Optica Applicata, Vol. LIII, No. 1, 2023
DOI: 10.37190/oa230104

Infrared and visible image fusion with deep wavelet-dense network

YANLING CHEN[1,2,3], LIANGLUN CHENG[1,2,3], HENG WU[1,2,3,*], ZIYANG CHEN[1,2,3], FENG LI[1,2,3]

[1]Guangdong Provincial Key Laboratory of Cyber-Physical System, School of Automation, Guangdong University of Technology, Guangzhou 510006, China

[2]College of Optical Sciences, University of Arizona, Tucson, AZ 85721, USA

[3]School of Computer, Guangdong University of Technology, Guangzhou 510006, China

[*]Corresponding author: heng.wu@foxmail.com

We propose a high-quality infrared and visible image fusion method based on a deep wavelet-dense network (WT-DenseNet). The WT-DenseNet includes three network layers, the hybrid feature extraction layer, fusion layer, and image reconstruction layer. The hybrid feature extraction layer is composed of a wavelet and dense network. The wavelet network decomposes the feature map of the visible and infrared images into low-frequency and high-frequency components, respectively. The dense network extracts the salient features. A fusion layer is designed to integrate low-frequency and salient features. Finally, the fusion images are outputted by an image reconstruction layer. The experimental results demonstrate that the proposed method can realize high-quality infrared and visible image fusions, and the performance of the proposed method is better than that of the six recently published fusion methods in terms of contrast and detail performance.

Keywords: infrared image, image fusion, image processing, infrared image enhancement.

1. Introduction

The infrared thermal imaging technology utilizes the thermal radiation of objects to reconstruct images, and has many advantages over the visible imaging technology, such as imaging targets in the environments of the low light, wind, sand, smoke, and so on [1]. However, infrared images usually encouter the problems of lacking details of the scenario [2] and low resolution [3]. Unlike the infrared thermal imaging technology, the visible imaging technology can capture the details of the scene, such as edges and textures, making the visible images more in line with the human vision [4]. Infrared and visible image fusion can generate a high quality and resolution image with both thermal and detail information [5], which has wide potential applications, such as the military detection, medical diagnosis, and remote sensing [6].

In recent years, many infrared and visible image fusion methods have been proposed. These methods can be divided into three categories, the multi-scale transform [7], deep learning [8], and other methods. Multi-scale transform methods mainly rely on the predefined transforms and corresponding levels for the decomposition and reconstruction [9]. The feature extraction and fusion rules of these methods are manually designed. However, the designs of diverse feature extraction and fusion rule make the fusion methods complicated [10]. Consequently, the performance of multi-scale transform methods is limited [7]. Nevertheless, deep learning methods can extract various features from source images without the manual intervention, simplify the fusion rules. For instance, LIU *et al*. [11] proposed an infrared and visible image fusion method based on the convolutional neural network. This method only took the result of the last layer as the image features, which lose many useful features obtained by the middle layer. To solve this problem, LI *et al*. [12] developed a DenseFuse network, where a dense block is adopted to get effective features from the infrared and visible image. However, the down-sampling strategies before fusion still inevitably led to the blurring of low -frequency information and the losing of visible texture details.

To address the low-frequency information blurring and visible texture detail losing problems caused by down-sampling strategies, we propose an infrared and visible image fusion method with a deep wavelet-dense network (WT-DenseNet). The proposed method includes three steps. Firstly, both infrared and visible images are fed into WT-DenseNet. After that, the infrared and visible images are filtered by a wavelet transform which is used to extract the low-frequency features. Meanwhile, the salient features of infrared and visible images are extracted by the dense network. The low -frequency features are then added to salient features. Secondly, a weighted-addition strategy is used to integrate features extracted from visible and infrared images. Different from many algorithms that use a weighted-average strategy to assign weights [13], in WT-DenseNet, the visible and the infrared images are given different weights within a certain range, and the thermal information weight of the infrared image is used to keep the weight of the visible image unchanged. Finally, the fused features are reconstructed by the deconvolution layer. The effectiveness of the proposed method is qualitatively and quantitatively validated.

The main contributions are described as follows.

1) We propose a high-quality infrared and visible image fusion method by using a WT-DenseNet. The WT-DenseNet uses a dense block to extract deep features, retain the details of the image texture, and thermal information. However, the dense block blurs and loses some parts of the low-frequency information. We design a wavelet transform block to extract low-frequency information, which makes up for the loss of low-frequency information, smooths the redundant image information, and realizes the optimization of image fusion.

2) We develop a hybrid feature extractor to extract multiple features. Multiple features include shallow features, deep features and low frequency features of the original image, and the fusion of multiple features improves the quality of the fused image.

3) We design a data preprocessing method that is used for the infrared and visible image registration and build an infrared and visible imaging system used to collect infrared and visible images for expanding the dataset.

2. Method

2.1. Experimental setup construction

An infrared and visible imaging system is built, as shown in Fig. 1. The imaging system includes an infrared module (IM), visible module (VM), and personal computer (PC). The IM is the ThermoX1 (uncooled infrared thermal imaging module, focal length 4 mm) which is used to obtain infrared images (resolution 320×240 pixels). Here, the IM captures the infrared signals whose wavelengths are in the range of 8–14 μm. The VM is composed of a visible camera (FL3-U3-32S2C-CS) and a lens (focal length 6 mm, focal range 0.1 mm ~ ∞) which is utilized to obtain visible images (resolution 1920×1080 pixels). The fields of view (horizontal × vertical) of the infrared and visible cameras are 81.9°×61.2° and 78°×78°, respectively.

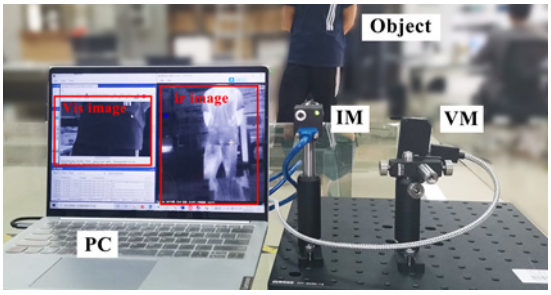

Fig. 1. Infrared and visible imaging system.

2.2. Data preprocessing

Before the image fusion, images from different source should be strictly aligned [14]. Since the resolution of the infrared and visible images collected by the experimental setup is different, thus the infrared and visible images are registered [15]. Figure 2 shows an example of the infrared and visible image registration, where the image (infrared and visible) fragmentary and saturated are not considered in the image registration process. Additionally, the image registration is achieved in a pixel-by-pixel manner. Image processing is divided into three steps. Due to the low-resolution problem of the infrared sensor, the bicubic interpolation alorithm [16] is firstly used to magnify the infrared image (320×240 pixels) by three times, and an infrared image with a resolution of 960×720 pixels is obtained. That is to say, the original IR image (320×240 pixels) is rescaled three times (×3) to a larger one (960×720 pixels). Secondly, because the region of interest area with coordinates (565, 200, 395, 160) in the

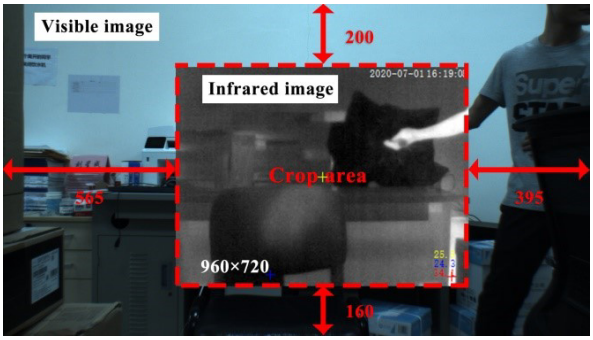

Fig. 2. Example of the infrared and visible image registration.

visible image (size 960×720 pixels) just coincides with the infrared image after three times magnification, hence the area of the visible image is cropped as the registered visible image. Note that the infrared images have lots of redundant information, and the bicubic interpolation also brings the redundant information to the image. To solve these problems, a non-local means denoising algorithm is used to smooth the infrared images [17]. Here, the standard deviation of the noise σ, patch window p, research window s and filtering parameter h are set as $\sigma = 40$, $p = 7 \times 7$, $s = 35 \times 35$, and $h = = 0.35\sigma = 14$ [17].

2.3. Deep wavelet-dense network

The network architecture of the proposed WT-DenseNet consists of three layers, the hybrid feature extraction layer, fusion layer, and image reconstruction layer, as shown in Fig. 3. The input preregistered infrared and visible images are denoted as K_1 and K_2, respectively. Here the infrared, visible, and fused images are all supposed to be grayscale images. The infrared and visible image fusion with WT-DenseNet includes three steps, which are shown as follows.

Step 1: Extracting multiple features from the infrared and visible images using a hybrid feature extraction layer.

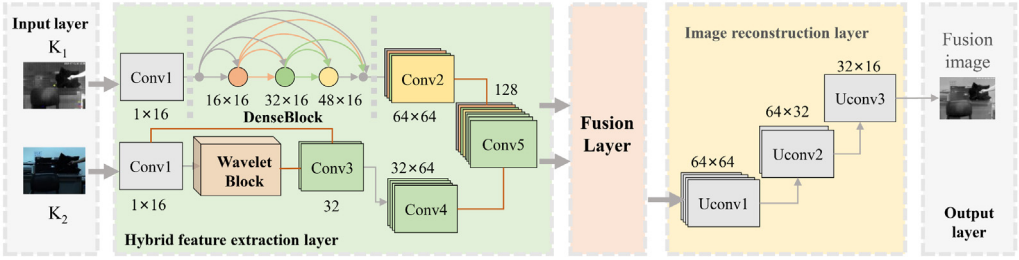

Fig. 3. The network structure of the proposed WT-DenseNet. K_1, infrared image; K_2, visible image; Conv, convolution layer; Uconv, deconvolution layer.

The hybrid feature extraction layer is shown in Fig. 3. Firstly, both the infrared and visible images are inputted into the first layer (Conv1) which uses 3×3 filters and rectified linear unit (ReLU) operations to extract rough features. This can solve the problem of low extraction efficiency and low feature quality in traditional manual feature extraction methods. Then DenseBlock and WaveletBlock are respectively used to obtain the low-frequency and salient features of infrared and visible images, by which the multiple features are constructed. Finally, feature channels of the DenseBlock and WaveletBlock are used as the splicing dimension. Specially, the feedforward connection method [18] is used for the sake of increasing feature contextual information, where short direct connections are established between each layer and all layers by feedforward.

The DenseBlock contains three dense layers, each of which uses 16 channels as the output of the high-dimensional mapping (stride: 1). The three dense layers are densely stacked. Besides, the dense layers are composed of 3×3 filters and ReLU. DenseNet can establish connections between different layers, thereby enhancing feature reuse. As shown in Fig. 4, the WaveletBlock that includes five wavelet (WT) layers (WT1, WT2, WT3, WT4, and WT5) is used to extract the useful information. The first four WT layers are composed of a convolution module with 3×3 filters, batch normalization (BN), and ReLU.

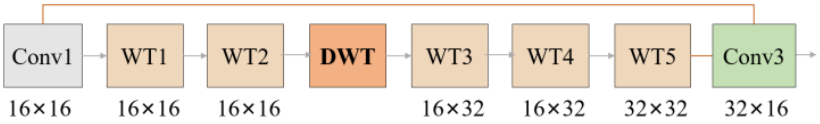

Fig. 4. The detail structure of WaveletBlock; Conv, convolution block; WT, wavelet block; DWT, discrete wavelet transforms.

The last layer WT5 only possesses a convolution layer which is used to generate the residual image. Here, the smooth features are first extracted by WT2 and then processed by the two-dimension (2D) discrete wavelet transform (DWT). DWT can separate the image into low frequency and high frequency components. Low-frequency components represent areas with slow changes in brightness and grayscale values, while high-frequency components correspond to drastic changes in the image, such as edges and details. After that, the low-frequency information is wrapped in the WT3 layer for output. Specifically, in DWT, four filters, *i.e.*, f_{LL}, f_{LH}, f_{HL}, and f_{HH}, are used to convolve the rough features [19]. Here, f_{LL} signifies the low-pass filter, f_{LH}, f_{HL}, and f_{HH} denote high-pass filters. Note that the Haar wavelet is adopted as the default wavelet function in WaveletBlock. In the two demensional Haar wavelet functions, the filters f_{LL}, f_{LH}, f_{HL}, and f_{HH} are defined as follows,

$$f_{LL} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad f_{LH} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}, \quad f_{HL} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}, \quad f_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{1}$$

The low-frequency filter f_{LL} and feature maps x are choosen for the convolution operation,

$$x_{\mathrm{LL}} = (f_{\mathrm{LL}} \otimes x)\downarrow_2 \tag{2}$$

where $\otimes$ is the convolution operator, $\downarrow_2$ denotes a standard down-sampling operator with a factor of 2. Given a feature map x with a size of $m \times n$, the (m, n)-th value of x_{LL} after 2D Haar transform can be written as,

$$x_{\mathrm{LL}}(i, j) = x(2i - 1, 2j - 1) + x(2i - 1, 2j) + x(2i, 2j - 1) + x(2i, 2j) \tag{3}$$

where i and j denote the index of pixel positions. Through the Haar transform in the low-frequency band, the features of the low-frequency band are integrated and outputted.

Step 2: Integrating the features of infrared and visible images by a weighted-addition strategy with the fusion layer.

In the testing phase, the infrared and visible images are fed into the hybrid feature extraction layer, and multiple feature maps are generated. A weighted-addition strategy is adopted to combine multiple feature maps, given by

$$f^m(i, j) = \alpha K_1^m(i, j) + K_2^m(i, j) \tag{4}$$

where (i, j) denotes the pixel coordinate in the feature maps, m is the number of feature maps, $m = \{1, 2, 3, ..., 128\}$, K_1^m and K_2^m are the infrared and visible feature maps, respectively. Note that α is a balance coefficient. The infrared thermal information in the fused image can be changed by adjusting the balance coefficient α.

Step 3: Reconstructing high-resolution feature maps by a deconvolution with the image reconstruction layer.

The image reconstruction layer is to reconstruct the feature maps outputted by the fusion layer. As shown in Fig. 3, the image reconstruction layer includes three deconvolution layers (Uconv1, Uconv2, Uconv3). In each layer, a deconvolution is used for up-sampling and to output the fused image.

3. Simulation results and analysis

In the training phase, the MS-COCO [20] dataset which contain 80000 color images is used as training images. The color images are changed into gray scale images. The reason for using the MS-COCO dataset is that the infrared images (gray scale) are insufficient. Since the purpose of the training is to extract the features of the images, thus training with MS-COCO dataset can achieve the same effect as that trained by the dataset containing the infrared and visible image pairs. To reduce the training time, the training images are resized to 256×256 pixels. Since the purpose of training is to obtain a feature extractor, thus only the hybrid feature extraction layer and image reconstruction layer are considered in the training phase. Here, the fusion layer is dis-

carded. As for the configurations of the loss function, we follow the method in DenseFuse [12], given by

$$\text{loss} = L_{\text{ssim}} + L_{\text{p}} \tag{5}$$

where L_{ssim} and L_{p} are respectively the structural similarity loss and pixel loss,

$$\begin{cases} L_{\text{ssim}} = 1 - \text{SSIM}(O, I) \\ L_{\text{p}} = \|O - I\|_2 \end{cases} \tag{6}$$

where O and I respectively represent output and input images, SSIM$(\cdot)$ denotes structural similarity, which represents the structural similarity of O and I; $\|\cdot\|_2$ indicates the Euclidean distance between O and I. The batch size and epochs are set as 4 and 100, respectively. To balance the thermal information and detailed information of the fusion results, α is set as 0.5 which is obtained by various experiments.

To test the effectiveness and robustness of the WT-DenseNet, ten infrared-visible image pairs from the public dataset [21] are tested. The results of WT-DenseNet are compared with those of other six fusion methods, including DRTV [22], BGR [23], weighted least square (WLS) [24], convolutional traditional wavelet transform (WT), U2Fusion [25] and DenseFuse [12]. To show the original and default fusion performance of each method, we do not modify or improve the parameter configurations of

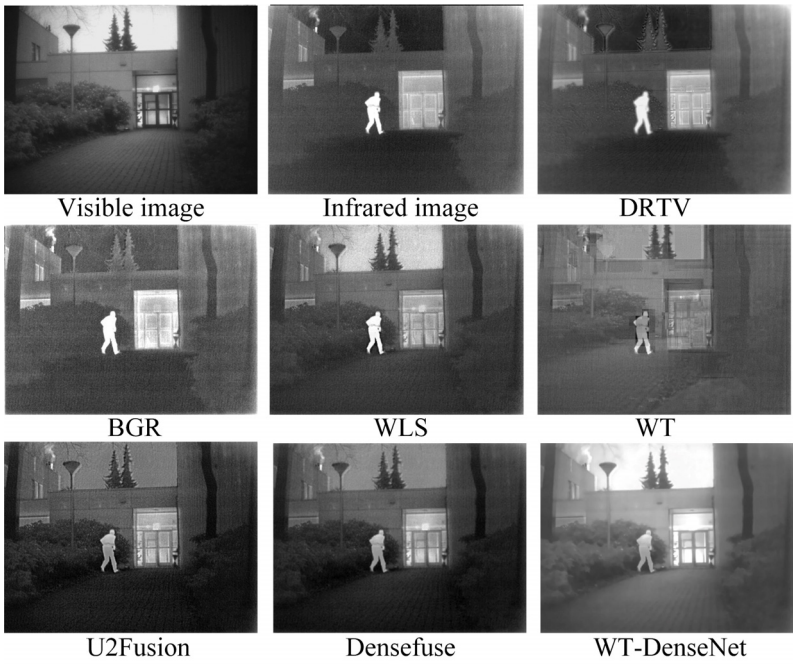

Fig. 5. Test images and fusion results of “person and door”.

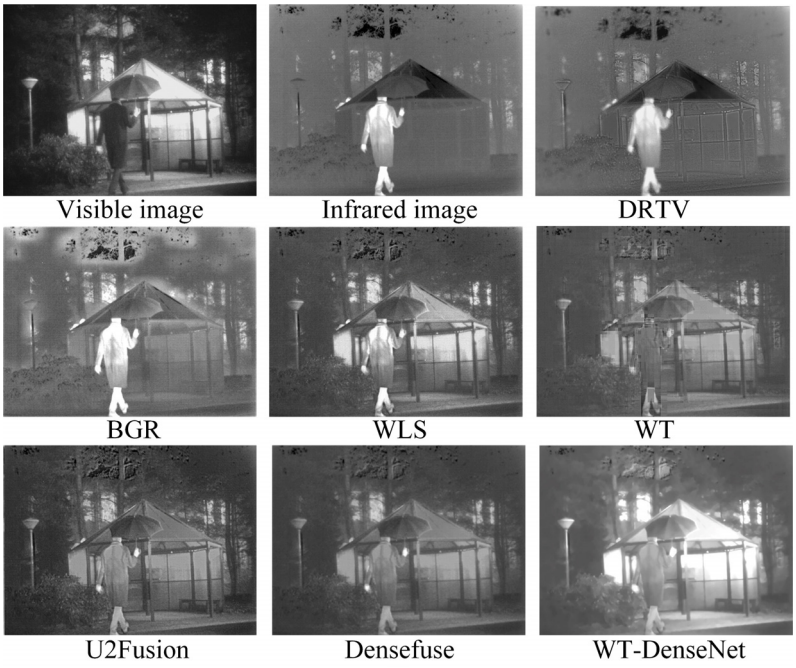

Fig. 6. Test images and fusion results of “person and tent”.

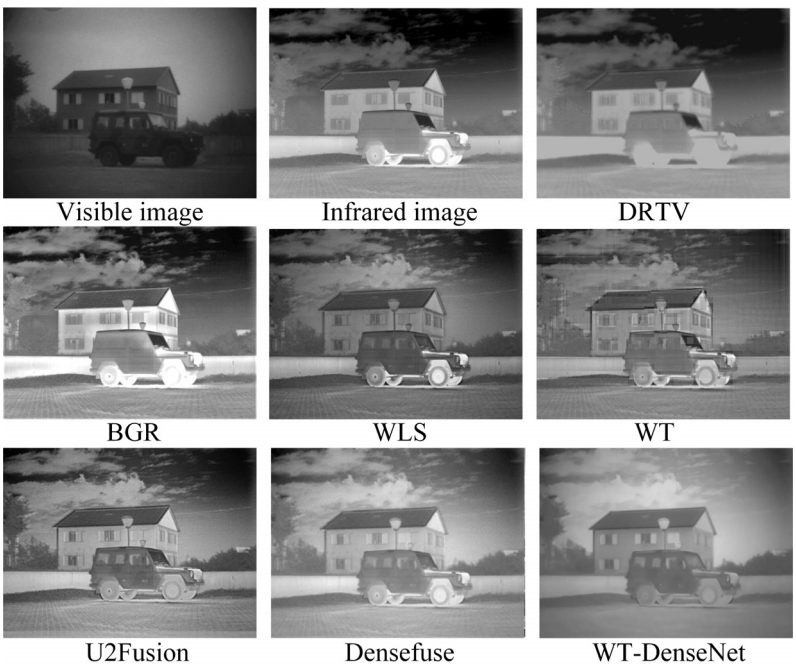

Fig. 7. Test images and fusion results of “car and house”.

the compared methods (DRTV, BGR, WLS, U2Fusion and DenseFuse). We use the default parameters of each method. As the proposed WT-DenseNet mainly is based on the wavelet fusion and DenseFuse, thus we compare the wavelet fusion and DenseFuse as the ablation experiments in comparative experiments. The fusion results are shown in Figs. 5–7. In Fig. 5, the fusion result of WT-DenseNet has the clearest and most ground textures, indicating that WT-DenseNet has best detail extraction capabilities. On the contrary, DRTV and BGR cannot restore the texture of the ground textures, and the details are not satisfactory. In addition, DRTV, BGR, WLS, WT, and U2Fusion contain much redundant information, especially BGR and WT.

However, WT-DenseNet can reduce the redundant information, and the fusion result clarity of WT-DenseNet is the best. Image details of the fusion results of DenseFuse are similar and in visual effects are not as good as WT-DenseNet. Although WT-DenseNet performs well in terms of details and clarity, its thermal information recognition ability is poor compared to WLS. WT-DenseNet can easily extract the highlight area of the visible image as the heat source information and take it into the fusion image together, which makes the thermal information (*e.g.*, person and door) from the infrared image blurred. As shown in Fig. 5, the sky appears as a highlight on the visible image, and the brightness intensity is also highlighted in the fusion result of WT-DenseNet. The same result is also found in Figs. 6 and 7.

4. Experimental results and analysis

4.1. The experimental details

The experimental infrared and visible source images are collected from the imaging system in Fig. 1. During the experiments, 300 pairs of source images are adopted from different scenes including indoor and outdoor. The resolutions of the visible and infrared images are 1920×1080 pixels and 320×240 pixels, respectively. The visible and infrared images are registered by the method in Section 2.2. The parameter configurations and training process of the WT-DenseNet are the same as in Section 3. The project of WT-DenseNet is implemented in a workstation (Intel Core i910900X, 32GB RAM, GTX2080Ti) with PyTorch.

4.2. Fusion method evaluation

Figures 8 and 9 show the fusion results of outdoor (strong light) and indoor scenarios (low light), respectively. In Fig. 8, DRTV, WLS, WT and U2Fusion do not improve the fusion image quality because they all look blurry. The fusion results of BGR and DenseFuse are softer, but the edge information is not obvious enough. Note that the WT fusion method has some limitations, such as generating the blockiness in the fused image, producing artifacts on diagonal edges, and outputting low-quality image.

In Fig. 9, DRTV has the obvious contour information, but the edge line looks a little abrupt. DRTV, BGR and WLS can effectively retain the thermal information. U2Fusion has difficulty seeing details in low light but the insufficient visible information also

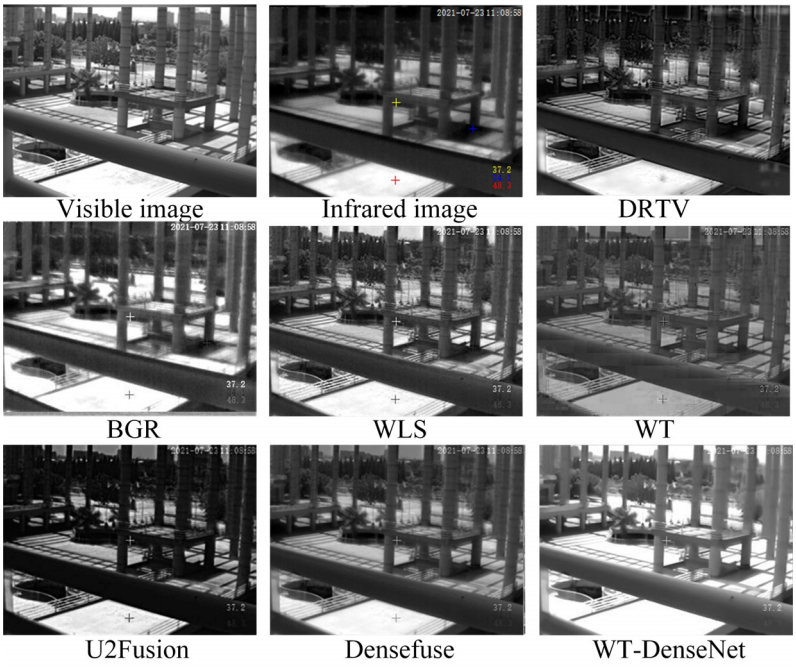

Fig. 8. Test images and fusion results of “outdoor”.

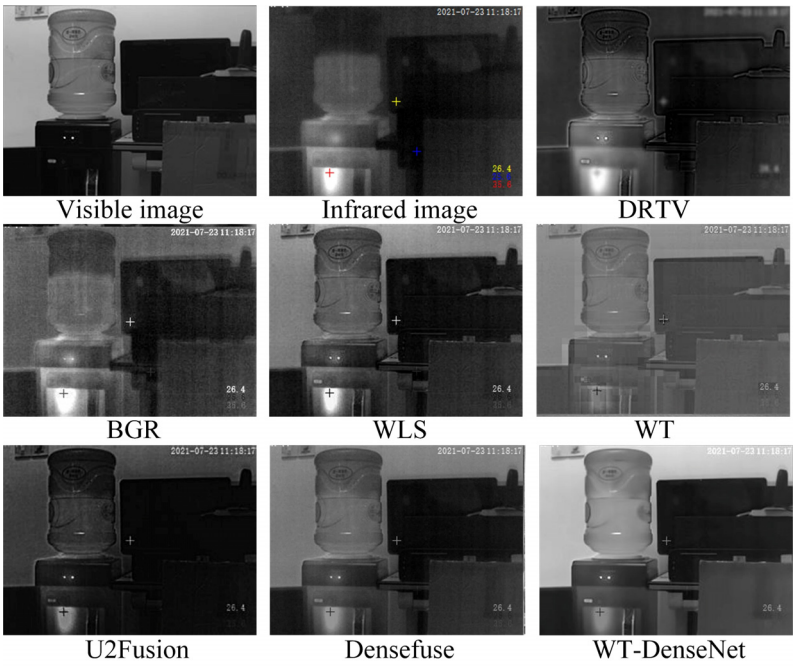

Fig. 9. Test images and fusion results of “indoor”.

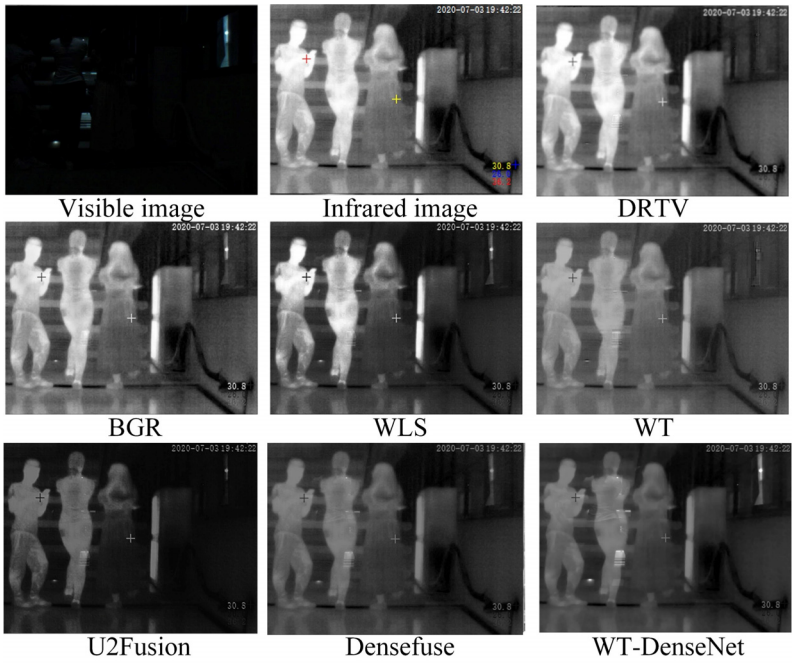

Fig. 10. Test images and fusion results of "people" in night environment.

reduces the image quality. In Figs. 8 and 9, the results of WT-DenseNet are the softest, and have the sharpest edges and the least redundant information. Moreover, WT-DenseNet is slightly inferior to DRTV, BGR and WLS in terms of the thermal information extraction. Compared with DenseFuse, WT-DenseNet has higher contrast and clearer details. In short, WT-DenseNet performs well at different scenarios in Figs. 8 and 9, indicating that WT-DenseNet performs better in contrast and detail.

Figure 10 shows the fusion results of the night environment, where the lights are turned off. Here, the visible image lacks effective imaging information. In Fig. 10, DRTV, BGR and WLS can well keep the infrared information, but they lose most of the visible image information. WT loses most of the visible image information and some infrared image information, while U2Fusion performs even worse than WT. DenseFuse and WT-DenseNet preserve most of the infrared and visible image information. Nonetheless, the fused image details and contrast of WT-DenseNet are superior to DenseFuse.

Besides, another group of experimental results is shown in Fig. 11. For better comparison, a small feature area is cropped from the source images and put in the lower right corner of the corresponding source images. The feature area image is surrounded by a red rectangle border. In Figs. 11 (C3 and D3), people's facial features are not clear enough. Figures 11 (BGR) are slightly better than Figs. 11 (DRTV and U2Fusion). Figure 11 (WT) has more artificial redundant information and the infrared thermal information is lost. From the five images in Fig. 11 (WT and U2Fusion), the low contrast

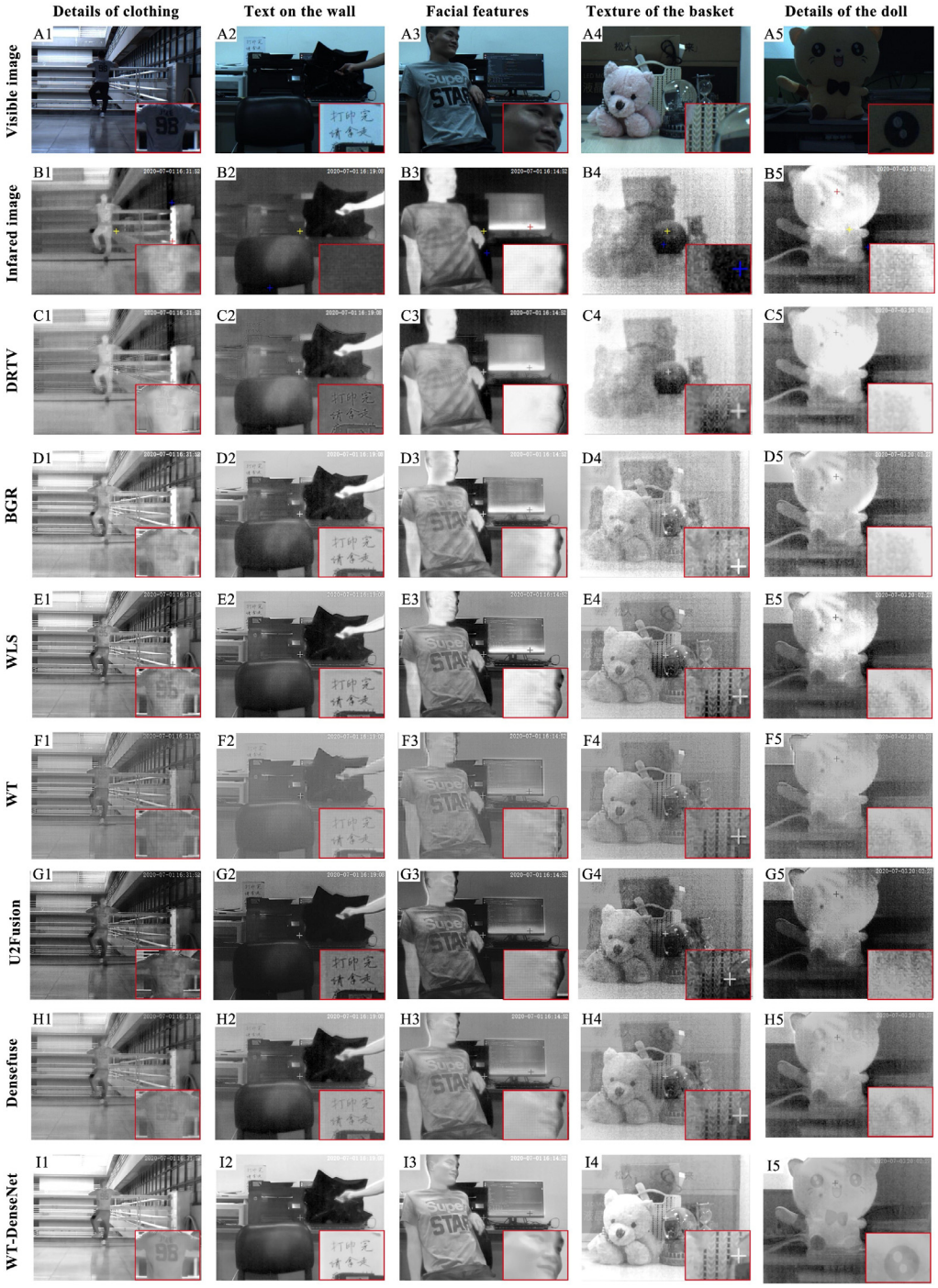

Fig. 11. Experimental results obtained by DRTV, BGR, WLS, WT, U2Fusion, DenseFuse and WT-DenseNet.

makes the image look worse. Figures 11 (WLS and DenseFuse) are visually good but they are not sufficient to provide all the details of the scene and contain a lot of redundant information.

In Fig. 11 (A1), the number on the clothes is 98, but Figs. 11 (C1-H1) can only get a number of 96. On the contrary, Fig. 11 (I1) can accurately observe the number 98, which means that WT-DenseNet can accurately identify tiny details. Since the target intensity of the infrared image is determined by temperature and radiated heat [26], Fig. 11 (B5) can clearly observe the palm print (thermal information) on the doll. However, the visible image Fig. 11 (B5) cannot reflect the thermal information. The thermal information of Fig. 11 (E5) is the most obvious but lacks the detail information, such as the eyes of a doll. In Figs. 11 (H5 and I5), the details and the highlighted thermal information can be observed. Compared with the DenseFuse, WT-DenseNet has richer image details and texture information.

As the results shown in Fig. 11, WT-DenseNet preserves more detailed information, such as details of clothing, text on the wall, facial features, texture of the basket, and details of the doll. Figures 11 (DRTV and BGR) cannot get enough details of visible images. For instance, as Figs. 11 (C1 and D1) shown, the numbers cannot be observed. Objectively, the fusion results of WT-DenseNet in Fig. 11 are visually clearer, smoother, and more natural than the other six methods. The clarity of WT-DenseNet is considerably improved. The reason is that WT-DenseNet can keep the low-frequency information which is used to optimize the fused image, making the fusion image more in line with human visual observation. The existing deep-learning-based image fusion

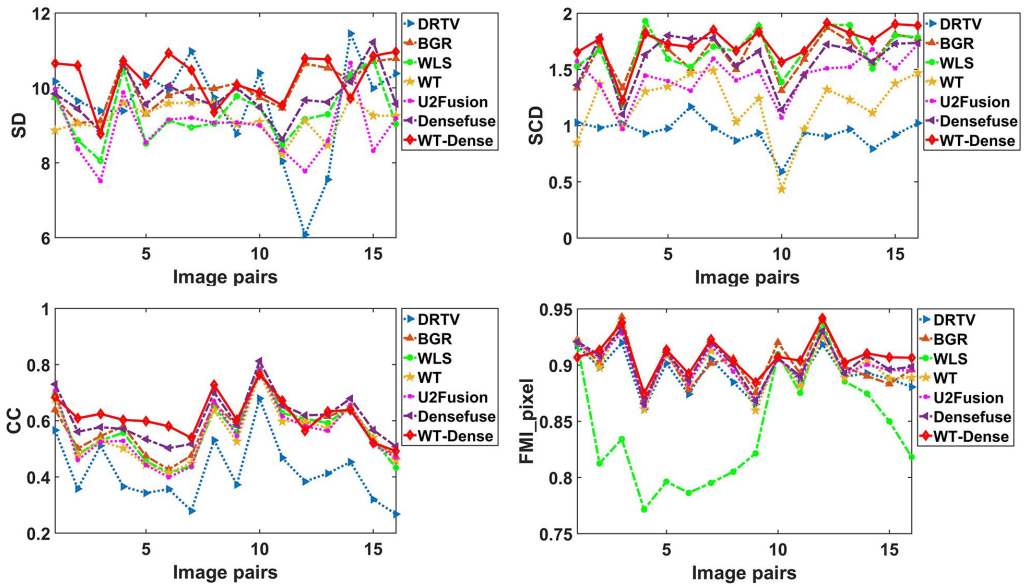

Fig. 12. Quantitative comparisons of seven fusion methods on sixteen infrared and visible image pairs. WT-Dense: WT-DenseNet.

T a b l e. The average of quality metrics for 50 fused images. WT-Dense: WT-DenseNet.

	FMIpixel	SD	CC	SCD
DRTV	0.92556	9.64703	0.42592	0.90979
BGR	0.91236	9.94664	0.59034	1.6523
WLS	0.93200	9.34177	0.57271	1.65863
WT	0.92856	9.00591	0.56137	1.15669
U2Fusion	0.92976	8.90272	0.56499	1.40703
DenseFuse	0.93659	9.48152	0.61681	1.58489
WT-Dense	**0.94278**	**10.3814**	**0.62010**	**1.74526**

methods typically extract the salient features of infrared and visible images, and often ignore the acquisition of low-frequency features. However, WT-DenseNet adds low -frequency features without reducing salient features, which is beneficial to reduce the redundant information of the fused image and improve the image quality.

Four quality metrics are utilized to objectively evaluate the performance of WT-DenseNet and other six methods, including the concept reflecting distribution and contrast (SD) [23], sum of the correlations of differences (SCD) [24], pixel feature mutual informations (FMIpixel) [26], and correlation coefficient (CC) [27]. Here, a larger value of the metrics (SD, SCD, FMIpixel and CC) means a better performance. Figure 12 shows quantitative comparison results on sixteen infrared and visible image pairs. The Table shows the average quality metrics for sixteen fused images and the best values for FMIpixel, SD, SCD, and CC are marked in bold. As shown in Fig. 12, the four metric values of WT-DenseNet are not always the biggest. However, it can be seen from the Table that WT-DenseNet can produce the largest average values and exhibits the best performance for the four metrics. From Figs. 5–12 and the Table, WT-DenseNet are strongly correlated with the source image, and the performance of WT-DenseNet is superior to the other six methods in detail and smoothness.

5. Conclusion

We have experimentally presented an infrared and visible fusion method by WT-DenseNet, which offers high-quality fusion images. The implementation process of WT-DenseNet is introduced in detail. The WT-DenseNet is trained with grayscale images from MS-COCO dataset and tested using the experimental infrared and visible images. The image fusion performance of the proposed method is compared with six fusion methods, DRTV, BGR, WLS, WT, U2Fusion and DenseFuse. The results show that image quality of the proposed method is better than that of the six methods in terms of clarity and detail. The proposed method can be applied in many areas, such as military detection, medical diagnosis and remote sensing. However, the infrared image thermal information recognition ability of the proposed method needs to be enhanced. In future work, we will pay more attention to the recognition and retention of thermal information. Besides, the future work will also focus on the optimizations and practical applications (*e.g.*, automatic detection of targets) of the proposed method.

Funding
National Natural Science Foundation of China (62173098, U20A6003, U2001201, U1801263, U1701262), Guangdong Provincial Key Laboratory of Cyber-Physical System (2020B1212060069).

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request.

Disclosures
The authors declare no conflicts of interest.

References

[1] Feng Y.F., Lu H.Q., Bai J.B., Cao L., Yin H., *Fully convolutional network-based infrared and visible image fusion*, Multimedia Tools and Applications **79**, 2020: 15001–15014, DOI: 10.1007/s11042-019-08579-w.
[2] Liu Y., Dong L., Ji Y., Xu W., *Infrared and visible image fusion through details preservation*, Sensors **19**(20), 2019: 4556, DOI: 10.3390/s19204556.
[3] Ma J., Ma Y., Li C., *Infrared and visible image fusion methods and applications: A survey*, Information Fusion **45**, 2019: 153–178, DOI: 10.1016/j.inffus.2018.02.004.
[4] Ma J., Zhang H., Shao Z., Liang P., Xu H., *GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion*, IEEE Transactions on Instrumentation and Measurement **70**, 2020: 5005014, DOI: 10.1109/TIM.2020.3038013.
[5] Li H., Qi X.B., Xie W.Y., *Fast infrared and visible image fusion with structural decomposition*, Knowledge-Based Systems **204**, 2020: 106182, DOI: 10.1016/j.knosys.2020.106182.
[6] Zhang H., Ma J., *SDNet: A versatile squeeze-and-decomposition network for real-time image fusion*, International Journal of Computer Vision **129**, 2021: 2761–2785, DOI: 10.1007/s11263-021-01501-8.
[7] Zhang Q., Liu Y., Blum R.S., Han J.G., Tao D.C., *Sparse representation based multi-sensor image fusion for multi-focus and multi-modality images: A review*, Information Fusion **40**, 2018: 57–75, DOI: 10.1016/j.inffus.2017.05.006.
[8] Zhang H., Xu H., Tian X., Jiang J.J., Ma J.Y., *Image fusion meets deep learning: A survey and perspective*, Information Fusion **76**, 2021: 323–336, DOI: 10.1016/j.inffus.2021.06.008.
[9] Zhu P., Ding L., Ma X.Q., Huang Z.H., *Fusion of infrared polarization and intensity images based on improved toggle operator*, Optics & Laser Technology **98**, 2018: 139–151, DOI: 10.1016/j.optlastec.2017.07.054.
[10] Ma J.Y., Xu H., Jiang J.J., Mei X.G., Zhang X.P., *DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion*, IEEE Transactions on Image Processing **29**, 2020: 4980–4995, DOI: 10.1109/TIP.2020.2977573.
[11] Liu Y., Chen X., Cheng J., Peng H., Wang Z.F., *Infrared and visible image fusion with convolutional neural networks*, International Journal of Wavelets, Multiresolution and Information Processing **16**(3), 2018: 1850018, DOI: 10.1142/S0219691318500182.
[12] Li H., Wu X.J., *DenseFuse: A fusion approach to infrared and visible images*, IEEE Transactions on Image Processing **28**(5), 2019: 2614–2623, DOI: 10.1109/TIP.2018.2887342.
[13] Garcia F., Mirbach B., Ottersten B., Grandidier F., Cuesta A., *Pixel weighted average strategy for depth sensor data fusion*, 2010 IEEE International Conference on Image Processing, 2010: 2805–2808, DOI: 10.1109/ICIP.2010.5651112.
[14] Zitova B., Flusser J., *Image registration methods: A survey*, Image and Vision Computing **21**(11), 2003: 977–1000, DOI: 10.1016/S0262-8856(03)00137-9.
[15] Cun X.D., Pun C.M., Gao H., *Applying stochastic second-order entropy images to multi-modal image registration*, Signal Processing: Image Communication **65**, 2018: 201–209, DOI: 10.1016/j.image.2018.03.021.

[16] LIN C.C., SHEU M.H., CHIANG H.K., LIAW C., WU Z.C., *The efficient VLSI design of BI-CUBIC convolution interpolation for digital image processing*,2008 IEEE International Symposium on Circuits and Systems (ISCAS), 2008: 480, DOI: 10.1109/ISCAS.2008.4541459.

[17] BUADES A., COLL B., MOREL J.M., *Non-local means denoising*, Image Processing On Line **1**, 2011: 208–212, DOI: 10.5201/ipol.2011.bcm_nlm.

[18] HUANG G., LIU Z., VAN DER MAATEN L., WEINBERGER K.Q., *Densely connected convolutional networks*, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 2261–2269, DOI: 10.1109/CVPR.2017.243.

[19] MALLAT G.S., *A theory for multiresolution signal decomposition: the wavelet representation*, IEEE Transactions on Pattern Analysis and Machine Intelligence **11**(7), 1989: 674–693, DOI: 10.1109/34.192463.

[20] LIN T.Y., MAIRE M., BELONGIE S., HAYS J., PERONA P., RAMANAN D., DOLLAR P., ZITNICK C.L., *Microsoft COCO: common objects in context*, [In] *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars [Eds.], ECCV 2014, Lecture Notes in Computer Science, Vol. 8693, Springer, Cham, 2014: 740–755, DOI: 10.1007/978-3-319-10602-1_48.

[21] TOET A., TNO Image Fusion Dataset, 2014.

[22] DU Q., XU H., MA Y., HUANG J., FAN F., *Fusing infrared and visible images of different resolutions via total variation model*, Sensors **18**(11), 2018: 3827, DOI: 10.3390/s18113827.

[23] ZHANG Y., ZHANG L.J., BAI X.Z., ZHANG L., *Infrared and visual image fusion through infrared feature extraction and visual information preservation*, Infrared Physics & Technology **83**, 2017: 227–237, DOI: 10.1016/j.infrared.2017.05.007.

[24] MA J.L., ZHOU Z.Q., WANG B., ZONG H., *Infrared and visible image fusion based on visual saliency map and weighted least square optimization*, Infrared Physics & Technology **82**, 2017: 8–17, DOI: 10.1016/j.infrared.2017.02.005.

[25] XU H., MA J.Y., JIANG J.J., GUO X.J., LING H.B., *U2Fusion: A unified unsupervised image fusion network*, IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(1), 2022: 502–518, DOI: 10.1109/TPAMI.2020.3012548.

[26] HAGHIGHAT M., RAZIAN M.A., *Fast-FMI: Non-reference image fusion metric*, 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), Astana, Kazakhstan, Oct. 2014, DOI: 10.1109/ICAICT.2014.7036000.

[27] DING M., YAO Y., LI W., CAO Y., *Visual tracking using locality-constrained linear coding and saliency map for visible light and infrared image sequences*, Signal Processing: Image Communication **68**, 2018: 13–25, DOI: 10.1016/j.image.2018.06.019.

Received February 24, 2022