

一种基于 ResNet152 的红外与可见光图像融合算法

李恒^{1*}, 张黎明^{2,3**}, 蒋美容², 李玉龙¹

¹兰州交通大学电子与信息工程学院, 甘肃 兰州 730070;

²兰州交通大学测绘与地理信息学院, 甘肃 兰州 730070;

³地理国情监测技术应用国家地方联合工程研究中心, 甘肃 兰州 730070

摘要 为进一步提高红外与可见光融合图像的细节信息, 并降低伪影和噪声, 提出一种基于 ResNet152 深度学习模型的红外与可见光图像融合算法。首先, 将红外与可见光图像分解成低频部分和高频部分; 然后, 运用平均权重策略对低频部分进行融合, 生成新的低频部分, 使用 ResNet152 网络对高频部分进行特征提取, 得到多个特征层, 对特征层分别进行 L1 正则化、卷积运算、双线性插值上采样和最大选择策略得到最大权重层, 由最大权重层和高频部分相乘得到新的高频部分; 最后, 由新的低频部分和高频部分对图像进行重建, 获得融合后的红外与可见光图像。实验结果表明, 该算法在保留图像显著特征的同时使融合图像获得了更多的细节纹理信息, 且有效地降低了伪影和噪声, 其主观评价和客观评价都优于对比算法。

关键词 图像处理; 图像融合; ResNet152; 红外图像; 可见光图像

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP57.081013

An Infrared and Visible Image Fusion Algorithm Based on ResNet152

Li Heng^{1*}, Zhang Liming^{2,3**}, Jiang Meirong², Li Yulong¹

¹School of Electronics and Information Engineering, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China;

²Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou, Gansu 730070, China;

³Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou, Gansu 730070, China

Abstract In order to improve the details of fusion image from infrared and visible image and reduce artifacts and noise, an infrared and visible image fusion algorithm based on ResNet152 deep learning model is proposed. Firstly, the source image is decomposed into the low frequency part and the high frequency part. The low frequency part is fused by the average weighting strategy to put a new low frequency part. The high frequency part is extracted by ResNet152 to obtain multiple feature layers. The L1 regularization, convolution operation, bilinear interpolation upsampling, and maximum selection strategy for the feature layers to obtain the maximum weight layer. Multiplying the maximum weight layer and the high frequency part to obtain a new high frequency part. Finally, the image is reconstructed by the new low frequency part and high frequency part. Experimental results show that the proposed algorithm can get more texture information while retaining the significant features of the image, and effectively reduces artifacts and noise. The subjective evaluation and objective evaluation are better than the comparison algorithm.

Key words image processing; image fusion; ResNet152; infrared image; visible image

OCIS codes 100.2980; 100.2000

1 引言

图像融合是一种增强技术, 目的是结合不同类型传感器捕捉到的信息生成内容更加丰富、细节更

加清晰的图像, 以便为后续处理和决策提供帮助。红外线和可见光的信号来自不同的模式, 红外图像捕获的是热辐射信息, 而可见光图像捕获的是光信息, 它们的组合信息比使用各自单独的模式得到的

收稿日期: 2019-07-01; 修回日期: 2019-07-31; 录用日期: 2019-09-16

基金项目: 国家自然科学基金(41761080)、兰州市人才创新创业科技计划资助(2016-RC-59)

* E-mail: 453984016@qq.com; ** E-mail: zhanglm8@gmail.com

信息更丰富^[1]。可见光图像具有高空间分辨率和细腻的纹理信息,但易被弱光照、雾和其他恶劣天气条件所干扰;红外图像能够抵抗这些干扰,但分辨率较低,纹理较差^[2]。因此,对红外与可见光图像进行融合可以实现信息互补^[3]。目前,红外与可见光图像融合在遥感探测、军事侦察、安全监控、医疗健康、工业生产等领域都有着重要的应用^[4]。有关红外与可见光图像融合方法主要有:基于多尺度变换的方法、基于稀疏表示的方法、基于神经网络的方法。

多尺度变换方法在红外与可见光图像融合中应用广泛。Huang 等^[5]提出了一种基于曲线变换和视觉注意机制的红外可见光图像融合方法,其模型可以提高融合图像的信噪比,突出模糊目标。Zhu 等^[6]提出了一种改进的多尺度高帽变换模型的红外可见光图像融合方法,该模型能够突出红外图像的目标,更好地保留可见光图像中的细节部分。基于稀疏表示的图像融合方法是从大量的高质量图像中学习一个过完备的字典,该方法关键在于过完备字典和稀疏编码的构造。Yin 等^[7]提出了一种将小波变换与字典学习相结合的多尺度字典学习方法,该方法充分利用了多尺度表示和字典学习的优点。Kim 等^[8]提出了一种基于图像块聚类 and 主成分分析(PCA)的字典学习方法,该模型不仅可以消除学习字典的冗余,同时还保证了红外和可见光图像融合的质量,但存在字典学习耗时过长的缺点。

近年来,深度学习在图像融合中得到了广泛的应用,多数基于深度学习的图像融合将深度模型作为特征提取的手段,利用提取到的深度特征和待融合图像得到融合后图像。Prabhakar 等^[9]提出一种基于卷积神经网络的多曝光图像融合方法,该网络结构是一种权重共享的孪生网络,源图像输入编码器后得到两个特征映射序列,之后使用加法策略对其进行融合得到融合特征图,最后融合特征图由解码器的三层卷积层进行重构。Liu 等^[10]提出了一种基于卷积神经网络的图像融合方法,使用输入图像的图像块及其模糊处理来训练网络,使网络具有分类能力,网络输出为分类得分表,得分表经过平均重叠块、二值化和两个一致性检验策略得到决定映射,最后将决定映射作为源图权重来重建图像。尽管该方法获得了更好的性能,但仍存在两个主要缺点:1)该方法目前只适用于多聚焦图像融合,使用范围有限;2)仅利用神经网络最后一层的计算结果,很多有用的中间层信息并未得到充分利用。

本文基于 ResNet152^[11]深度学习网络,设计了

一种简单高效的红外可见光图像融合方法。其基本过程是将红外与可见光图像通过均值滤波进行二尺度分解^[12],得到包含大尺度特征的低频部分和包含纹理特征的高频部分。低频部分使用平均权重策略获得新的低频部分;使用 ResNet152 网络对高频部分做多层特征的提取,每个特征层经过 L1 正则化、卷积运算和双线性插值上采样得到权重层,使用最大选择策略对多个权重层进行运算得到最大权重层,最大权重层与高频部分相乘得到新的高频部分,最后用新的低频部分和低频部分重建图像。

2 相关工作

神经网络的网络层数对图像特征的提取至关重要,然而简单的增加网络层数很容易导致梯度消失和精度饱和后极速下降问题^[11]。ResNet 提出的带有短连接的残差块成功地解决了这个问题^[11],并加快了训练速度。其中,残差块的结构如图 1 所示。

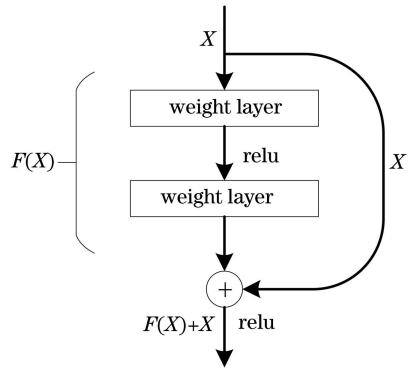


图 1 残差块结构

Fig. 1 Structure of residual block

图 1 中 X 表示残差块的输入, $F(X)$ 表示 X 经过两层权重后的映射结果,relu 为修正线性单元,残差块的输出为 $F(X)+X$ 。该结构可以有效地利用网络的多层信息,其中一个 152 层的 ResNet 网络,其深度是 VGG-19^[13] 的 8 倍,但其复杂度更低、提取特征的能力更强^[14]。因此,使用 ResNet152 作为图像融合特征提取的基础网络。

3 融合算法

本文将源图像记作 $I_k, k \in \{1, 2\}$ 。图 2 为本文算法的框架图,均值滤波可消除图像尖锐噪声,使图像变平滑。 I_k 通过均值滤波进行二尺度分解^[12],得到低频部分 I_k^b 和高频部分 I_k^d , 表示为

$$I_k^b = I_k * Z, \quad (1)$$

$$I_k^d = I_k - I_k^b, \quad (2)$$

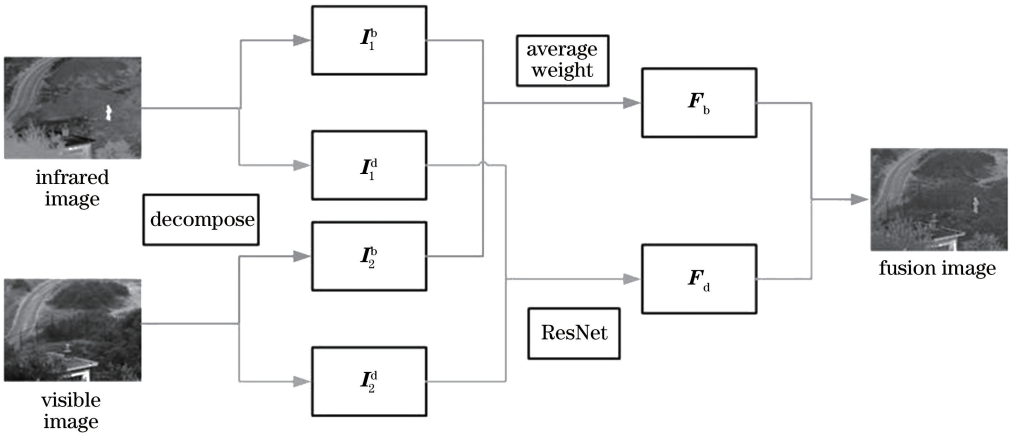


图2 本文方法的框架图

Fig. 2 Framework of proposed method

式中: Z 为均值滤波器,大小为 31×31 。低频部分使用平均权重的方法进行融合得到 F_b ,高频部分经过ResNet152处理后得到最大权重层,并与源图高频部分运算得到高频融合图像 F_d 。最后,由融合后的低频部分 F_b 和低频部分 F_d 重建得到融合图像。

3.1 低频部分融合

本文使用平均权重策略对低频部分进行融合,公式如下

$$F_b(x, y) = \lambda_1 I_1^b(x, y) + \lambda_2 I_2^b(x, y), \quad (3)$$

式中: $I_1^b(x, y)$ 和 $I_2^b(x, y)$ 为两种源图像低频部分 (x, y) 处的像素值; $F_b(x, y)$ 为融合后低频部分 (x, y) 处的像素值; λ_1 和 λ_2 为图像 I_1^b 和 I_2^b 像素的权重,为尽量保留低频部分中的大尺度特征,本文 λ_1 和 λ_2 的取值分别为0.6和0.4。

3.2 高频部分融合

对于高频部分 I_1^d 和 I_2^d ,先使用ResNet152提取特征,得到特征层。然后由特征层通过最大选择策略得到最大权重层。最后将最大权重层分别作为 I_1^d 和 I_2^d 权重系数得到高频部分融合图像 F_d 。具体流程如图3所示。

高频部分融合具体步骤如下。

1)提取特征。为均匀全面地获得ResNet152提取到的特征,分别选取ResNet152的conv1、conv2_1、conv3_1、conv4_1和conv5_1作为本文算法的特征提取层,定义 $H_i(\cdot)$ 为ResNet152网络的特征提取层,提取到的特征层为 $L_k^{i,m}$, $i \in \{1, 2, 3, 4, 5\}$ 。

$$L_k^{i,m} = H_i(I_k^d). \quad (4)$$

2)L1正则化。 $L_k^{i,m}$ 表示第 k 个高频部分经过 $H_i(\cdot)$ 提取特征得到的第 i 个特征层, m 表示第 i

个特征层的通道数, $m = 64 \times 2^{i-1}$,故 $L_k^{i,m}(x, y)$ 表示一个 m 维向量。 $L_k^{i,m}(x, y)$ 经L1正则化得到 T_k^i , T_k^i 表示第 k 个高频部分图像的第 i 个特征层经L1正则化后的结果。

$$T_k^i = \|L_k^{i,m}(x, y)\|_1. \quad (5)$$

3)卷积运算。为了使高频部分融合图像纹理更加自然,本文使用卷积核 A (步长为1)在 T_k^i 上进行卷积运算得到 \hat{T}_k^i ,流程如图4所示。之后由 \hat{T}_k^i 得到权重层 $W_k^i(x, y)$,其中 $n=2$ 。 $\hat{T}_k^i(x, y)$ 为第 k 个高频部分图像的第 i 个特征层经卷积运算后在 (x, y) 处的值, $W_k^i(x, y)$ 表示第 k 个高频部分图像的第 i 个权重层在 (x, y) 处的权重值。

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}, \quad (6)$$

$$W_k^i(x, y) = \frac{\hat{T}_k^i(x, y)}{\sum_{k=1}^n \hat{T}_k^i(x, y)}. \quad (7)$$

4)双线性插值上采样。 W_k^i 经ResNet152网络得到,记 W_k^i 的宽高为 (m_i, n_i) , I_k^d 的宽高为 (M, N) 。为使 (m_i, n_i) 与 (M, N) 相等,对 W_k^i 进行双线性插值上采样得到 \hat{W}_k^i , $i \in \{1, 2, 3, 4, 5\}$ 。

$$(m_i, n_i) = (M, N) \times \frac{1}{2^i}. \quad (8)$$

5)最大选择策略。 $\hat{W}_k^i(x, y)$ 表示第 k 个高频部分图像的第 i 个权重层经双线性插值上采样后在 (x, y) 处的值,故 $\hat{W}_k^{1:5}(x, y)$ 是一个5维向量。 $\hat{W}_k^{1:5}(x, y)$ 经最大选择策略得到 $\hat{W}_k(x, y)$, \hat{W}_k 即为第 k 个高频部分图像的最大权重层。最后,得到融合后的高频部分 F_d 。

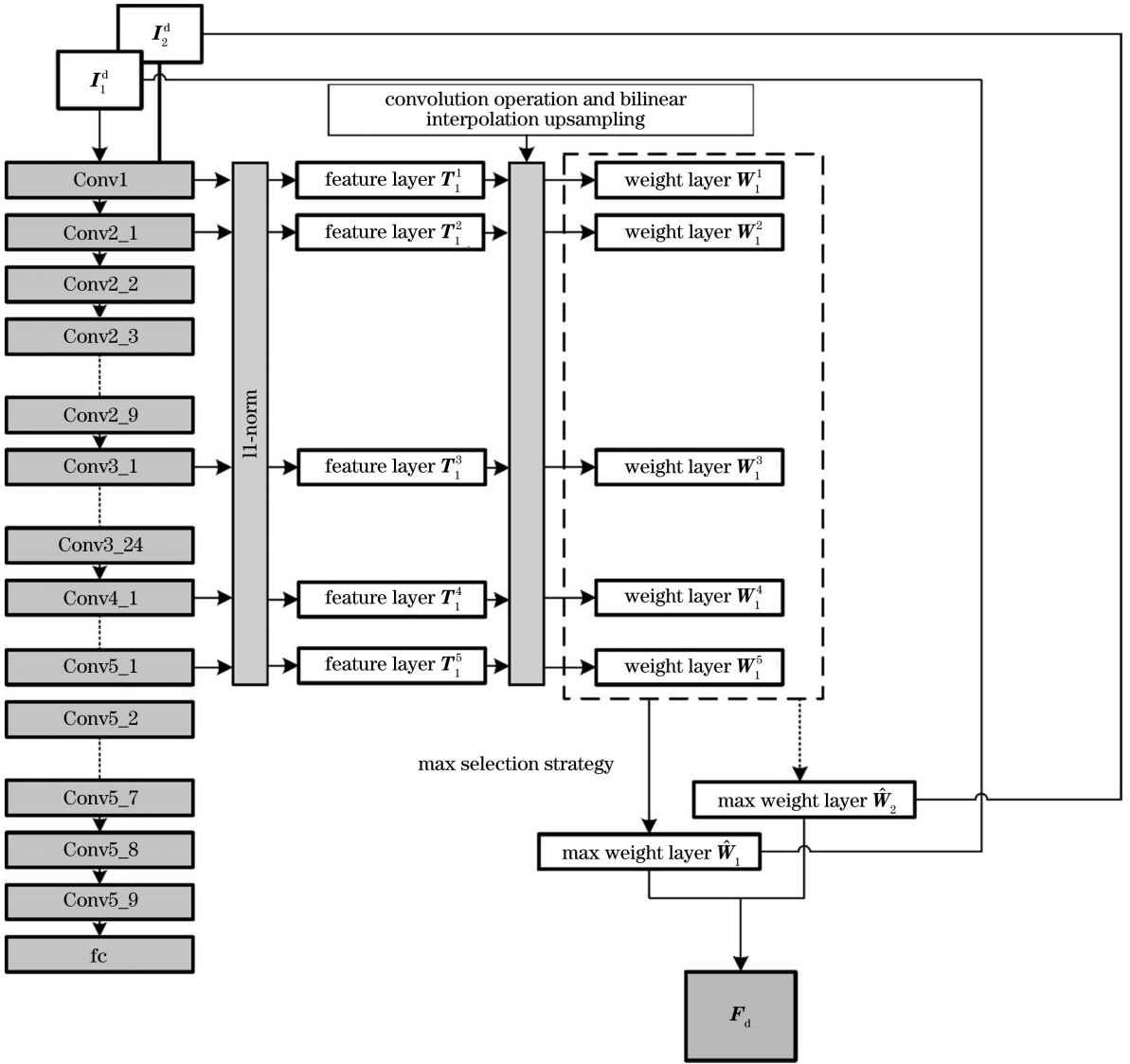


图 3 高频部分融合流程图

Fig. 3 Flowchart of high frequency part fusion

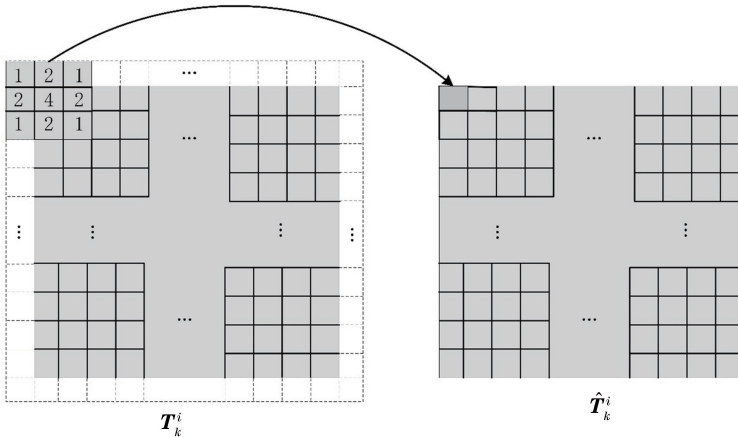


图 4 卷积运算图

Fig. 4 Convolution operation

$$\hat{W}_k(x, y) = \max[\hat{W}_k^{1:5}(x, y)], \quad (9)$$

$$F_d = \sum_{k=1}^n \hat{W}_k I_k^d. \quad (10)$$

3.3 重建图像

获得了融合后的低频部分 F_b 和高频部分 F_d 后,再对图像进行重建得到最后的融合图像,

$$F(x, y) = F_b(x, y) + F_d(x, y), \quad (11)$$

式中: $F_b(x, y)$ 表示融合后的低频部分在 (x, y) 处的值; $F_d(x, y)$ 表示融合后的高频部分在 (x, y) 处的值; $F(x, y)$ 表示最终融合图像在 (x, y) 处的像素值。

4 实验结果及分析

4.1 实验设定

本文实验中红外与可见光源图像来自公共数据

集 TNO,选择其中 20 组进行实验。本文选取 5 种常用的方法进行对比,分别是:交叉双边滤波融合法(CBF)^[15]、联合稀疏表示模型(JSR)^[16]、基于显著检测的联合稀疏表示模型(JSRSD)^[17]、加权最小二乘优化方法(WLS)^[18]、卷积稀疏表示模型(ConvSR)^[19],以上 5 种算法参数取值与各文献保持一致。本文实验环境为 Inter Core i5-4210U,内存 12 GB,操作系统为 Win 8.1,编程软件为 Matlab2017b。

4.2 主观评价

对上述 5 种方法和本文方法的实验结果进行对比分析,选择其中 5 组图像进行说明,编号分别为 a、b、c、d、e。如图 5 所示,5 组图像中算法 CBF 的实验结果噪声和晕影较多,显著特征不清晰。JSR 除

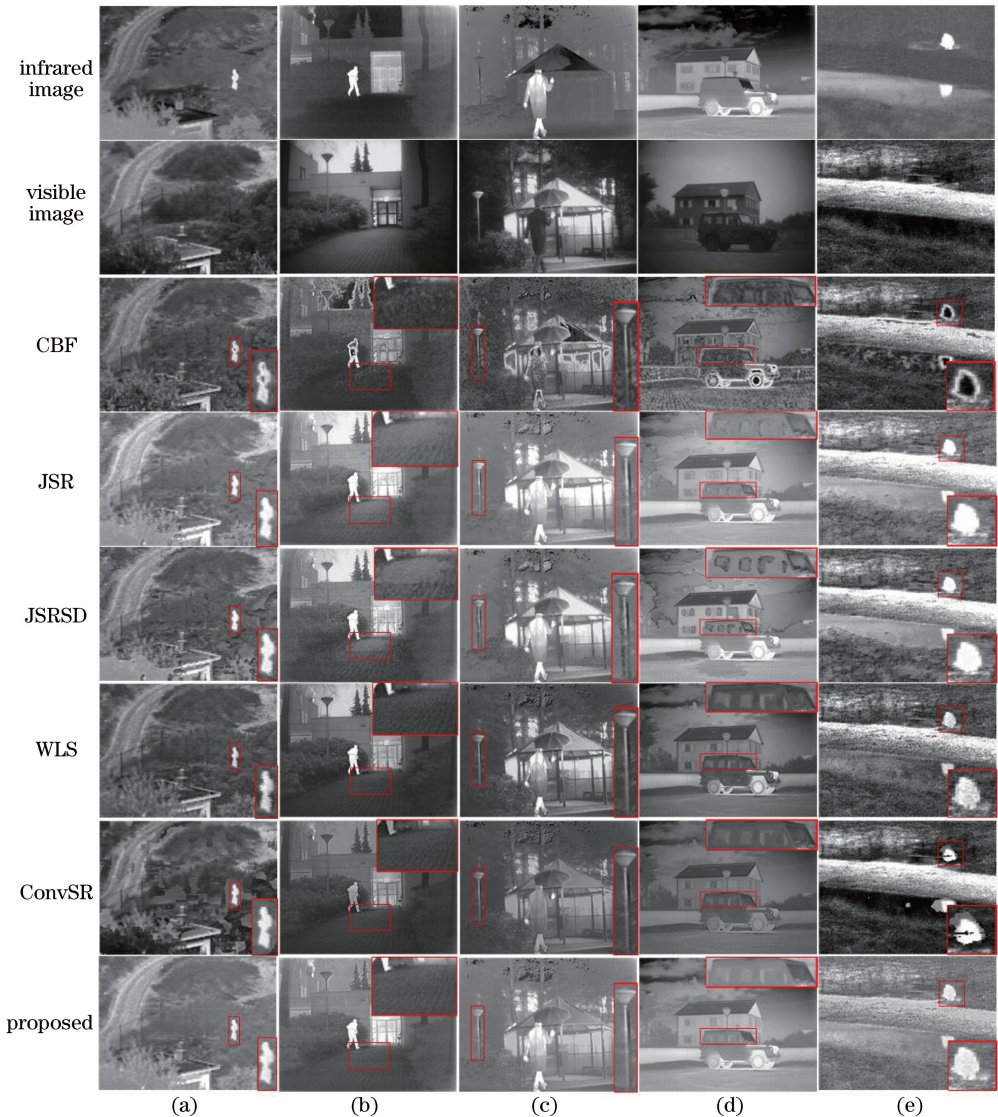


图 5 不同融合方法的比较

Fig. 5 Comparison of different fusion methods

显著特征外其他区域细节模糊,对比度不高。JSRSD和WLS融合后图像较多地保留了红外图像的特征,亮度过高,区域过度不自然。ConvSR在显著特征周围伪影较多且块效应明显,观感不佳。对比这5种方法,本文方法融合结果细节更加清晰,对比度更高,且晕影和块效应不明显,更适合人眼视觉的观察。

4.3 客观评价

为了定量的比较本文方法和对比方法,选用

表1 图像融合的客观评价

Table 1 Image fusion objective evaluation

Image	Metrics	CBF	WLS	JSR	JSRSD	ConvSR	Proposed
a	E_{FMIPixel}	0.8701	0.8847	0.8528	0.8339	0.8336	0.8856
	E_{FMIdct}	0.2450	0.2745	0.1538	0.1315	0.1422	0.3378
	E_{FMlw}	0.2984	0.3351	0.2113	0.1812	0.3205	0.3828
	N_{abf}	0.2317	0.1621	0.2332	0.3285	0.1224	0.0250
	E_{SSIM}	1.2476	1.4442	1.2146	1.0792	1.1968	1.5324
b	E_{FMIPixel}	0.8968	0.9305	0.9189	0.8932	0.9160	0.9345
	E_{FMIdct}	0.2229	0.2890	0.1686	0.1443	0.1388	0.3563
	E_{FMlw}	0.2920	0.3495	0.2326	0.2036	0.3813	0.3906
	N_{abf}	0.2554	0.2091	0.2430	0.3419	0.0572	0.0374
	E_{SSIM}	1.2994	1.4618	1.2562	1.1330	1.3608	1.5454
c	E_{FMIPixel}	0.8645	0.8918	0.8792	0.8632	0.8477	0.8927
	E_{FMIdct}	0.2299	0.2981	0.1699	0.1454	0.1342	0.3477
	E_{FMlw}	0.2633	0.3464	0.2205	0.1956	0.3424	0.3836
	N_{abf}	0.3607	0.2156	0.1769	0.2417	0.0868	0.0387
	E_{SSIM}	1.0740	1.3834	1.2560	1.1874	1.2084	1.4620
d	E_{FMIPixel}	0.8386	0.8977	0.8885	0.8653	0.8380	0.9128
	E_{FMIdct}	0.1835	0.3236	0.1591	0.1325	0.1065	0.3649
	E_{FMlw}	0.2407	0.3773	0.2266	0.2078	0.2992	0.4007
	N_{abf}	0.5278	0.2821	0.2109	0.2801	0.1145	0.0376
	E_{SSIM}	0.9150	1.3538	1.2506	1.1616	1.1500	1.4478
e	E_{FMIPixel}	0.8569	0.8471	0.8481	0.8434	0.8629	0.8634
	E_{FMIdct}	0.3781	0.3908	0.2360	0.2099	0.3061	0.4429
	E_{FMlw}	0.4650	0.4428	0.2840	0.2565	0.5144	0.4476
	N_{abf}	0.1523	0.2820	0.1873	0.2198	0.0348	0.0483
	E_{SSIM}	1.0472	1.0916	0.9764	0.9374	1.1292	1.1808

5种指标中实验结果最优的以粗体标注。从实验结果可以看出,本文方法在图a、b、c、d中5种指标均取得了较好的结果。在图e中只有ConvSR的 E_{FMlw} 、 N_{abf} 优于本文算法,其他指标均不如本文算法。这说明相较于其他5种对比方法,本文方法得到的红外与可见光融合图像在很大程度上保留了源图像的纹理细节特征,且降低了融合图像中的伪影和噪声,使得融合图像看起来更加清晰、自然,这与主观评价是一致的。

$E_{\text{FMIPixel}}^{[20]}$ 、 $E_{\text{FMIdct}}^{[20]}$ 、 $E_{\text{FMlw}}^{[20]}$ 、 $E_{\text{SSIM}}^{[3]}$ 、 $N_{\text{abf}}^{[21]}$ 5种指标进行评价, E_{FMIPixel} 、 E_{FMIdct} 和 E_{FMlw} 分别计算图像像素特征、离散余弦特征和小波特征的特征互信息,其值越大,表明源图与融合图像的信息相关度越高,融合过程中信息损失越少。 E_{SSIM} 表示源图与融合图像的相似程度。 N_{abf} 表示融合图像中产生的噪声或伪影,其值越小,表明融合图像包含的伪影和噪声越少。本文选取图5中的5组图像,定量评估本文方法和对比方法,结果如表1所示。

5 结论

本文使用ResNet152作为特征提取的基础网络来融合红外线和可见光图像。首先,将图像分解成包含大尺度特征的低频部分和包含纹理信息的高频部分。低频部分使用平均权重进行融合。使用ResNet152网络对高频部分做特征提取,得到最大权重层,之后高频部分和最大权重层相乘得到融合后的高频部分。最后由融合后的低频部分和高频部

分对图像进行重建。实验结果表明,本文方法不仅很好地保留了源图像的纹理特征,且很大程度上降低了融合图像的伪影和噪声,在主观评价和客观评价两个方面均取得了较好的融合效果,因此本文方法具有一定的实用价值与理论意义。

参 考 文 献

- [1] Liu S P, Fang Y. Infrared image fusion algorithm based on contourlet transform and improved pulse coupled neural network[J]. *Journal of Infrared and Millimeter Waves*, 2007, 26(3): 217-221.
刘盛鹏, 方勇. 基于 Contourlet 变换和 IPCNN 的融合算法及其在可见光与红外线图像融合中的应用[J]. *红外与毫米波学报*, 2007, 26(3): 217-221.
- [2] Wu Y Q, Wang Z L. Infrared and visible image fusion based on target extraction and guided filtering enhancement[J]. *Acta Optica Sinica*, 2017, 37(8): 0810001.
吴一全, 王志来. 基于目标提取与引导滤波增强的红外与可见光图像融合[J]. *光学学报*, 2017, 37(8): 0810001.
- [3] Ma J Y, Ma Y, Li C. Infrared and visible image fusion methods and applications: a survey [J]. *Information Fusion*, 2019, 45: 153-178.
- [4] Xu L, Cui G M, Zheng C P, et al. Fusion method of visible and infrared images based on multi-scale decomposition and saliency region extraction[J]. *Laser & Optoelectronics Progress*, 2017, 54(11): 111003.
许磊, 崔光芒, 郑晨浦, 等. 基于多尺度分解和显著性区域提取的可见光红外图像融合方法[J]. *激光与光电子学进展*, 2017, 54(11): 111003.
- [5] Huang Y D, Gao K, Gong C, et al. Infrared and visible image fusion with the target marked based on multi-resolution visual attention mechanisms [J]. *Proceedings of SPIE*, 2017, 10255: 102552V.
- [6] Zhu P, Ma X Q, Huang Z H. Fusion of infrared-visible images using improved multi-scale top-hat transform and suitable fusion rules [J]. *Infrared Physics & Technology*, 2017, 81: 282-295.
- [7] Yin H T. Sparse representation with learned multiscale dictionary for image fusion [J]. *Neurocomputing*, 2015, 148: 600-610.
- [8] Kim M, Han D K, Ko H. Joint patch clustering-based dictionary learning for multimodal image fusion [J]. *Information Fusion*, 2016, 27: 198-214.
- [9] Prabhakar K R, Sai Srikar V, Babu R V. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs [C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE, 2017: 4724-4732.
- [10] Liu Y, Chen X, Peng H, et al. Multi-focus image fusion with a deep convolutional neural network[J]. *Information Fusion*, 2017, 36: 191-207.
- [11] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [12] Li S T, Kang X D, Hu J W. Image fusion with guided filtering [J]. *IEEE Transactions on Image Processing*, 2013, 22(7): 2864-2875.
- [13] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J/OL]. (2015-04-10) [2019-05-20]. <https://arxiv.org/abs/1409.1556>.
- [14] Liu S P, Tian G H, Xu Y. A novel scene classification model combining ResNet based transfer learning and data augmentation with a filter [J]. *Neurocomputing*, 2019, 338: 191-206.
- [15] Kumar B K S. Image fusion based on pixel significance using cross bilateral filter[J]. *Signal, Image and Video Processing*, 2015, 9(5): 1193-1204.
- [16] Zhang Q H, Fu Y L, Li H F, et al. Dictionary learning method for joint sparse representation-based image fusion[J]. *Optical Engineering*, 2013, 52(5): 057006.
- [17] Liu C H, Qi Y, Ding W R. Infrared and visible image fusion method based on saliency detection in sparse domain[J]. *Infrared Physics & Technology*, 2017, 83: 94-102.
- [18] Ma J L, Zhou Z Q, Wang B, et al. Infrared and visible image fusion based on visual saliency map and weighted least square optimization [J]. *Infrared Physics & Technology*, 2017, 82: 8-17.
- [19] Liu Y, Chen X, Ward R K, et al. Image fusion with convolutional sparse representation[J]. *IEEE Signal Processing Letters*, 2016, 23(12): 1882-1886.
- [20] Haghghat M, Razian M A. Fast-FMI: non-reference image fusion metric [C] // 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT), October 15-17, 2014, Astana, Kazakhstan. New York: IEEE, 2014: 14916890.
- [21] Wang Z, Bovik A C. A universal image quality index[J]. *IEEE Signal Processing Letters*, 2002, 9(3): 81-84.