A meta-upscale based end-to-end depth framework for infrared and visible image fusion

Wemrui Niu
Xihua University
Mingwei Tang (tang@26.com)
Xihua University

Research Article

Keywords: Image Fusion, Convolutional Neural Network, Infrared and Visible Image, Multi-scale Feature

Posted Date: June 16th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1742199/v1

A meta-upscale based end-to-end depth framework for infrared and visible image fusion

Wenrui Niu · Mingwei Tang

the date of receipt and acceptance should be inserted later

Abstract In the recent years, infrared and visible image fusion has been actively explored due to its advantages in various vision-based applications. However, the existing fusion methods only use the output of the last layer in the coded network. The results of fusion are not good as most of the important information in the intermediate layers is lost. In this paper, a meta-upscale based end-to-end depth framework for infrared and visible image fusion is proposed. The meta-upscale module can increase the resolution of the multi-scale features without repeatedly training the whole model. Firstly, in the proposed Multi-scale Feature Extraction Module (MFEM), the multi-scale deep features of each source image are extracted and up-scaled by a meta-upscale module. Secondly, in the Multi-scale Feature fusion Module (MFM), the *L1-Norm* strategy-based feature fusion method is developed to fuse the features based on the scaling. Thirdly, the Multi-scale Feature Compensation Module (MFCM), it has dense skip connections, which can preserve significant amounts of information from the input data in a multi-scale perspective. In addition, a new content retention loss is proposed to further improve the contrast enhancement approach. Experiments demonstrate that the proposed method can achieve significant results compared tothe state-of-the-art algorithms.

Wenrui Niu
School of Computer and Software Engineering, Xihua University, Chengdu, Sichuan Province, China
Tel.: 18143768893
E-mail: 18143768893@163.com

Mingwei Tang
School of Computer and Software Engineering, Xihua University, Chengdu, Sichuan Province, China
Tel.: 18081041450
E-mail: tang4415@126.com

Keywords Image Fusion · Convolutional Neural Network · Infrared and Visible Image · Multi-scale Feature

1 Introduction

Image fusion is a kind of enhancement technology, which fuses the images collected by different types of sensors to get an image with stronger explanatory power and richer information. Convolutional Neural Network (CNN) is used to obtain the image features and reconstruct the fused image . In 2017, Liu *at al.* [20] proposed multi-focus image fusion based on deep learning. Prabhakar *at al.* [26] proposed a multi exposure fusion problem based on CNN network system structure, because the coding network only uses the results of the last layer, the important features in the middle are lost. And with the increase of network depth, the more information lost, the worse the result. He *at al.* [9] introduced a deep residual learning framework. To further improve the information flow between layers, Huang *at al.* [10] proposed a new dense block structure, in which any layer is directly connected to the subsequent layers. Li *at al.* [13] proposed a new deep learning structure, which uses the pretrained VGG network to extract multi-scale depth features. The image fusion method based on convolutional neural network has a deep network structure, which leads to the loss of most information. It also has some limitations, such as the large amount of calculation and the need to design the fusion rules manually. Therefore, we propose a meta-upscale based end-to-end depth framework for infrared and visible image fusion, Fig.1 shows the framework, the main contributions of this work are summarized into the following four folds:

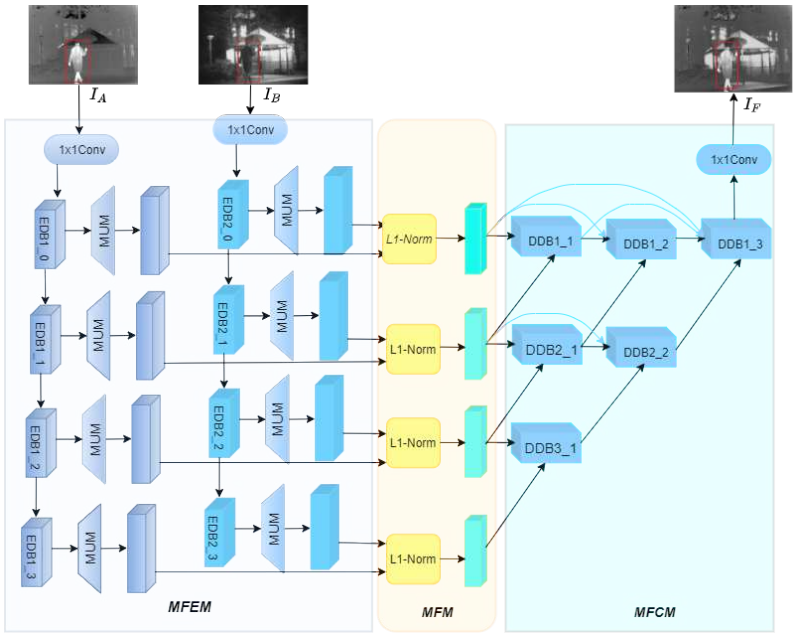


Fig. 1 The framework of the proposed a meta-upscale based end-to-end depth framework for infrared and visible image fusion.

(1)Unlike most existing image fusion methods, our method extracts multi-scale features and performs super-resolution techniques before fusion, and achieves resulting in better fusion results.

(2)Dataset: To enhance the diversity and robustness of the dataset, the collected source images are cropped into 58923 images of $256 * 256$ size to form a new training dataset.

(3)The dense skip connection manners are adopted to further improve the capacity of feature extraction and retain more information at different scales.

(4)We formulate he content retention-based loss function, which is helpful to learn more effective features, enhance the contrast of the fused images.

2 Related works

Image fusion algorithms are broadly classified into two types: traditional method and deep learning-based method. The traditional method includes multi-scale decomposition method , and sparse representation method . Multi-scale decomposition method [17] usually decomposes the source image into different scales to extract features by using specific transformation tools and the appropriate fusion strategy is used to fuse the features of each scale, and the inverse operator is used to reconstruct the fused image. Although these methods show good fusion performance, their performance highly dependents on the multi-scale approach of extracting feature images manually, which increases the complexity of the algorithm. Recently, the sparse coding method has been applied to various signal processing fields [31] to process two-dimensional images. Usually, the sparse representation based fusion scheme for infrared and visible images consists of three steps [31], firstly, each source image is decomposed into several overlapping modules by using a sliding window strategy. Secondly, an over complete dictionary is learned from multiple high-quality natural images, and sparse coding is performed on each module to obtain sparse representation coefficients, which transforms high-dimensional data into low dimensional space. Thirdly, the sparse representation coefficients are fused according to *L1* parameterization, choose-max, weighted average and other fusion rules, but the image quality after fusion is poor. Lu *at al.* [24] proposed a fusion scheme based on target separation and sparse representation. Yang *at al.* [6] constructed a discrete cosine transform dictionary to represent and fuse the source images. Li *at al.* [12] designed a fusion method combining low-rank representation and dictionary learning. The SR fusion methods achieve good performance, but these methods have shortcomings: (1) the running time of fusion algorithm is very dependent on dictionary learning algorithm; (2) when the source image is very complex, it will lead to performance degradation.

In order to overcome these shortcomings, convolutional neural network (CNN) has made many breakthroughs in the field of computer vision and image processing in recent years. In 2017, Liu *at al.* [20] proposed a multi focus image fusion scheme, which uses siamese network to classify focusing and defocusing modules. Li

Table 1 The network settings of encoder and decoder network.

	Layer	Size	Stride	Channel(input)	Channel(out)	Activatation
Encoder	Conv	1	1	2	16	ReLU
	EDB1-0	-	-	16	64	-
	EDB2-0	-	-	64	112	-
	EDB3-0	-	-	112	160	-
	EDB4-0	-	-	160	208	-
Decoder	DDB3-1	-	-	368	160	-
	DDB2-1	-	-	272	112	-
	DDB2-2	-	-	384	112	-
	DDB1-1	-	-	176	64	-
	DDB1-2	-	-	240	64	-
	DDB1-3	-	-	240	64	-
EDB	Conv	-	-	304	64	ReLU
	Conv	-	-	2	16	ReLU
	Max pooling	-	-	16	1	-
DDB	Conv	-	-	Nin	16	ReLU
	Conv	-	-	16	Nout	ReLU

at al [32] decomposed the source image into two parts, fused the infrared image and visible image, and designed a suitable fusion strategy based on deep learning. In ICCV 2017, Prabhakar *at al.* [35] performed a CNN-based approach to the exposure fusion problem. They proposed a simple CNN-based architecture, in which the encoder contains convolution layers and the decoder contains three convolution layers. Two images are input into the network, and two feature mapping sequences are obtained by adding strategy, and then fused. Finally, the fused feature map is reconstructed. Although this method achieves better performance, there is still main drawbacks: only use the calculation results of the last layer in the coding network, and the useful information obtained in the middle layer is lost. With the increase of network depth, more information is lost.

Li *at al.* [13] used VGG networks to extract multi-layer deep features and constructed robust weight maps for fusion. However, they use the pretrained CNN model as a feature extraction tool, and CNN has not learned to integrate or select depth features adaptively. These methods have achieved good results, but they do not provide an end-to-end solution and still rely on manual settings. Most fusion methods need to design fusion rules manually, which increases the complexity of the algorithm. In order to solve the problem of important information loss and need to design fusion rules manually. we propose a meta-upscale based end-to-end depth framework for infrared and visible image fusion. Our approach can retain more useful information from the middle layer and is easy to train.

3 Methodology

Our method aims to generate images with rich thermal information and texture details. In this section, we will demonstrate a meta-upscale based end-to-end depth framework for infrared and visible image fusion. Our method consists of three parts: feature extraction module, fusion model and feature compensation module. The detail of process is shown in Fig.1.

3.1 Multi-scale feature extraction module

The multi-scale feature extraction module (MFEM) consists of a feature extraction network (FEN) and a meta-upscale(MUM) module, which is equivalent to the encoder. The multi-scale feature extraction network extracts features from each source input image, and the meta-upscale module upgrades the feature map to different multi-scale resolution.

Our Feature Network (FEN) has two branches, each of which contains a 1x1 convolution layer and four feature Extraction DenseBlocks. (EDB0, EDB1, EDB2, EDB3). Each block contains two convolution layers and a max-pooling operator, which ensures that the coding network can extract multi-scale depth features. The parameters are shown in Table I.

Meta-Upscale Model:In order to upgrade the feature maps of different scales, the meta-upscale module (MUM) is adopted as the up-sampling layer. Specifically, the MUM consists of three steps, including location projection, weight prediction, and feature mapping. Fig.2 shows the positional projection of pixels between the low resolution image and its corresponding

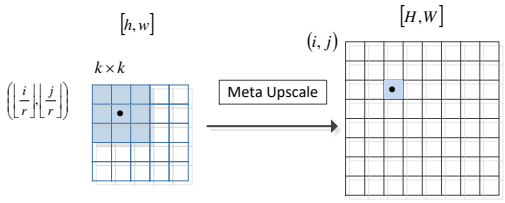


Fig. 2 Location projection of LR to HR

high resolution version, with a scale factor of r. which is calculated by Eq.1.

$$I^{SR}(i,j) = \Phi\left(F^{LR}(i',j'), W(i,j)\right) \tag{1}$$

where SR is the high-resolution feature, and LR is the low-resolution feature. Location Projection: For each pixel in SR, the position projection module should find the corresponding LR image. The two pixel mapping is calculated by Eq.2.

$$(i',j') = T(i,j) = \left(\left\lfloor \frac{i}{r} \right\rfloor, \left\lfloor \frac{j}{r} \right\rfloor\right) \tag{2}$$

where T is the transpose function, and for each pixel in SR, we find a unique pixel LR, which we will be considered as the two most relevant pixels.

Weight Prediction: A typical up-sampling module can pre-define the number of convolution kernels and the weights obtained from the training set learning. In the Meta-Upscale module, we use a network to predict the weights of the convolution kernels, which is calculated by Eq.3.

$$W(i,j) = \varphi(V_{ij};\theta) \tag{3}$$

where W is the weight of the convolution kernel corresponding to pixel SR. φ is the weight prediction network and θ is the weight of the weight prediction network. V_{ij} is the vector associated with i, j, the V_{ij} is calculated by Eq.4.

$$V_{ij} = \left(\frac{i}{r} - \left\lfloor \frac{i}{r} \right\rfloor, \frac{j}{r} - \left\lfloor \frac{j}{r} \right\rfloor, \frac{1}{r}\right) \tag{4}$$

Feature Mapping: Finally, the feature is projected to the pixel value of the SR. The feature mapping process is shown in Fig.3. Corresponding pixel. The matrix product is used as a Feature Mapping function, as shown in Eq.5.

$$\Phi\left(F^{LR}(i',j'), W(i,j)\right) = F^{LR}(i',j')\,W(i,j) \tag{5}$$

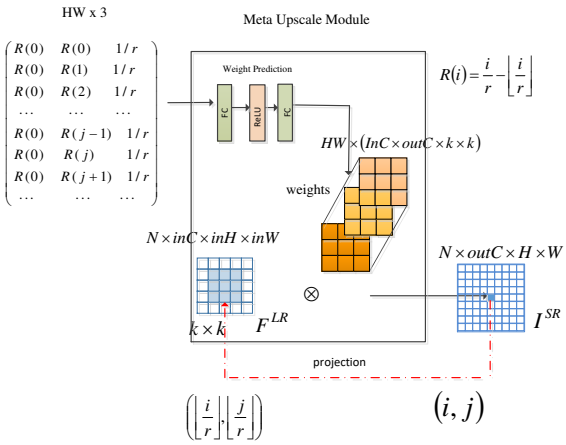


Fig. 3 The structure of the meta-upscale module

3.2 Multi-scale fusion module

L1-Norm Strategy: In order to improve the fusion efficiency, the block-based averaging method based on *L1-Norm* and softmax operation is applied to our network, as shown in Figure 5. Firstly, the activity level of the upgraded feature map is measured to get the initial activity level feature map. Thus, the initial activity level map C is calculated by Eq.6.

$$C(x,y) = \left\|\phi_i^{1:M}(x,y)\right\|_1 \tag{6}$$

where $\phi_i^{1:M}(x,y)(i=1,\cdots,k)$ indicates the upgraded feature map, and M is the number of channels. $k \geq 2$ indicates the index of feature maps which are obtained from input images. Then block-based average operator

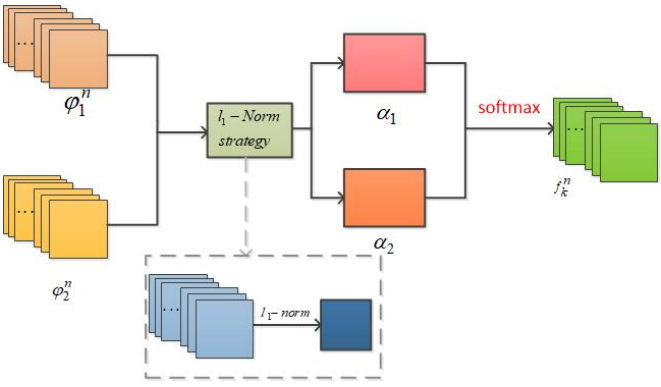


Fig. 4 *L1-Norm* Strategy

is utilized to calculate the final activity level map $\hat{C}$ by Eq.7.

$$\hat{C}(x,y) = \frac{\sum_{a=-r}^{r}\sum_{b=-r}^{r} C_i(x+a,y+b)}{(2r+1)^2} \tag{7}$$

Where r determines the block size and and the value of r is 1.

After we get the final activity level map $\hat{C}$, f^m is calculated by Eq.8.

$$f^m(x,y) = \sum_{i=1}^{k} \omega_i(x,y) \times \phi_i^m(x,y), \\ \omega_i(x,y) = \frac{\hat{C}_i(x,y)}{\sum_{n=1}^{k} \hat{C}_n(x,y)} \tag{8}$$

The final fused image will be reconstructed by a feature compensation module in which the fused feature maps f^m as the input.

3.3 Loss function

Finally, The loss function consists of Structural Similarity (SSIM) and Content Retention Loss (CRL) to ensure that the rich information in the final fused image is retained.In this work, to improve the effect of feature learning, I_A and I_B are fed into the VGG16 pretrained network. The output feature map is before each Max-pooling layer, which are shown in Fig.5 as $\varphi c1(I) \ldots \varphi c5(I)$. Then information measurements of

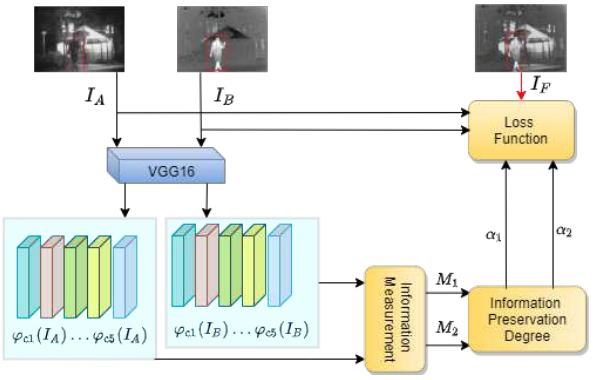


Fig. 5 The loss function calculation process

these feature maps are calculated to generate two measurements. $M_I(I = 1, 2)$ is calculated by Eq.9.

$$M_I = \frac{1}{5} \sum_{i=1}^{5} \left\| c_i^k(I) \right\|_F^2 \tag{9}$$

Where $c_i(I)$ denotes the layer of features. k denotes the feature map in the *kth* channel of channels. And $\|\|_F$ denotes the Frobenius norm.

Two adaptive weights are calculated as the content retention. α_1 and α_2 are calculated by Eq.10.

$$[\alpha_1, \alpha_2] = softmax(M_1, M_2) \tag{10}$$

The sum of α_1 and α_2 is 1. Then the loss function is calculated by α_1 and α_2. The L_{ssim} is obtained by Eq.11.

$$L_{ssim} = \alpha_1 \left(1 - ssim\left(I_A, I_F\right)\right) + \alpha_2 \left(1 - ssim\left(I_B, I_F\right)\right) \tag{11}$$

where $SSIM(.)$ denotes the structural similarity measure. Content loss is calculated for fused and input images. L_{CRL} is obtained as Eq.12.

$$L_{CRL} = \alpha_1 \left(\|I_F - I_A\|_F^2 \right) + \alpha_2 \left(\|I_F - I_B\|_F^2 \right) \tag{12}$$

I_F represents the fused image, I_A represents infrared images, I_B represents visible light images, and $\|\|_F^2$ represents the F-parameter. λ is a hyperparameter. The loss function equation is shown in Eq.13.

$$Loss = L_{ssim} + \lambda L_{CRL} \tag{13}$$

3.4 Dataset

In the field of image fusion, the aligned source image is used as the input image, and the final image fusion is realized through feature extraction and reconstruction. However, limited by imaging equipment and military use of such images, publicly available infrared and visible image pairs are very sparse. In this experiment, 25 pairs of infrared and visible images are obtained from TNO dataset. The infrared image has obvious intensity contrast, while the visible image has rich detail information. It is difficult to get stable and robust model only by training the original image. As a result, we decided to expand the dataset and use it as a training base. The most straight forward method is to split the source image into $256 * 256$ size images, and finally get 98523 images to form the dataset. The process of making the dataset is shown in Fig.6.

4 Experimental results and analysis

Fig.7 and Fig.8 illustrates the fused results. The structure of CBF [1], ConvSR [33] loses a lot of information, such as billboards. LRR [15], WLS [22], Resnet50 [16], IFCNN [34], Deep Learning [13], DeepFuse [26], DenseFuse [11] fusion results are results in low contrast and resolution.Our method not only preserves texture detail better, the red box has a nice visual effect, the fusion results have more reasonable brightness information.In most cases, the subjective visual perception system is easily affected by human factors, such as personal emotion and vision. Here, we quantitatively compare our method with other existing methods. Five evaluation metrics, including information entropy (EN) [2], standard deviation (SD) [28], visual information fidelity (VIF) [8],mutual information [25](MI), structural similarity index measure (SSIM) [36], FMI_{dct} and FMI_M [7].

For the objective evaluation, seven objective metrics to evaluate our method. The average values of the fused

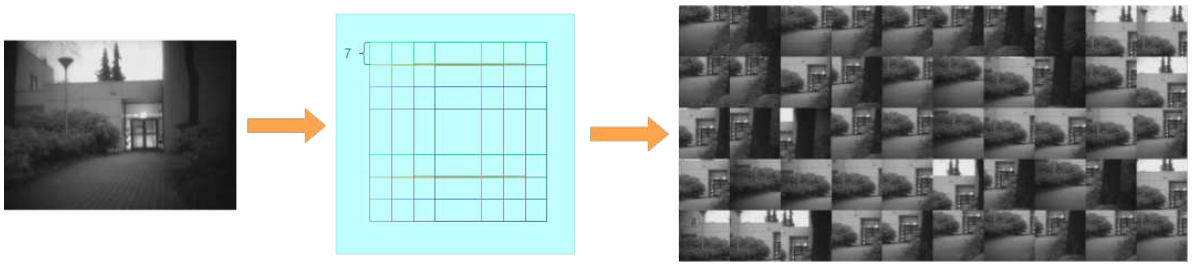
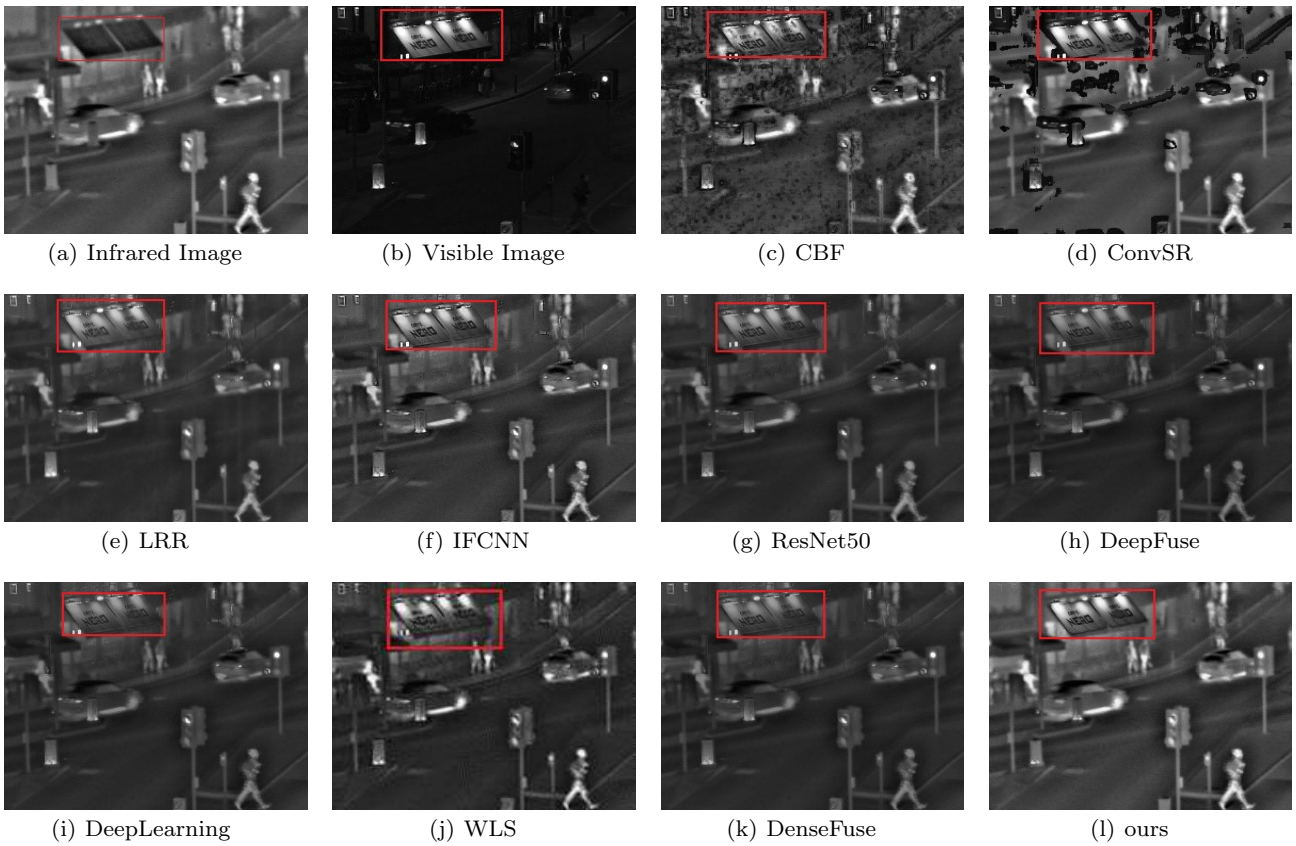


Fig. 6 The process of making a new dataset



(a) Infrared Image (b) Visible Image (c) CBF (d) ConvSR

(e) LRR (f) IFCNN (g) ResNet50 (h) DeepFuse

(i) DeepLearning (j) WLS (k) DenseFuse (l) ours

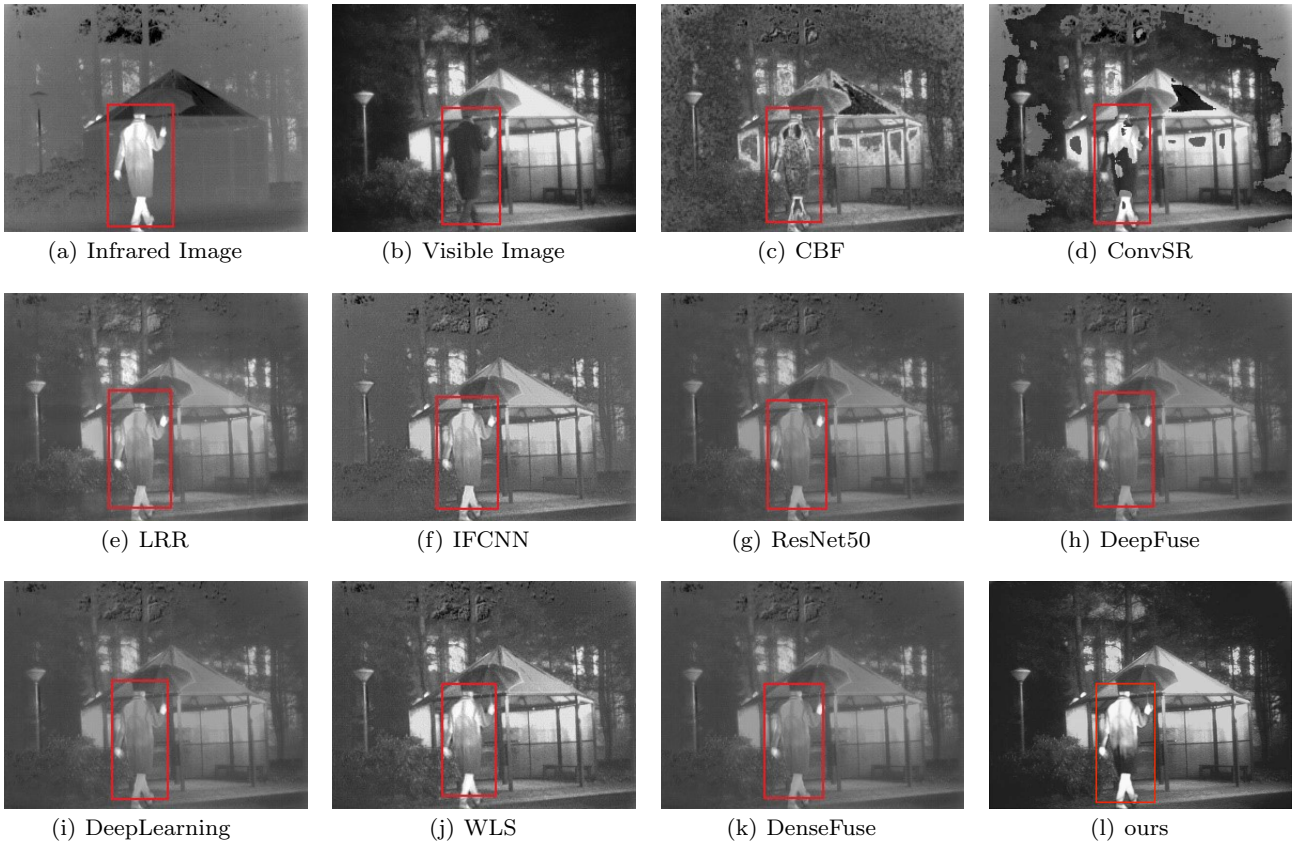
Fig. 7 Fused results of the "street" with different methods,(a) Infrared image; (b) Visible image; (c) CBF; (d) ConvSR; (e) LRR; (f) IFCNN; (g) ResNet50; (h)DeepFuse; (i) DeppLearning;(j) WLS; (k) DenseFuse; (l) Ours.

image quality measures are shown in Table II, where the best values are shown in bold.With the increase of MI, FMI_{dct} and FMI_M, the fusion method can retain more information and features of the original image. For SSIM and VIF, our method retains more structural information of the source image and generates more natural features,which shows that our method has a more advanced level of fusion.

5 Conclusion

In this paper, we proposed a meta-upscale based end-to-end depth framework for infrared and visible image fusion. The novelty mainly includes two aspects: 1) our method performs super-resolution processing on multi-scale features before fusion to obtain better multi-scale fusion effect. 2) In the feature compensation module, the short jump connection is used to improve the unit extraction ability. In addition, the content retention loss function enhance the contrast of the fused images, and improve the quality of final fusion results. Through the comprehensive comparison of nine latest methods, our method not only provide rich texture details, but also eliminate noise and holes.

Declarations Ethics approval and consent to participate: Not applicable.



(a) Infrared Image (b) Visible Image (c) CBF (d) ConvSR

(e) LRR (f) IFCNN (g) ResNet50 (h) DeepFuse

(i) DeepLearning (j) WLS (k) DenseFuse (l) ours

Fig. 8 Fused results of the "Kaptein" with different methods,(a) Infrared image; (b) Visible image; (c) CBF; (d) ConvSR; (e) LRR; (f) IFCNN; (g) ResNet50; (h)DeepFuse; (i) DeppLearning;(j) WLS; (k) DenseFuse; (l) Ours.

TABLE 2 The average values of quality metrics for 21 fused images

Methods	En	SD	MI	FMI_{dct}	FMI_M	$SSIM$	VIF
CBF [1]	6.86	76.82	13.71	0.26	0.32	0.60	0.71
ConvSR [33]	7.08	**98.32**	14.15	0.18	0.38	0.67	0.93
LRR [15]	6.35	52.2	12.7	0.43	0.43	0.72	0.31
WLS [22]	6.63	71.50	13.28	0.33	0.38	0.71	0.75
ResNet50 [16]	6.20	48.71	12.4	0.41	0.42	0.79	0.29
IFCNN [34]	6.60	67.29	13.18	0.37	0.40	0.73	0.86
DeepLearning [13]	6.19	48.16	12.36	0.40	0.41	0.77	0.30
DeepFuse [26]	6.70	66.92	13.40	0.41	0.42	0.74	0.66
DenseFuse [11]	6.17	47.82	12.35	0.41	0.42	0.78	0.27
Ours	**7.2**	80.03	**13.9**	**0.56**	**0.51**	**0.85**	**0.80**

Consent for publication: Not applicable.

Availability of data and materials:Analyses during the current study are available in the TNO.

Competing interests:The authors declare that they have no competing interests

Funding:This work is supported by the Scientic Research Funds project of Science and Technology Department of Sichuan Province (No.2019GFW131, 2020JY, 2021**), Funds Project of Chengdu Science and Technology Bureau (No. 2017-RK00-00026-ZF), the National Natural Science Foundation of China (No. 61902324), the Fund of Sichuan Educational Committee (17ZA0360), the Foundation of Cyberspace Security Key Laboratory of Sichuan Higher Education Institutions (No.sjzz2016-73) and Sichuan Youth Science and technology innovation research team(2021).

Authors' contributions:Niu Wenrui wrote the code, completed the laboratory, data analysis and wrote the paper. Tang Mingwei revised the manuscript.

Acknowledgements:Teacher Tang revised the guidance paper for my information. Thank you for helping

me when I encountered difficulties in the experiment. Thank you for my parents' cultivation.

References

1. Image fusion based on pixel significance using cross bilateral filter. Signal,Image & Video Processing (2015)
2. Aardt, V., Jan: Assessment of image fusion procedures using entropy, image quality, and multispectral classification. Journal of Applied Remote Sensing **2**(1), 1–28 (2008)
3. Ahmed, N.N., Natarajan, T., Rao, K.R.: Discrete cosine transform. IEEE Transactions on Computers **C-23**(1), 90–93 (2006)
4. Burt, P.J., Adelson, E.H.: The laplacian pyramid as a compact image code. Readings in Computer Vision **31**(4), 671–679 (1987)
5. Da, C., Arthur, L., Zhou, J.P., Do, M.N.: The nonsubsampled contourlet transform: Theory, design, and applications. imageprocessing (2006)
6. Gs, A., Af, B., Fcm, A., Sbs, C., Lb, D.: Image fusion techniques for remote sensing applications. Information Fusion **3**(1), 3–15 (2002)
7. Haghighat, M., Razian, M.A.: Fast-fmi: Non-reference image fusion metric. In: 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), pp. 1–3. IEEE (2014)
8. Han, Y., Cai, Y., Cao, Y., Xu, X.: A new image fusion performance metric based on visual information fidelity. Information fusion **14**(2), 127–135 (2013)
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. IEEE (2016)
10. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4700–4708 (2017)
11. Hui, Xiao-Jun: Densefuse: A fusion approach to infrared and visible images. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society (2018)
12. Hui, L., Wu, X.J.: Multi-focus Image Fusion Using Dictionary Learning and Low-Rank Representation. Image and Graphics (2017)
13. Hui, L., Wu, X.J., Kittler, J.: Infrared and visible image fusion using a deep learning framework. In: International Conference on Pattern Recognition 2018 (2018)
14. Jia, D., Wei, D., Socher, R., Li, L.J., Kai, L., Li, F.F.: Imagenet: A large-scale hierarchical image database. Proc of IEEE Computer Vision & Pattern Recognition pp. 248–255 (2009)
15. Li, H., Wu, X.J.: Infrared and visible image fusion using latent low-rank representation. arXiv preprint arXiv:1804.08992 (2018)
16. Li, H., Wu, X.j., Durrani, T.S.: Infrared and visible image fusion with resnet and zero-phase component analysis. Infrared Physics & Technology **102**, 103039 (2019)
17. Li, S., Kang, X., Fang, L., Hu, J., Yin, H.: Pixel-level image fusion: A survey of the state of the art. Information Fusion **33** (2017)
18. Li, S., Yang, B., Hu, J.: Performance comparison of different multi-resolution transforms for image fusion. In: Asia-pacific Computer Systems Architecture Conference (2008)
19. Li, S., Yin, H., Fang, L.: Group-sparse representation with dictionary learning for medical image denoising and fusion. IEEE Transactions on Biomedical Engineering **59**(12), 3450–3459 (2012)
20. Liu, Y., Chen, X., Peng, H., Wang, Z.: Multi-focus image fusion with a deep convolutional neural network. Information Fusion **36**, 191–207 (2017)
21. Luo, X., Zhang, Z., Zhang, B., Wu, X.: Image fusion with contextual statistical similarity and nonsubsampled shearlet transform. IEEE Sensors Journal **PP**(6), 1–1 (2017)
22. Ma, J., Zhou, Z., Wang, B., Zong, H.: Infrared and visible image fusion based on visual saliency map and weighted least square optimization. Infrared Physics & Technology (2017)
23. Nejati, M., Samavi, S., Shirani, S.: Multi-focus image fusion using dictionary-based sparse representation. Information Fusion **25**, 72–84 (2015)
24. Pei, H., Zhao, Y., Liu, H., Zhang, B., Lu, X.: The infrared and visible image fusion algorithm based on target separation and sparse representation. Infrared physics and technology (2014)
25. Peng, H., Long, F., Ding, C.: Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on pattern analysis and machine intelligence **27**(8), 1226–1238 (2005)
26. Prabhakar, K.R., Srikar, V.S., Babu, R.V.: Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs. IEEE Computer Society (2017)
27. Pu, Tian: Contrast-based image fusion using the discrete wavelet transform. Optical Engineering **39**(8), 2075–2082 (2000)
28. Rao, Y.J.: In-fibre bragg grating sensors. Measurement science and technology **8**(4), 355 (1997)
29. Vetterli, M.: Ieee transactions on image processing 1 the contourlet transform: An efficient directional multiresolution image representation (2008)
30. Wang, J., Peng, J., Feng, X., He, G., Fan, J.: Fusion method for infrared and visible images by using non-negative sparse representation. Infrared Physics & Technology **67**, 477–489 (2014)
31. Yang, B., Li, S.: Visual attention guided image fusion with sparse representation. Optik - International Journal for Light and Electron Optics **125**(17), 4881–4888 (2014)
32. Yang, J., Wright, J., Huang, T.S., Yi, M.: Image super-resolution as sparse representation of raw image patches. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2008), 24-26 June 2008, Anchorage, Alaska, USA (2008)
33. Yu, L., Xun, C., Ward, R.K., Wang, Z.J.: Image fusion with convolutional sparse representation. IEEE Signal Processing Letters **PP**(99), 1–1 (2016)
34. Yu, Z.A., Yu, L.B., Peng, S.C., Han, Y.A., Xz, D., Li, Z.A.: Ifcnn: A general image fusion framework based on convolutional neural network. Information Fusion **54**, 99–118 (2020)
35. Zhang, Z., Blum, R.S.: A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application. Proceedings of the IEEE **87**(8), 1315–1326 (1999)
36. Zhou, W., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process **13**(4) (2004)