete fractional gradient of $s \in \mathbb{R}^{n\times n}$ [16].

$$D^v s = \begin{pmatrix} Ms \\ sM \end{pmatrix} \in \mathbb{R}^{2n\times n}, M = \begin{pmatrix} 2c_1^v & c & c_3^v & \cdots & c_n^v \\ c & 2c_1^v & \ddots & \ddots & \vdots \\ c_3^v & \ddots & \ddots & \ddots & c_3^v \\ \vdots & \ddots & \ddots & 2c_1^v & c \\ c_n^v & \cdots & c_3^v & c & 2c_1^v \end{pmatrix} \in \mathbb{R}^{n\times n}, \tag{3}$$

and

$$c_k^v = (-1)^k \binom{v}{k}, c_0^v = 1, c = c_0^v + c_2^v, (D^v)^T \begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = Ms_1 + s_2 M. \tag{4}$$

The first term of Eq.(2) ensures that the fused image f and infrared image f_{ir} has similar contrast, and the second term ensures that the fused image f retains the fractional gradient features in infrared image f_{ir} and visible image f_{vis}. Here, we set $v_f = 0.5v_{f_{ir}} + 0.5v_{f_{vis}}$. Furthermore, the impact of two parameters $v_{f_{ir}}$ and $v_{f_{vis}}$ on model performance will be discussed in Section 4.4.

For the optimization solution of Eq.(2), we can obtain from the first-order condition:

$$f - f_{ir} + \mu (D^{v_f})^T (D^{v_f} f - D^{v_{f_{ir}}} f_{ir} - D^{v_{f_{vis}}} f_{vis}) = 0, \tag{5}$$

then based on Eq.(3) and Eq.(4), we can obtain that:

$$f + \mu (M_{v_f}^2 f + f M_{v_f}^2) = f_{ir} + \mu (D^{v_f})^T (D^{v_{f_{ir}}} f_{ir}) + \mu (D^{v_f})^T (D^{v_{f_{vis}}} f_{vis}), \tag{6}$$

and

$$\begin{aligned} (E + \mu M_{v_f}^2) f + f(\mu M_{v_f}^2) = (E + \mu M_{v_f} M_{v_{f_{ir}}}) f_{ir} + f_{ir}(\mu M_{v_{f_{ir}}} M_{v_f}) \\ + f_{vis}(\mu M_{v_{f_{vis}}} M_{v_f}) + (\mu M_{v_f} M_{v_{f_{vis}}}) f_{vis}, \end{aligned} \tag{7}$$

where E denotes the identity matrix. We can find that Eq.(7) is equivalent to the form of $AX + XB = C$, namely Sylvester equation. It can be solved by the Bartels-Stewart algorithm [3]. More details of the optimization model are provided in the Supplementary Materials.

The pre-fused image f_{pre} can be obtained by solving Eq.(2). In addition, to obtain multi-level features, we perform MDLatLRR decomposition on the source images and the pre-fused image to obtain the base layer and the detail layers, respectively. The base layer of the pre-fused image is directly used as the fused base layer, because the pre-fused image retains a large amount of energy information of the source images and controls global contrast. Therefore, the fused base layer can be obtained by:

$$I_{bf} = MDLatLRR(f_{pre}), \tag{8}$$

where MDLatLRR is denoted by Eq.(1).

3.2 The fusion of detail layers

Compared with the base layer, detail layers contain a large amount of texture details of the image. Here, we use the infrared image and visible image to perform the fusion operation of the detail layers. Because fractional differential operators can enhance high-frequency features while retaining low-frequency information [17], we put forward a new fractional gradient energy function to extract detail features, which is defined as follows:

$$FGE_{dk}^i = \sum_{x=1}^{M-1} \sum_{y=1}^{N-1} \{[G_1 * F_{dk}^i(m,n)]^2 + [G_2 * F_{dk}^i(m,n)]^2\}, \tag{9}$$

where G_1 and G_2 are two masks in negative x- and y- directions of fractional differential operator *YiFeiPU-1* [17]. '$*$' denotes the convolution operation. Here, we set $k = [1,2]$. It is worth noting that the determination of the order is very important for fractional differential

operator. Therefore, we propose an adaptive fractional order calculation method based on modified spatial frequency and average gradient. The calculation formula is described as:

$$\begin{aligned} v &= (a-b)\cdot\frac{f(U)-min(U)}{max(U)-min(U)}+b,\\ f(U) &= tanh(U)=\frac{e^{U}-e^{-U}}{e^{U}+e^{-U}},\\ U &= \frac{1}{\pi}\cdot(tan(MSF)+tan(AG)), \end{aligned} \tag{10}$$

where $a=0.6$, $b=0.5$ [26], MSF and AG denote modified spatial frequency and average gradient, respectively.

In our method, the fractional gradient energy function is used to calculate the weights of corresponding features. The calculation is as follows:

$$\begin{aligned} \omega^{i}_{dk}(m,n) &= \frac{FGE^{i}_{dk}(m,n)}{\sum_{k=1}^{K} FGE^{i}_{dk}(m,n)},\\ F^{i}_{df}(m,n) &= \sum_{k=1}^{K}\omega^{i}_{dk}(m,n)\times F^{i}_{dk}(m,n),\\ I^{i}_{df} &= R(F^{i}_{df}), i=[1,2,\cdots,L]. \end{aligned} \tag{11}$$

After obtaining the fused detail layers and base layer, we can obtain the fused image by the following formula.

$$I_f = I_{bf}+\sum_{i=1}^{L} I^{i}_{df}. \tag{12}$$

4 Experimental Results

4.1 Datasets and implementation details

We evaluate the performance of the proposed method on two public datasets, namely TNO natural scene dataset [21] and RoadScene dataset [23]. The datasets used are pre-registered images. And for the decomposition layers of MDLatLRR, we set $L=4$ similar to Ref [8].

Our method is compared with 10 state-of-the-art image fusion methods. These methods include non-deep learning methods (DCHWT [5], CVT [10], NSCT [10], ConvSR [12], GTF [13], MDLatLRR [8]) and deep learning methods (DenseFuse [6], FusionGAN [14], IFCNN [25], U2Fusion [23]). The parameters of each of 10 comparison methods are set according to the optimal parameter settings reported in the corresponding papers.

To quantitatively compare the proposed method with other existing fusion methods, six quality evaluation indexes are used. These are: entropy (EN) [1]; standard deviation (SD) [18]; correlation coefficient (CC); mutual information (MI) [4] indicates how many features are preserved in the fused image; MS-$SSIM$ [15] focuses on structural information; and the sum of the correlations of differences (SCD) [2]. The larger the values of these metrics, the better the fusion performance.

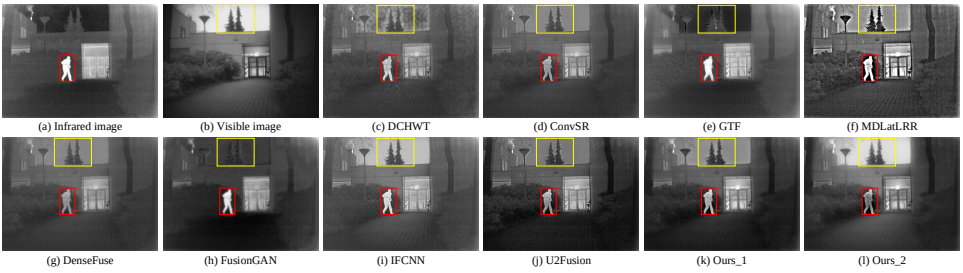

Figure 3: The fusion results of one pair of images on TNO dataset. Note "Ours_1" and "Ours_2" denote the proposed methods using parameters ($v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.3) and ($v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.9), respectively.

Table 1: Average quantitative results on TNO dataset (red: the best, blue: the second best).

Methods		EN	SD	MI	CC	MS-SSIM	SCD
DCHWT [6]		6.5678	64.9789	13.1355	0.4272	0.8433	1.6099
ConvSR [12]		6.2587	50.7437	12.5174	0.4922	0.9028	1.6482
GTF [13]		6.6343	67.5436	13.2687	0.3228	0.8084	1.0049
MDLatLRR [8]		6.9774	83.7741	13.9555	0.4393	0.8517	1.6332
DenseFuse [6]		6.6716	54.3575	13.3431	0.4994	0.9290	1.8350
FusionGAN [14]		6.3629	54.3575	12.7257	0.4257	0.7318	1.4569
IFCNN [25]		6.5955	66.8758	13.1909	0.4659	0.9053	1.7138
U2Fusion [23]		6.7571	64.9116	13.5142	0.5010	0.9253	1.7984
Ours	$v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.3	6.8174	72.2746	13.6348	0.4880	0.9281	1.8291
	$v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.9	7.0294	93.1529	14.0588	0.3935	0.9065	1.5680

4.2 Performance comparison on TNO dataset

Figure 3 shows the qualitative fusion results of different methods on TNO dataset. The results show that all fusion methods can preserve the information of the source images to varying degrees. However, the target in the fused images obtained by DCHWT, ConvSR, DenseFuse and U2Fusion are not effectively highlighted (see the red regions). This is because these algorithms cannot effectively extract the target information of the infrared image into the fused image. Although GTF, MDLatLRR, FusionGAN and IFCNN can effectively highlight the target in the image, they cannot preserve the texture details of the source images (see the yellow regions). Especially for GTF and FusionGAN, too many texture features are lost in fused images. Compared with other methods, our proposed method shows a good balance between preserving details and highlighting targets.

Table 1 presents the quantitative results of all fusion methods on 21 pairs of images on TNO dataset. The results show that the proposed method performs the best on three evaluation indexes (*EN*, *SD* and *MI*). Moreover, the performance on *MS-SSIM* and *SCD* is sub-optimal. Especially, the objective indexes values of MDLatLRR [8] is lower than that of our method, and its *EN*, *SD* and *MI* are lower than those of *Ours*_2 ($v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.9). Its *CC*, *MS-SSIM* and *SCD* are lower than those of *Ours*_1 ($v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.3). For comparison with deep learning-based methods, our method is higher than U2Fusion in all evaluation indexes, except *CC*.

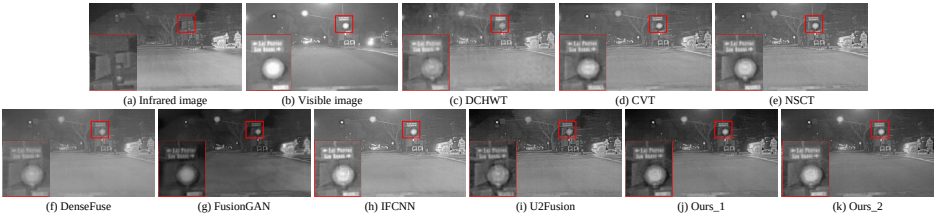

Figure 4: The fusion results of one pair of images on RoadScene dataset. Note "Ours_1" and "Ours_2" denote the proposed methods using parameters ($v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.3) and ($v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.9), respectively.

Table 2: Average quantitative results on RoadScene dataset (red: the best, blue: the second best).

Methods		EN	SD	MI	CC	MS-SSIM	SCD
DCHWT [6]		7.1710	63.0711	14.3420	0.4687	0.8283	1.4725
CVT [11]		7.0159	57.2032	14.0319	0.5255	0.8721	1.5710
NSCT [11]		6.9471	56.0915	13.8942	0.5387	0.9282	1.5965
DenseFuse [8]		6.6755	48.7363	13.3510	0.5583	0.8531	1.5607
FusionGAN [14]		7.1753	67.0645	14.3507	0.4416	0.7352	1.3753
IFCNN [25]		6.9730	56.8367	13.9461	0.5322	0.8798	1.5889
U2Fusion [23]		7.1969	68.0394	14.3938	0.5295	0.9250	1.7984
Ours	$v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.3	7.2357	70.8168	14.4714	0.5498	0.9058	1.8250
	$v_{f_{ir}}$=0.3,$v_{f_{vis}}$=0.9	7.2378	70.5262	14.4756	0.4825	0.8680	1.5497

The maximum *EN* and *SD* indicate that the fused images obtained by our method contain more information. The maximum *MI* shows that the fusion results obtained by the proposed method have higher mutual information similarity with the source images. The sub-optimal values on *MS-SSIM* and *SCD* indicate that the structural similarity and correlation between the fused images obtianed by our method and the source images is lower than that of DenseFuse, but better than the remaining methods.

4.3 Performance comparison on RoadScene dataset

Figure 4 shows the qualitative fusion results of different methods on RoadScene dataset. As shown in Figure 4, due to the introduction of too much interference information from the visible image, the 7 comparison methods lost some details, e.g., the car, the traffic light, the sign. DCHWT and FusionGAN introduce excessive noise into the fused images. The fusion results obtained by CVT, NSCT, DenseFuse and IFCNN have different degrees of contrast reduction, which are closer to the infrared images. It is not conductive to human observation. While the fused image obtained by U2Fusion has high contrast, the detials of the traffic light are some blurry (see the red regions). In contrast, our method provides more details and reduces noise effects through fractional optimization, thus preserving the complementary information of the source images more completely.

Table 2 shows the quantitative fusion results of different methods on RoadScene dataset. As shown in Table 2, our method achieves the best or sub-optimal results on all evaluation

Table 3: Analysis on the key components of the proposed method on TNO dataset.

Methods	SD	CC	MS-SSIM	SCD
MDLatLRR [8]	83.7741	0.4393	0.8517	1.6332
MDLatLRR+*model-1*	88.9609	0.4626	0.8567	1.7399
MDLatLRR+*model-2*	91.3343	0.4225	0.7877	1.5675
Ours	93.1529	0.4880	0.9281	1.8291

Table 4: Fusion results on TNO dataset using different parameter combinations ($v_{f_{ir}}$,$v_{f_{vis}}$).

$v_{f_{ir}}$	$v_{f_{vis}}$	EN	SD	MI	MS-SSIM	SCD
0.3	0.3	6.8174	72.2746	13.6348	0.9281	1.8291
	0.6	6.9920	87.9021	13.9841	0.9196	1.7472
	0.9	7.0294	93.1529	14.0588	0.9065	1.5680
0.6	0.6	6.8527	74.1245	13.7054	0.9267	1.7948
	0.9	6.8359	77.1384	13.6718	0.9175	1.6886
	1.2	6.7987	76.3516	13.5973	0.9167	1.7118
0.9	0.9	6.8537	79.2122	13.7074	0.9231	1.7675
	1.2	6.8265	77.4651	13.6530	0.9222	1.7574
	1.5	6.8398	77.8568	13.6796	0.9216	1.7610
1.2	1.2	6.8243	77.0069	13.6486	0.9236	1.7516
	1.5	6.8292	78.3813	13.6584	0.9237	1.7552
	1.8	6.8427	78.5622	13.6854	0.9235	1.7604
1.5	1.5	6.8210	76.7004	13.6421	0.9234	1.7483
	1.8	6.8285	77.4367	13.6570	0.9233	1.7521
1.8	1.8	6.8216	76.9553	13.6433	0.9230	1.7494

metrics except *MS-SSIM*. Since our method removes some interference structural information through fractional optimization control, it does not have many advantages in the index *MS-SSIM* based on structural information, but its practical application effect is far greater than this. Furthermore, although neural network methods are commonly used for image fusion today, as shown in Table 2, the performance of DenseFuse, FusionGAN, IFCNN and U2Fusion is generally not better than our proposed method. The experimental results can illustrate that our method has high fidelity to the source images and highlights the target.

4.4 Analysis and discussion

The key components of the proposed method are the generation of the pre-fused image based on fractional optimization function (*model-1*) and detail layers fusion based on fractional gradient energy function (*model-2*). This section discusses the performance contribution of these components.

We first select MDLatLRR [8] as the baseline method. Then, we transplant *model-1* and *model-2* to MDLatLRR respectively to verify their effectiveness on fusion performance. The experimental results are shown in Table 3. It can be seen that the indexes of the transplanted methods are mostly higher than those of MDLatLRR. Moreover, our final model achieves

the best among the three methods. This further illustrates the effectiveness of our method in infrared and visible image fusion.

Furthermore, the parameter settings are very important in the fractional optimization function. Since different combinations of $v_{f_{ir}}$ and $v_{f_{vis}}$ lead to different fusion results, we analyze the impact of these two parameters on the fusion results from the range of $[0.3, 1.8]$. The objective evaluation values obtained by different combinations are shown in Table 4. As can be seen from Table 4, when $v_{f_{ir}}$=0.3, $v_{f_{vis}}$=0.3/0.9, the fusion result obtained by our method performs the best on five metrics. Therefore, based on this analysis, we set the parameters of the model as $v_{f_{ir}}$=0.3, $v_{f_{vis}}$=0.3/0.9.

5 Conclusions

In this paper, we propose an infrared and visible image fusion method based on fractional optimization model. The method consists of two stages, which can efficiently extract target features and detail textures of source images. Firstly, a pre-fused image is generated by the fractional gradient optimization function. Secondly, the base layer of the pre-fused image obtained by using the multi-layer decomposition tool is used as the fused base layer. For the fusion of detail layers, the detail information is evaluated by the designed fractional gradient energy function to retain important detail features. The experimental results show that the proposed method has good qualitative and quantitative performance on two public datasets (TNO and RoadScene). Overall, our method can better highlight targets while retaining more useful details from source images. Moreover, the proposed method can also be extended to various computer vision applications such as infrared target detection and RGB-T tracking.

Acknowledgement

This work is partially supported by the National Natural Science Foundation of China (No. 61773210).

References

[1] V. Aardt. Assessment of image fusion procedures using entropy, image quality, and multispectral classification. *Journal of Applied Remote Sensing*, 2(1):1–28, 2008.

[2] V. Aslantas and E. Bendes. 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, 2015.

[3] R. Bartels and G. Stewart. Solution of the matrix equation AX + XB = C [F4]. *Communications of the ACM*, 15(9):820–826, 1972.

[4] M. Hossny, S. Nahavandi, and D. Creighton. Comments on 'information measure for performance of image fusion'. *Electronics Letters*, 44(18):1066–1067, 2008.

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

[6] H. Li and X. Wu. DenseFuse: A fusion approach to infrared and visible images. *IEEE Transactions on Image Processing*, 28(5):2614–2623, 2018.

[7] H. Li, X. Wu, and T. Durrani. NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models. *IEEE Transactions on Instrumentation and Measurement*, 69(12):9645–9656, 2020.

[8] H. Li, X. Wu, and J. Kittler. MDLatLRR: A novel decomposition method for infrared and visible image fusion. *IEEE Transactions on Image Processing*, 29:4733–4746, 2020.

[9] H. Li, X. Wu, and J. Kittler. RFN-Nest: An end-to-end residual fusion network for infrared and visible images. *Information Fusion*, 73:72–86, 2021.

[10] G. Liu and S. Yan. Latent low-rank representation for subspace segmentation and feature extraction. In *2011 International Conference on Computer Vision*, pages 1615–1622, 2011.

[11] Y. Liu, S. Liu, and Z. Wang. A general framework for image fusion based on multi-scale transform and sparse representation. *Information Fusion*, 24:147–164, 2015.

[12] Y. Liu, X. Chen, R. Ward, and Z. Wang. Image fusion with convolutional sparse representation. *IEEE Signal Processing Letters*, 23(12):1882–1886, 2016.

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

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

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

[16] J. Mei, Y. Dong, and T. Huang. Simultaneous image fusion and denoising by using fractional-order gradient information. *Journal of Computational and Applied Mathematics*, 351:212–227, 2019.

[17] Y. Pu, J. Zhou, and X. Yuan. Fractional differential mask: A fractional differential-based approach for multiscale texture enhancement. *IEEE Transactions on Image Processing*, 19(2):491–511, 2010.

[18] Y. Rao. Review article: In-fibre bragg grating sensors. *Measurement Science & Technology*, 8(4):355–375, 1997.

[19] A. Song, H. Duan, H. Pei, and L. Ding. Triple-discriminator generative adversarial network for infrared and visible image fusion. *Neurocomputing*, 483:183–194, 2022.

[20] L. Tang, J. Yuan, and J. Ma. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network. *Information Fusion*, 82:28–42, 2022.

[21] A. Toet. TNO image fusion dataset. 2014.

[22] J. Wang, J. Peng, X. Feng, G. He, and J. Fan. Fusion method for infrared and visible images by using non-negative sparse representation. *Infrared Physics & Technology*, 67:477–489, 2014.

[23] H. Xu, J. Ma, J. Jiang, X. Guo, and H. Ling. U2Fusion: A unified unsupervised image fusion network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44 (1):502–518, 2022.

[24] K. Zhang, Y. Huang, X. Yuan, H. Ma, and C. Zhao. Infrared and visible image fusion based on intuitionistic fuzzy sets. *Infrared Physics & Technology*, 105:103124, 2020.

[25] Y. Zhang, Y. Liu, P. Sun, H. Yan, X. Zhao, and L. Zhang. IFCNN: A general image fusion framework based on convolutional neural network. *Information Fusion*, 54: 99–118, 2020.

[26] Kai Zuo, Tongjing Sun, Zhenhua Li, and Liang Tao. 2D fractional kalman filter and its application to image process. *Journal of Electronics & Information Technology*, 32 (12):3027–3031, 2010.