引用格式：LOU Xicheng, FENG Xin. Infrared and Visible Image Fusion in Latent Low Rank Representation Framework Based on Convolution Neural Network and Guided Filtering[J]. Acta Photonica Sinica, 2021, 50(3):0310004
娄熙承,冯鑫. 潜在低秩表示框架下基于卷积神经网络结合引导滤波的红外与可见光图像融合[J]. 光子学报,2021,50(3):0310004

潜在低秩表示框架下基于卷积神经网络结合引导滤波的红外与可见光图像融合

娄熙承,冯鑫
(重庆工商大学 机械工程学院 制造装备机构设计与控制重庆市重点实验室, 重庆 400067)

摘　要:为提高融合图像的可视性,解决传统红外与可见光图像融合算法中存在的边缘特征缺失、细节模糊的问题,提出了一种潜在低秩表示框架下基于卷积神经网络结合引导滤波的红外与可见光图像融合算法。该算法首先利用潜在低秩表示对源图像进行分解,得到源图像的低秩分量和显著分量。其次,利用卷积神经网络根据源图像的特征信息,得到权值图。再次,通过引导滤波算法对权值图进行边缘锐化,然后再将优化后的权值图分别与源图像的低秩分量和显著分量融合,得到融合图像的低秩分量和显著分量。最后,将融合图像的低秩分量和显著分量叠加,得到最终的融合图像。实验结果表明,该算法在主观评价和客观指标上均优于传统的红外与可见光图像融合算法。

关键词:红外与可见光图像;图像融合;潜在低秩表示;卷积神经网络;引导滤波
中图分类号:TP391　　**文献标识码**:A　　**doi**:10.3788/gzxb20215003.0310004

Infrared and Visible Image Fusion in Latent Low Rank Representation Framework Based on Convolution Neural Network and Guided Filtering

LOU Xicheng, FENG Xin
(*Key Laboratory of Manufacturing Equipment Mechanism Design and Control of Chongqing, College of Mechanical Engineering, Chongqing Technology and Business University, Chongqing* 400067, *China*)

Abstract: In order to improve the visibility of fused images and solve the problems of missing edge features and fuzzy details in traditional infrared and visible image fusion algorithms, an novel image fusion algorithm in latent low rank representation framework based on convolution neural network and guided filtering was proposed. Firstly, the source images are decomposed to low-rank parts and saliency parts by latend low-rank representation. Secondly, according to the pixel activity information of source images, the weight maps are obtained through the convolution neural network. Thirdly, the weight maps are improved by guided filtering according to the source images and through that the weight maps of low-rank parts and saliency parts can be obtained respectively. Then, the weight maps are fused with low-rank parts and saliency parts of original images to obtain the low-rank part and the saliency part of fused image. Finally, the final fused image can be obtained by adding the fused low-rank part and the fused significant part. Compared with other fusion algorithms, the experimental result shows that the proposed algorithm is superior to the traditional infrared and visible image fusion algorithms in terms of subjective visual effects

基金项目:国家自然科学基金(Nos. 31501229, 61861025),重庆市基础研究与前沿探索项目(No. cstc2018jcyjAX0483),重庆市教育委员会科学技术研究项目(Nos. KJQN201900821, KJQN202000803)
第一作者:娄熙承(1986—),男,硕士研究生,主要研究方向为图像融合。 Email: 2019611007@email.ctbu.edu.cn
导师(通讯作者):冯鑫(1982—),男,副教授,博士,主要研究方向为智能信息处理、图像融合。 Email: 149495263@qq.com
收稿日期:2020-11-22;**录用日期**:2020-12-23

and objective indexes.

Key words: Infrared and visible image; Image fusion; Latent low-rank representation decomposition; convolutional neural networks; Guided filtering

OCIS Codes: 100.2000; 100.2960; 100.2980; 100.3020

0 引言

可见光图像需要良好的照明条件才能获得有用的信息,在光线昏暗或是有浓雾的环境中,可见光图像所能反映的有用信息是极其有限的。而红外图像依靠其独特的热成像原理,即便是在夜间也能收集到能辐射热量的目标的信息。但由于成像原理的因素,红外图像很难有效表达目标的背景信息。因此,对可见光图像和红外图像的融合,可以充分利用两者之间的互补关系,使融合图像同时具备可见光图像良好的细节和背景信息以及红外图像良好的目标信息。这一图像融合技术正广泛应用于诸如航空侦查[1]、目标识别[2]和视频监控[3]等领域。

传统的红外与可见光图像融合算法主要是基于多尺度变换(Multi-Scale Transformation, MST),即先将源图像转换到变换域,然后在变换域中对分解系数按照预先设定的融合规则进行融合,最后再通过反变换得到最终的融合图像。该类型算法中比较典型的有离散小波变换(Discrete Wavelet Transformation, DWT)[4]、Contourlet变换[5]和Shearlet变换[6]。由于这些算法是利用预先定义的基函数对图像进行分析和融合,所以源图像的一些诸如边缘和纹理信息等重要特征往往不能很好的提取[7]。上述算法经过完善,出现了相应的改进版本,比如Tetrolet变换[8]、非下采样轮廓波变换(Non-Subsampled Contourlet Transform, NSCT)[9]和非下采样剪切波变换(Non-Subsampled Shearlet Transform, NSST)[10],但相同的问题始终没有得到根本的解决。YANG Bin等[11]将稀疏表示(Sparse Representation, SR)应用在图像处理上,将图像信号用预先学习的字典矩阵中部分列向量的线性组合来近似。由于字典矩阵是过完备的,该线性组合就可以表示成稀疏系数向量。按l_0范数最小的原则找到源图像的稀疏系数向量,再按l_1范数最大的原则确定融合图像的稀疏系数向量。该算法相对于MST不仅能改善保留源图像细节信息的能力,由于滑动窗口的使用,也增进了抗噪声的性能。但由于字典矩阵不可能完全涵盖源图像的细节信息,所以SR在融合图像的细节和边缘特征反映上仍然显得不足。为了处理这样的问题,有学者提出先用MST将源图像分解,再借助SR和其他方法分别处理分解后的图像低频和高频部分。比如SONAL G等[12]用NSST分解图像,LIU Feiqiang[13]等用拉普拉斯金字塔(Laplacian Pyramid, LP)分解图像,然后再对图像的低频和高频部分分别采用SR和取最大值的方法来处理。这类算法在改善SR无法很好提取图像显著特征的缺点上有明显的效果,但融合图像的细节展示仍然存在缺陷。LIU Guancan等[14]提出了潜在低秩表示(Latent Low-Rank Representation, LatLRR)算法,可以将图像分解成低秩分量、显著分量和噪音分量,LI Hui等[15]将LatLRR算法运用在图像融合上取得了良好的抗噪效果。该算法原理与SR类似,区别是采用图像信号本身作为字典矩阵。由于该算法可以把图像的噪音部分分离出来,所以该算法的抗噪性能良好,细节提取能力也较SR有明显的优势。但是该算法将图像的显著分量系数作为增加的隐藏项求解,这导致了融合图像显著分量的缺失。LI Shutao等[16]用引导滤波(Guided Filtering, GF)将源图像本身作为引导图像,能很好的保持图像的边缘信息(显著分量)。但正因为强调对图像显著分量的保持,该算法对噪音比较敏感。LIU Yu等[17]用卷积神经网络(Convolutional Neural Networks, CNN)根据源图像的特征信息得到权值图,再分别对源图像和权值图进行拉普拉斯金字塔分解和高斯金字塔分解,最后根据相应的判别值进行图像融合。CNN虽然具备良好的特征提取能力,但是金字塔算法的使用会造成图像细节的模糊。

综上所述,LatLRR算法在保留源图像细节信息上具有优势,CNN在提取图像特征信息上具有优势。本文提出的算法首先利用LatLRR将源图像进行分解;然后由CNN根据源图像的特征信息生成权值图;再由GF以源图像作为引导图像对权值图进行边缘锐化;最后用经GF完善的权值图与源图像的低秩和显著分量进行融合。将融合后的低秩和显著分量叠加,就得到了融合图像。这样就能将LatLRR和CNN各自的优点结合起来,兼顾融合图像边缘及细节信息的显示。

1 潜在低秩表示分解

LIU Guancan等[18]提出的低秩表示（Low-Rank Representation，LRR）在给定字典$\boldsymbol{A}$条件下，找到秩最小的矩阵$\boldsymbol{Z}$，将图像数据$\boldsymbol{X}$表示成字典中向量的线性组合，即

$$\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_* \quad \text{s.t. } \boldsymbol{X}=\boldsymbol{A}\boldsymbol{Z} \tag{1}$$

式中，$\|\cdot\|_*$表示核范数，也就是矩阵奇异值的和。若将图像数据$\boldsymbol{X}$本身作为字典，式(1)即可写成

$$\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_* \quad \text{s.t. } \boldsymbol{X}=\boldsymbol{X}\boldsymbol{Z} \tag{2}$$

但是用$\boldsymbol{X}$本身作为字典要有两个前提条件，一是$\boldsymbol{X}$的数据向量必须足够的完备；二是$\boldsymbol{X}$的噪声必须控制在很小的范围。在很多实际条件下，这样的要求是难以达到的。为了解决这一问题，文献[14]提出了在字典中增加隐藏项的方法，即

$$\min_{\boldsymbol{Z}} \|\boldsymbol{Z}\|_* \quad \text{s.t. } \boldsymbol{X}_{\mathrm{O}}=\left[\boldsymbol{X}_{\mathrm{O}},\boldsymbol{X}_{\mathrm{H}}\right]\boldsymbol{Z} \tag{3}$$

式中，$\boldsymbol{X}_{\mathrm{O}}$表示已知的图像数据，$\boldsymbol{X}_{\mathrm{H}}$表示未知的隐藏数据。由于字典中包含隐藏数据，所以这种改进算法就称为潜在低秩表示。考虑到噪声的影响，将式(3)改写成

$$\min_{\boldsymbol{Z},\boldsymbol{E}} \|\boldsymbol{Z}\|_*+\lambda\|\boldsymbol{E}\|_1 \quad \text{s.t. } \boldsymbol{X}_{\mathrm{O}}=\left[\boldsymbol{X}_{\mathrm{O}},\boldsymbol{X}_{\mathrm{H}}\right]\boldsymbol{Z}+\boldsymbol{E} \tag{4}$$

式中，$\lambda>0$表示平衡系数，$\|\cdot\|_1$表示l_1范数，$\boldsymbol{E}$表示噪声。按照文献[14]的步骤对上式化简，可到

$$\min_{\boldsymbol{Z},\boldsymbol{L},\boldsymbol{E}} \|\boldsymbol{Z}\|_*+\|\boldsymbol{L}\|_*+\lambda\|\boldsymbol{E}\|_1 \quad \text{s.t. } \boldsymbol{X}=\boldsymbol{X}\boldsymbol{Z}+\boldsymbol{L}\boldsymbol{X}+\boldsymbol{E} \tag{5}$$

式中，$\boldsymbol{L}$是显著系数，而矩阵$\boldsymbol{Z}$在这里就可以看作是低秩系数。用增广拉格朗日乘数法（Augmented Lagrangian Multiplier，ALM）对式(5)求解，解出系数$\boldsymbol{Z}$和$\boldsymbol{L}$，图像的低秩分量和显著分量就可以相应的表示成$\boldsymbol{X}\boldsymbol{Z}$和$\boldsymbol{L}\boldsymbol{X}$。

用LatLRR分解红外与可见光图像的结果如图1所示。其中，$\boldsymbol{I}_1$是源红外图像，$\boldsymbol{I}_1^{\mathrm{lrr}}$是$\boldsymbol{I}_1$经LatLRR分解后得到的低秩分量，由$\boldsymbol{I}_1$乘以式(5)中的低秩系数$Z$求得；$\boldsymbol{I}_1^{\mathrm{s}}$是$\boldsymbol{I}_1$经LatLRR分解后得到的显著分量，由式(5)中的显著系数$\boldsymbol{L}$乘以$\boldsymbol{I}_1$求得；$\boldsymbol{E}_1$是$\boldsymbol{I}_1$经LatLRR分解后得到噪声分量，也就是式(5)中的$\boldsymbol{E}$。$\boldsymbol{I}_2$是源可见光图像，同理，$\boldsymbol{I}_2^{\mathrm{lrr}}$、$\boldsymbol{I}_2^{\mathrm{s}}$和$\boldsymbol{E}_2$分别是$\boldsymbol{I}_2$经LatLRR分解后得到的低秩分量、显著分量和噪声分量。从图中可以看出，低秩分量$\boldsymbol{I}_1^{\mathrm{lrr}}$和$\boldsymbol{I}_2^{\mathrm{lrr}}$含有源图像主要的细节和光线信息；而显著分量$\boldsymbol{I}_1^{\mathrm{s}}$和$\boldsymbol{I}_2^{\mathrm{s}}$含有源图像主要的边缘信息和显著特征。另外，源图像在经LatLRR分解后噪声分量$\boldsymbol{E}_1$、$\boldsymbol{E}_2$被隔离出来，在往后的图像融合过程中，并没有将$\boldsymbol{E}_1$、$\boldsymbol{E}_2$包含进去，只叠加经过处理的低秩分量和显著分量，这也就是LatLRR能有良好去噪声能力的原因。

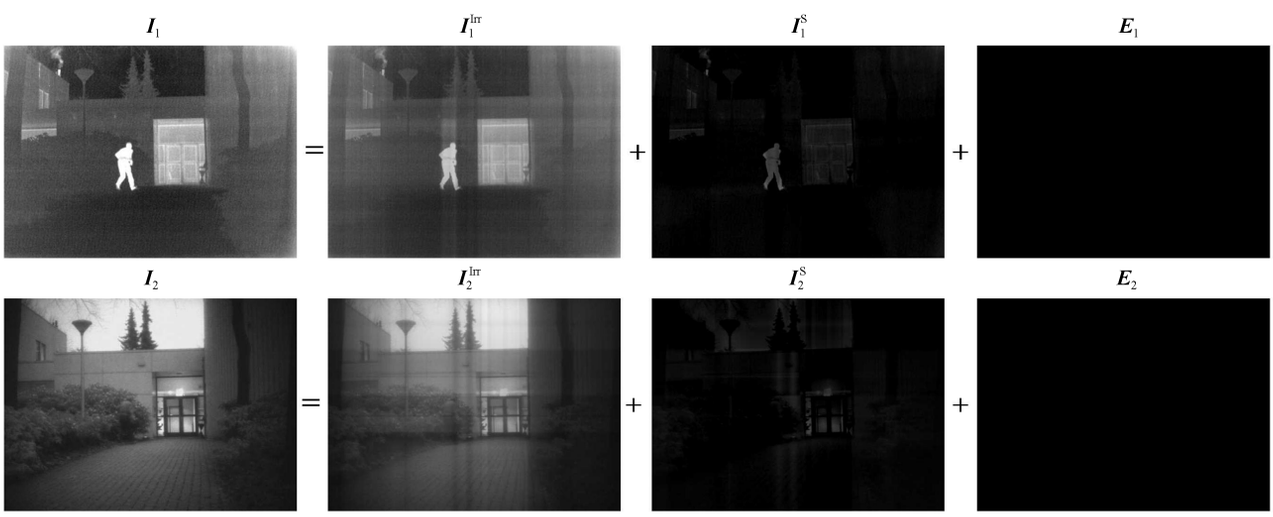


图1 LatLRR分解结果

Fig.1 LatLRR decomposition results

2 基于卷积神经网络的显著图学习

CNN是一种典型的深度学习模型，它具有表征学习能力，可以按层级结构对信号进行分类。CNN由输入层、隐藏层和输出层构成，其中隐藏层又由若干个卷积层、激活层和池化层组成。CNN的基本单元是神经元，结构如图2虚线方框部分所示。图中x_1、x_2、x_3表示输入量，w_1、w_2、w_3是权重，可以由反向求导(back-propagation)[19]得出。b是偏置量，y是输出量。输入层的数据在乘以各自的权重后，加上偏置值b送到卷积层。卷积层的数据在经过激活函数的处理后进入激活层，引入激活函数的目的是为了增加非线性因素，常用的激活函数是ReLU函数[20]。该函数功能是将负数值置为0，正数值保持不变，即$f(x)=\max(0,x)$。激活层的数据再经过池化处理后进入池化层，池化的目的是为了减少运算量，常用的池化Max-pool是用一个2×2的窗口选出目标中的最大值。池化层的数据再经过扁平化，也就是将池化层的数据压缩为一维后，由softmax函数处理转换成0至1之间的概率分布，它反映了神经网络判断的结果。关于CNN的更多信息参考文献[21]。

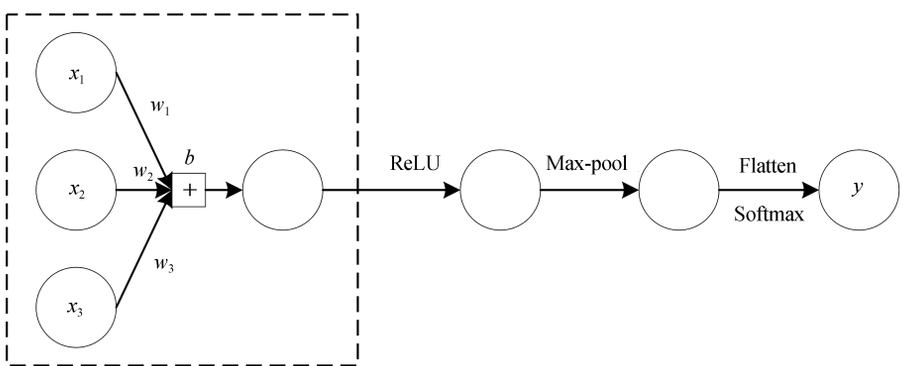


图2 神经元结构图

Fig.2 Neuron structure

本文用已经过训练的孪生神经网络(siamese network)[17]对源图像进行处理，其结构如图3所示，它由两条完全相同的CNN支路构成，包括3个卷积层和1个池化层。每条支路分别接收一个样本输入，然后将其映射至高维特征空间，并输出对应的表征。最后通过计算两个表征的距离，来比较两个样本的相似程度[22]。训练的样本由ILSVRC 2012图像集[23]中的高清图像生成。在将这些高清图像转换成灰度图像后，用高斯核为7×7，标准差为2的参数对高清图像先后进行5次滤波，这样可以得到5张模糊程度递增的模糊图像。训

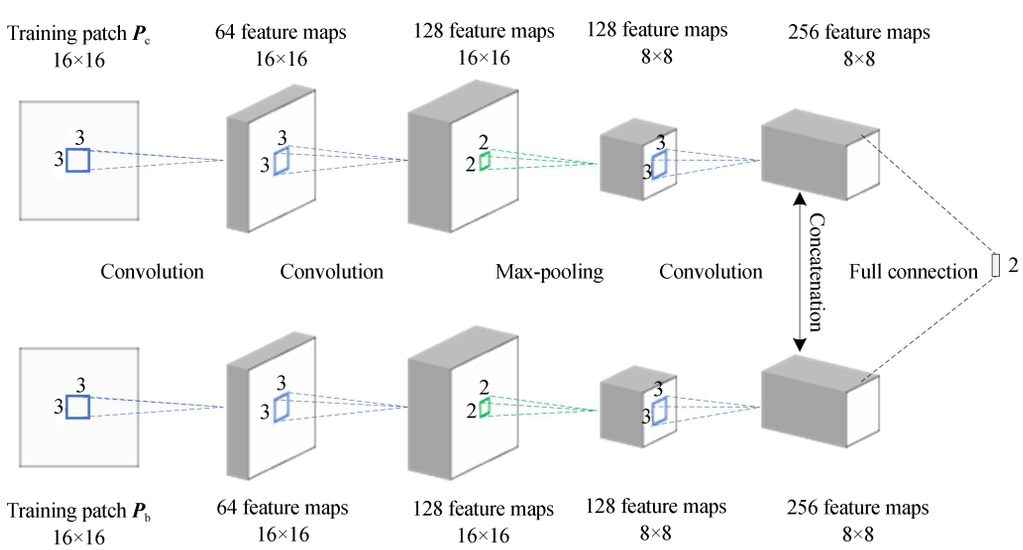


图3 孪生神经网络结构图

Fig.3 Siamese network structure

练中以高清图像分别和5张模糊程度不同的图像组成两两一组的实验组，然后分别在每个实验组对高清图像和模糊图像进行20次16×16的子块采样，这样每个实验组就有20对子块。以$\boldsymbol{P}_{\mathrm{c}}$和$\boldsymbol{P}_{\mathrm{b}}$分别代表高清和模糊的子块，以$\boldsymbol{P}_1$和$\boldsymbol{P}_2$分别代表孪生神经网络两条支路的输入，若$\boldsymbol{P}_{\mathrm{c}}=\boldsymbol{P}_1$，神经网络的标签值就为1，反之，标签值就为0。损失函数，即softmax函数（见式（7））的输出值与标签值的对数损失，由随机梯度下降法（Stochastic Gradient Descent，SGD）进行最小化处理。动量和权值衰减分别设置为0.9和0.000 5，权重值按式（6）更新

$$v_{i+1}=0.9\cdot v_i-0.0005\cdot\alpha\cdot w_i-\alpha\cdot\frac{\partial L}{\partial w_i} \tag{6}$$

式中，v是动量变量，i表示第i次重复，α表示学习速率，L表示损失函数，$w_{i+1}=w_i+v_{i+1}$。训练过程在深度学习框架Caffe[24]内完成，更加详细的神经网络训练过程请参考文献[17]。源图像在经神经网络的卷积处理后，得到2组各256张特征图。这些特征图通过1 024个8×8卷积核压缩成两个矩阵，再由softmax函数将矩阵内的各元素的值转换成0至1之间表示概率密度的数，softmax函数的表达式为

$$\mathrm{softmax}\left(z_i\right)=\frac{\mathrm{e}^{z_i}}{\sum_{c=1}^{C}\mathrm{e}^{z_c}} \tag{7}$$

式中，z_i为第i个节点的输出值，C为输出节点个数，也就是分类的类别数。如图4所示，神经网络的输出结果是两张权值图$\boldsymbol{S}_1$和$\boldsymbol{S}_2$。受神经网络池化和卷积的影响，这时的权值图尺寸不足源图像尺寸的二分之一。为便于后期融合处理，需要通过以子块图像叠加像素，再除以像素叠加数量求出像素均值的方法将权值图放大至与源图像一样的尺寸，得到$\boldsymbol{\omega}_1$和$\boldsymbol{\omega}_2$。

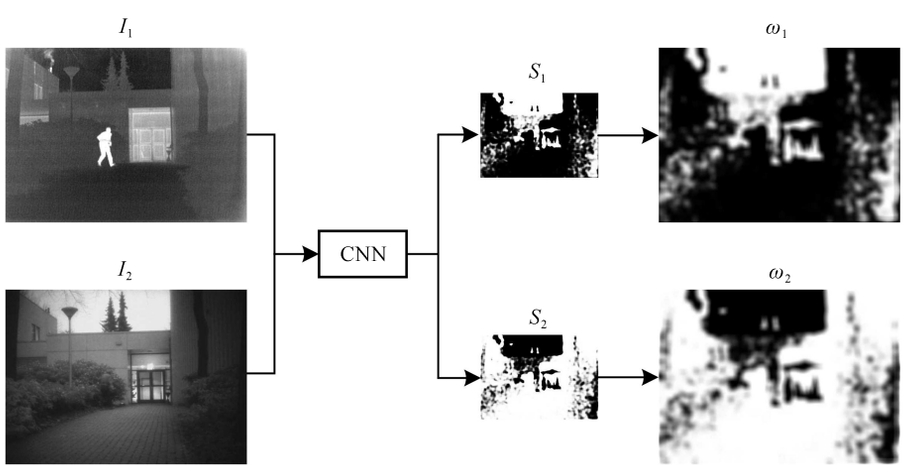


图4　CNN处理过程图

Fig.4　CNN processing diagram

3　基于引导滤波的规则设置

假设引导滤波器在引导图像$\boldsymbol{I}$和滤波输出图像$\boldsymbol{q}$之间的一个以像素k为中心的二维窗口$\boldsymbol{\omega}_k$内是一个局部线性模型，即

$$\boldsymbol{q}_i=a_k\boldsymbol{I}_i+b_k\quad\forall i\in\omega_k \tag{8}$$

式中，系数a_k和b_k在窗口$\boldsymbol{\omega}_k$内都是常数，建立一个最优化模型使窗口$\boldsymbol{\omega}_k$内输入和输出图像的差别$\boldsymbol{E}$最小

$$\boldsymbol{E}\left(a_k,b_k\right)=\sum_{i\in\omega_k}\left[\left(a_k\boldsymbol{I}_i+b_k-\boldsymbol{p}_i\right)^2+\epsilon a_k{}^2\right] \tag{9}$$

式中，ϵ是预先定义的正则化参数，$\boldsymbol{p}$是输入图像。通过线性回归的方法求解式（9）[25]得到线性模型系数分别为

$$a_k=\frac{\frac{1}{|\omega|}\sum_{i\in\omega_k}\boldsymbol{I}_i\boldsymbol{p}_i-\mu_k\bar{p}_k}{{\sigma_k}^2+\epsilon} \tag{10}$$

以及

$$b_k=\bar{p}_k-a_k\mu_k \tag{11}$$

式中，μ_k和σ_k分别为引导图像$\boldsymbol{I}$在窗口$\boldsymbol{\omega}_k$内的均值与方差，$|\omega|$是窗口$\boldsymbol{\omega}_k$内的像素数目，$\bar{p}_k$是输入图像$\boldsymbol{I}$在窗口$\boldsymbol{\omega}_k$内的均值。如果将输入图像作为引导图像，即$\boldsymbol{I}=\boldsymbol{p}$，式(10)和(11)可分别改写成

$$a_k=\frac{{\sigma_k}^2}{{\sigma_k}^2+\epsilon} \tag{12}$$

以及

$$b_k=\left(1-a_k\right)\mu_k \tag{13}$$

因为输入图像的每个像素会被多个窗口$\boldsymbol{\omega}_k$包含，而每个窗口$\boldsymbol{\omega}_k$又有单独的线性系数a_k和b_k，所以在计算输出图像每个像素数值的时候应该采用线性系数的平均值，即

$$\boldsymbol{q}_i=\frac{1}{|\omega|}\sum_{k\in\boldsymbol{\omega}_i}a_k\boldsymbol{I}_i+\frac{1}{|\omega|}\sum_{k\in\boldsymbol{\omega}_i}b_k \tag{14}$$

式中，$\boldsymbol{\omega}_i$也是一个二维窗口，表示包含像素i的所有$\boldsymbol{\omega}_k$窗口。将式(12)、(13)和$\boldsymbol{I}=\boldsymbol{p}$带入式(14)可以得到

$$\boldsymbol{q}_i=\frac{1}{|\omega|}\sum_{k\in\boldsymbol{\omega}_i}\frac{{\sigma_k}^2}{{\sigma_k}^2+\epsilon}\boldsymbol{p}_i+\frac{1}{|\omega|}\sum_{k\in\boldsymbol{\omega}_i}\left(1-a_k\right)\mu_k \tag{15}$$

在输入图像${\sigma_k}^2\gg0$的区域，也就是边界区域，$a_k\approx1$，式(15)可写成

$$\boldsymbol{q}_i=\boldsymbol{p}_i \tag{16}$$

这说明输出图像基本上保留了输入图像的边缘信息。而在输入图像${\sigma_k}^2\approx0$的区域，也就是平滑区域，式(15)可写成

$$\boldsymbol{q}_i=\frac{1}{|\omega|}\sum_{k\in\boldsymbol{\omega}_i}\boldsymbol{p}_k \tag{17}$$

这相当于是对平滑区域进行了均值滤波，以达到去除噪声的目的。

如图5所示，将CNN输出的权值图作二值化处理，方法就是将ω_1和ω_2按相同位置的像素做比较，像素较大的点置1，较小的点置0，得到的二值图像分别为$\boldsymbol{P}_1$和$\boldsymbol{P}_2$。即

$$\boldsymbol{P}_n^k=\begin{cases}1 & \text{if } \boldsymbol{\omega}_n^k=\max\left(\boldsymbol{\omega}_1^k,\boldsymbol{\omega}_2^k\right)\\ 0 & \text{otherwise}\end{cases}\quad n=1,2 \tag{18}$$

式中，$\boldsymbol{P}_n^k$表示二值图像中第n张图的第k个像素的值，$\boldsymbol{\omega}_n^k$表示权值图中第n张图的第k个像素的值。以源图像$\boldsymbol{I}_1$和$\boldsymbol{I}_2$作为引导图像，对$\boldsymbol{P}_1$和$\boldsymbol{P}_2$进行引导滤波

$$\begin{cases}\boldsymbol{W}_1^{\mathrm{lrr}}=G_{r_1,\epsilon_1}\left(\boldsymbol{P}_1,\boldsymbol{I}_1\right)\\ \boldsymbol{W}_2^{\mathrm{lrr}}=G_{r_1,\epsilon_1}\left(\boldsymbol{P}_2,\boldsymbol{I}_2\right)\end{cases} \tag{19}$$

$$\begin{cases}\boldsymbol{W}_1^{\mathrm{s}}=G_{r_2,\epsilon_2}\left(\boldsymbol{P}_1,\boldsymbol{I}_1\right)\\ \boldsymbol{W}_2^{\mathrm{s}}=G_{r_2,\epsilon_2}\left(\boldsymbol{P}_2,\boldsymbol{I}_2\right)\end{cases} \tag{20}$$

式中，$\boldsymbol{W}_1^{\mathrm{lrr}}$表示红外图像的平滑权值图，$\boldsymbol{W}_2^{\mathrm{lrr}}$表示可见光图像的平滑权值图，$\boldsymbol{W}_1^{\mathrm{s}}$表示红外图像的锐化权值图，$\boldsymbol{W}_2^{\mathrm{s}}$表示可见光图像的锐化权值图，$G$表示引导滤波函数，$r_1,\epsilon_1$和$r_2,\epsilon_2$是引导滤波的参数。考虑到源图像的低秩分量含有图像的平滑部分，而显著分量含有图像的边缘及显著特征部分，所以源图像的低秩分量和显著分量的权值图就需要根据它们各自的特点进行完善。用图像滤波的方式对权值图做进一步处理，r表示过滤器的尺寸大小，ε表示模糊程度。与低秩分量结合权值图应该是比较平滑的，如果过于锐化，就会造成融合图像出现形状扭曲现象；而与显著分量结合权值图应该是比较锐化的，如果过于平滑，会导致融合图像的边缘或特征不明显。参数r_1,ϵ_1和r_2,ϵ_2就是按照上述原则确定，本文采用文献[15]中的参数值，r_1,ϵ_1

和r_2,ε_2分别取45,0.3和7,10^{-6}。

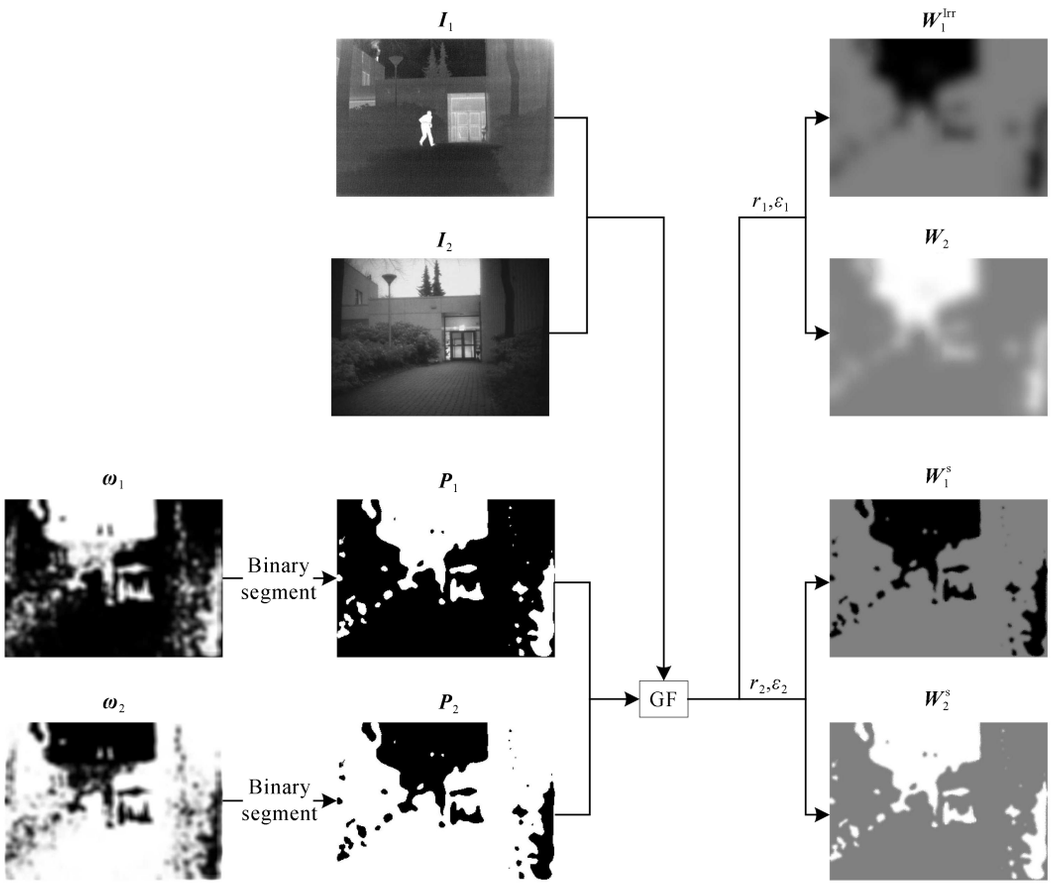


图5 GF处理过程图

Fig.5 GF processing diagram

4 融合分量组合

将引导滤波输出的平滑权值$\boldsymbol{W}_1^{\mathrm{lrr}}$、$\boldsymbol{W}_2^{\mathrm{lrr}}$和锐化权值$\boldsymbol{W}_1^{s}$、$\boldsymbol{W}_2^{s}$分别与源图像的低秩分量和显著分量分别做哈达玛积(Hadamard product),也就是两个矩阵内元素对应相乘。得到融合图像的低秩分量$\boldsymbol{I}_{\mathrm{lrr}}$为

$$\boldsymbol{I}_{\mathrm{lrr}}=\boldsymbol{W}_1^{\mathrm{lrr}}\circ\boldsymbol{I}_1^{\mathrm{lrr}}+\boldsymbol{W}_2^{\mathrm{lrr}}\circ\boldsymbol{I}_2^{\mathrm{lrr}} \tag{21}$$

式中,$\boldsymbol{I}_1^{\mathrm{lrr}}$和$\boldsymbol{I}_2^{\mathrm{lrr}}$分别表示红外图像的低秩分量和可见光图像的低秩分量。融合图像的显著分量$\boldsymbol{I}_{\mathrm{s}}$为

$$\boldsymbol{I}_{\mathrm{s}}=\boldsymbol{W}_1^{s}\circ\boldsymbol{I}_1^{s}+\boldsymbol{W}_2^{s}\circ\boldsymbol{I}_2^{s} \tag{22}$$

式中,$\boldsymbol{I}_1^{s}$和$\boldsymbol{I}_2^{s}$分别表示红外图像的显著分量和可见光图像的显著分量。最后的融合图像为

$$\boldsymbol{I}=\boldsymbol{I}_{\mathrm{lrr}}+\boldsymbol{I}_{\mathrm{s}} \tag{23}$$

5 算法主要结构

融合算法框图如图6所示。

假设红外图像$\boldsymbol{I}_1$和可见光图像$\boldsymbol{I}_2$已经过配准,本文融合算法的步骤为

1)用LatLRR将红外图像$\boldsymbol{I}_1$和可见光图像$\boldsymbol{I}_2$分解,可以分别获得红外图像的低秩分量$\boldsymbol{I}_1^{\mathrm{lrr}}$和显著分量$\boldsymbol{I}_1^{s}$以及可见光图像低秩分量$\boldsymbol{I}_2^{\mathrm{lrr}}$和显著分量$\boldsymbol{I}_2^{s}$。

2)将源图像输入CNN,得到两个权值分配图$\boldsymbol{S}_1$和$\boldsymbol{S}_2$。再将权值图$\boldsymbol{S}_1$和$\boldsymbol{S}_2$放大至与源图像一样的大小,分别记为为$\boldsymbol{\omega}_1$和$\boldsymbol{\omega}_2$。

3)把尺寸放大后的权值图$\boldsymbol{\omega}_1$和$\boldsymbol{\omega}_2$变成二值图像,得到的$\boldsymbol{P}_1$和$\boldsymbol{P}_2$。以$\boldsymbol{I}_1$和$\boldsymbol{I}_2$作为引导图像对$\boldsymbol{P}_1$和$\boldsymbol{P}_2$进行引导滤波,分别得到平滑的权值图$\boldsymbol{W}_1^{\mathrm{lrr}}$和$\boldsymbol{W}_2^{\mathrm{lrr}}$,以及锐化的权值图$\boldsymbol{W}_1^{s}$和$\boldsymbol{W}_2^{s}$。

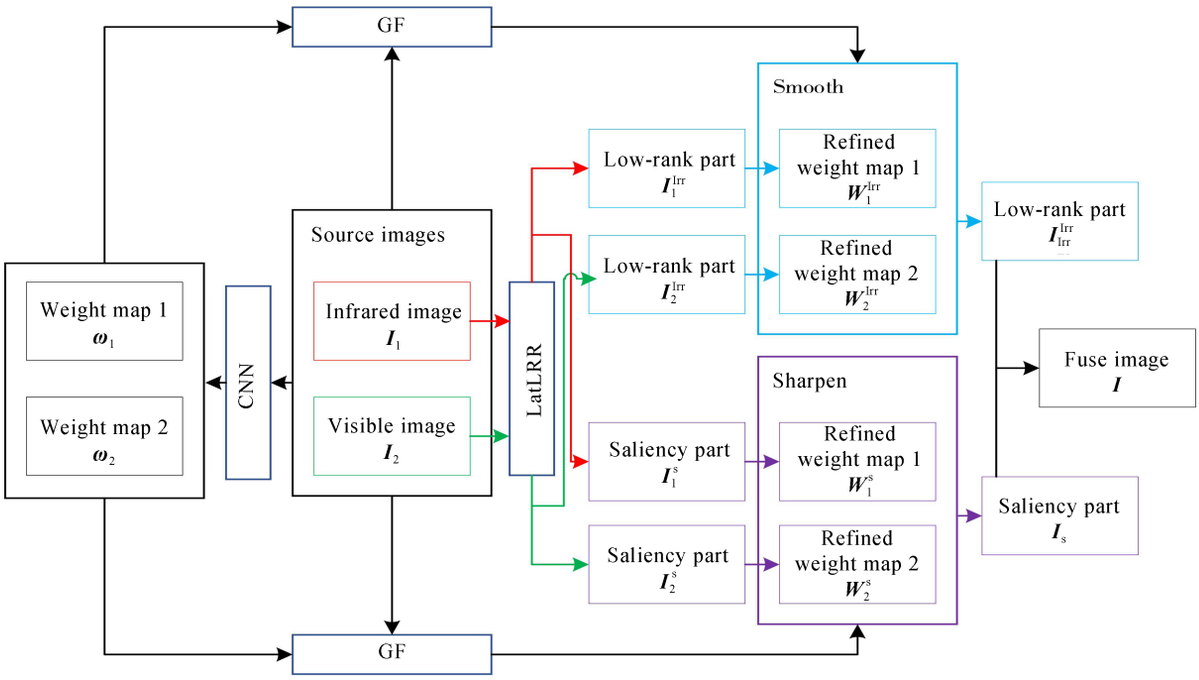


图 6 融合算法框图

Fig.6 The framework of proposed method

4)将第三步得到的权值图分别与源图像的低秩分量和显著分量做矩阵内元素对应相乘。得到融合图像的低秩分量I_{lrr}和融合图像的显著分量I_s。两者叠加,得到最终的融合图像。

6 实验结果

本文的仿真实验平台采用酷睿i5-4210M双核4线程CPU,显卡型号为AMD FirePro W4170M,内存12GB,操作系统为Win10,编程软件为Matlab2019b。选取几种比较典型的基于MST、基于SR和基于MST结合SR的红外与可见光融合算法:双树复小波变换(Dual-Tree Complex Wavelet Transform, DTCWT)[26]、曲波变换(Curvelet Transform, CVT)[27]、一种稀疏表示融合算法SR、一种NSCT结合SR融合算法NSCT_SR。还包括文献[17]和文献[15]的红外与可见光融合算法。其中,DTCWT算法分解层数为4层,高频部分按取最大值的规则进行融合,低频部分按取平均值的方法进行融合;CVT算法分解层数为4层,高频部分按取最大值的规则进行融合,低频部分按取平均值的方法进行融合;SR算法将两幅源图像按8×8的图像块分解,训练字典采用(K-Singular Value Decomposition, K-SVD),通过训练字典计算出的稀疏系数向量按最大l_1范数进行选择,最后再将选择的稀疏系数向量转换成图像块进行图像的重构,重构误差设置为0.1;NSCT结合SR算法先将两幅源图像以NSCT算法分解,分解层数为4层,方向滤波器设置为“vk”,分解滤波器设置为“pyrexc”,4层分解方向分别为4、8、8、16。经NSCT分解得到的高频部分按取最大值的规则进行融合,得到的低频部分按上述相同的SR算法进行分解和融合。实验采用4组红外和可见光图像对算法有效性进行验证。

图7为第一组红外与可见光图像融合的结果。图7(a)与图7(b)为源图像,图7(c)至图7(i)分别为不同方法的融合结果。可以看出,在不影响绿框处目标显示的情况下,本文方法在对可见光图像的红框处细节保留是最完整的。文献[17]的融合方法总体效果较好,尤其是目标的显示比较清晰,但可以看到红框内屋顶的像素与源可见光图像有很大的差异,这说明该方法将红外图像非目标的信息也过多的带入融合图像当中;文献[15]的融合方法对比度明显较差,可以看到图像中植物和道路的像素趋于一致,在图像中部也可以

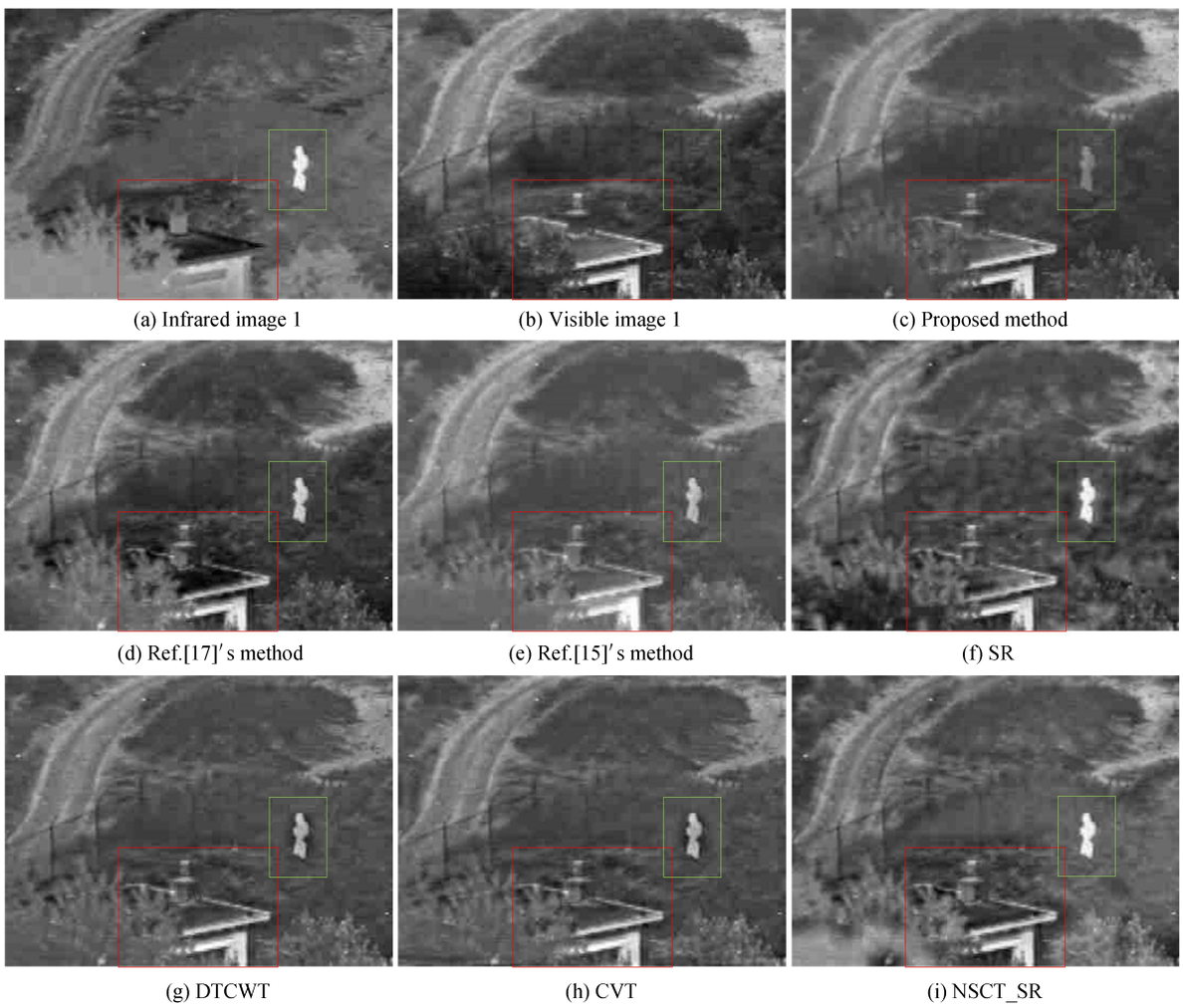
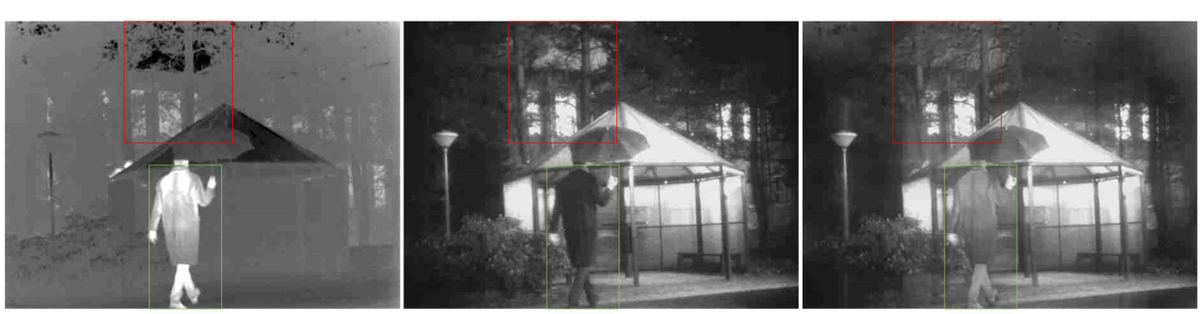


图 7　第一组红外与可见光图像融合结果对比

Fig.7　Comparison of the first set of infrared and visible image fusion results

看出细节比较模糊；DTCWT 和 CVT 融合方法目标周围存在比较明显的伪影，图像整体的对比度也比较差；而 SR 融合方法虽然目标显示最为清晰，但图像的特征显示非常差，道路和房檐出现扭曲，植物已经难以辨认，这说明 SR 的字典是不完善的；NSCT 结合 SR 融合方法的特征显示较 SR 有明显的改善，可以看出在高频系数的融合上取最大值的办法是有一定效果的。但图像的细节显示仍然非常模糊，左下角的植物难以辨认，图像中部的植物也只保留了轮廓特征。

图 8 为第二组红外与可见光图像融合的结果。图 8(a)与图 8(b)为源图像，图 8(c)至图 8(i)分别为不同方法的融合结果。可以看出，文献[17]的融合方法红框处天空的像素差别太大，显然不符合真实环境，这同样是因为融合图像过多的受到了红外图像非目标信息的影响。另外，绿框处目标与建筑的像素差别又太小，画面缺乏层次感；文献[15]的融合方法对比度较差，融合图像整体泛白；而 DTCWT 和 CVT 的融合方法



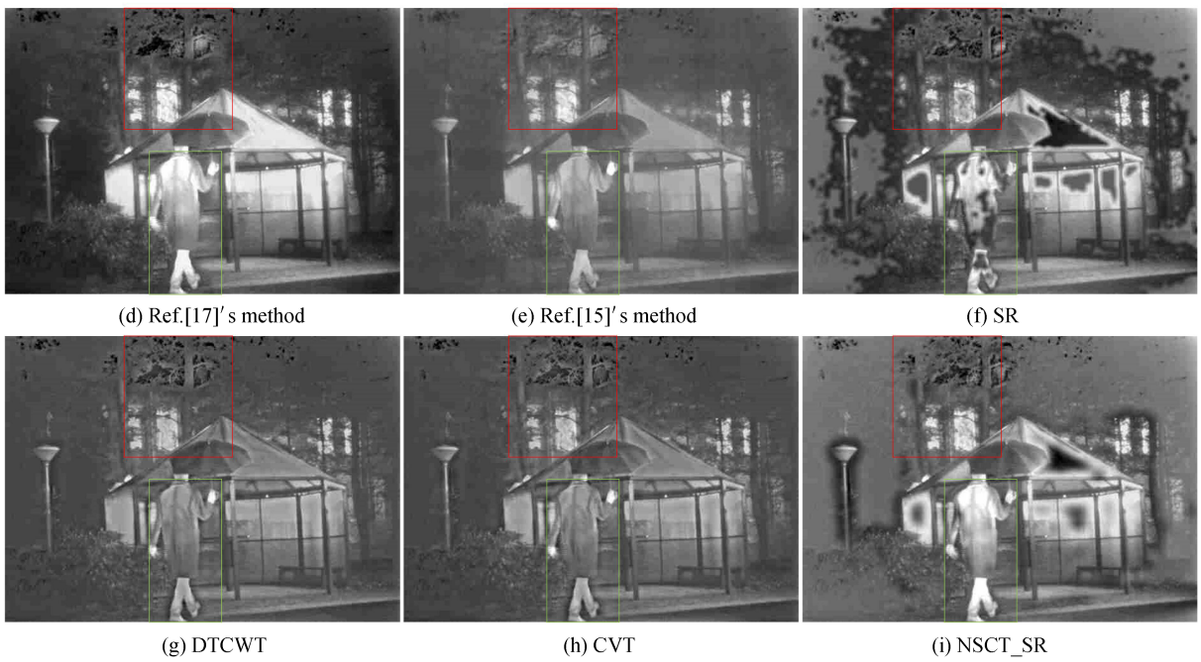
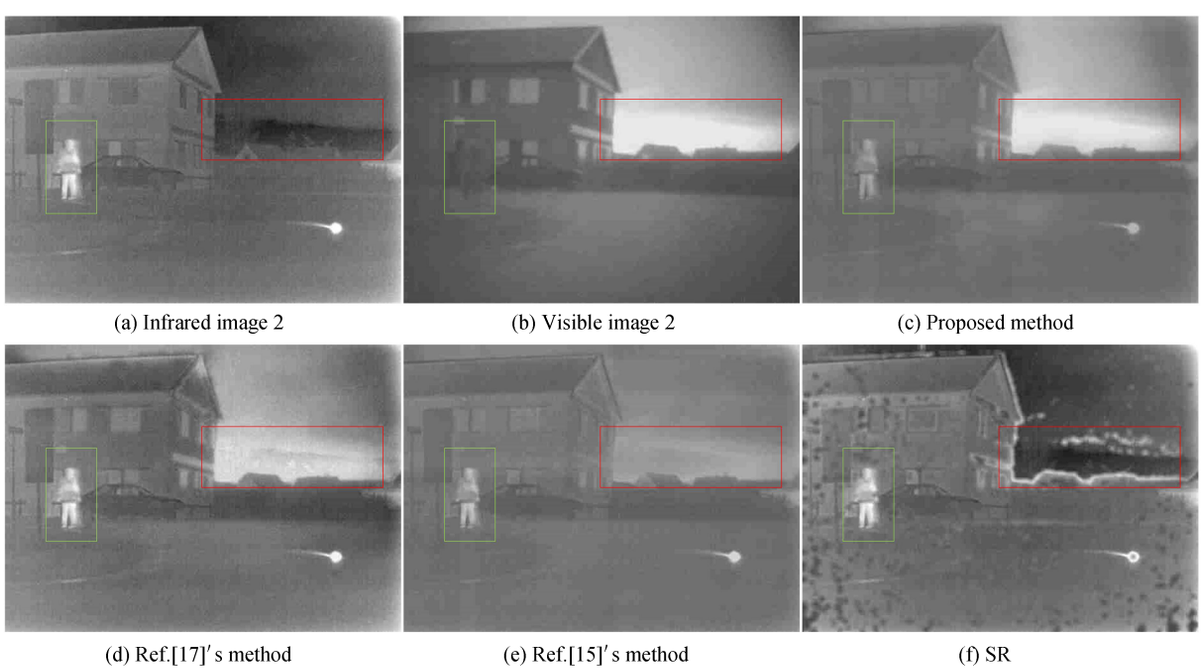


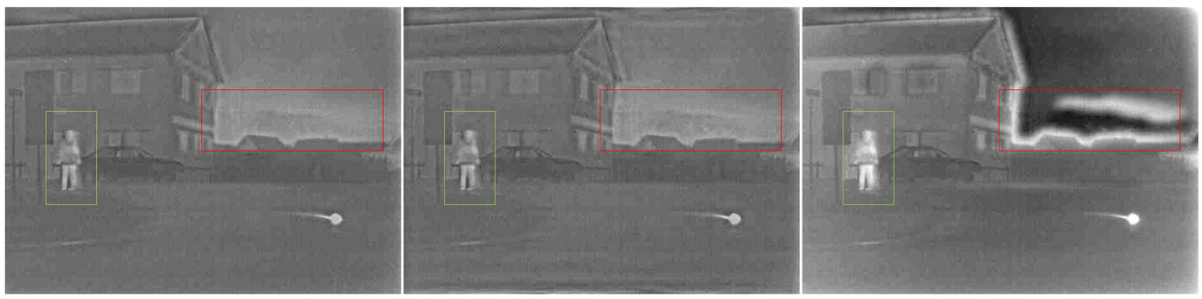
图 8 第二组红外与可见光图像融合结果对比

Fig.8 Comparison of the second set of infrared and visible image fusion results

在处于红框处的植物显示细节较差，边缘呈渐变效果，显然不符合真实环境；SR融合方法在图像的细节显示上有很明显的缺陷，建筑物和目标都出现了伪影，植物和环境之间的过渡也出现了很多边缘显示问题；而NSCT结合SR融合方法在伪影和边缘显示上虽然较SR有明显的进步，但还是没有减少到可以接受的程度。总体来看，本文方法在像素差异上与源可见光图像是最为接近的，清晰度与对比度有明显的优势，目标的显示状态也处于合理的范围内。

图9为第三组红外与可见光图像融合的结果。图9(a)与图9(b)为源图像，图9(c)至图9(i)分别为不同方法的融合结果。可以看出，除了本文方法外，其余几种融合方法在红框处都不同程度的出现了伪影。文献[17]融合方法在房顶处出现了变形，且边界显示效果较差；文献[15]融合方法由于对比度较低，楼房面朝右边的窗口已经很难辨认；DTCWT和CVT融合方法噪声比较明显，清晰度较差，而且建筑物边缘存在伪





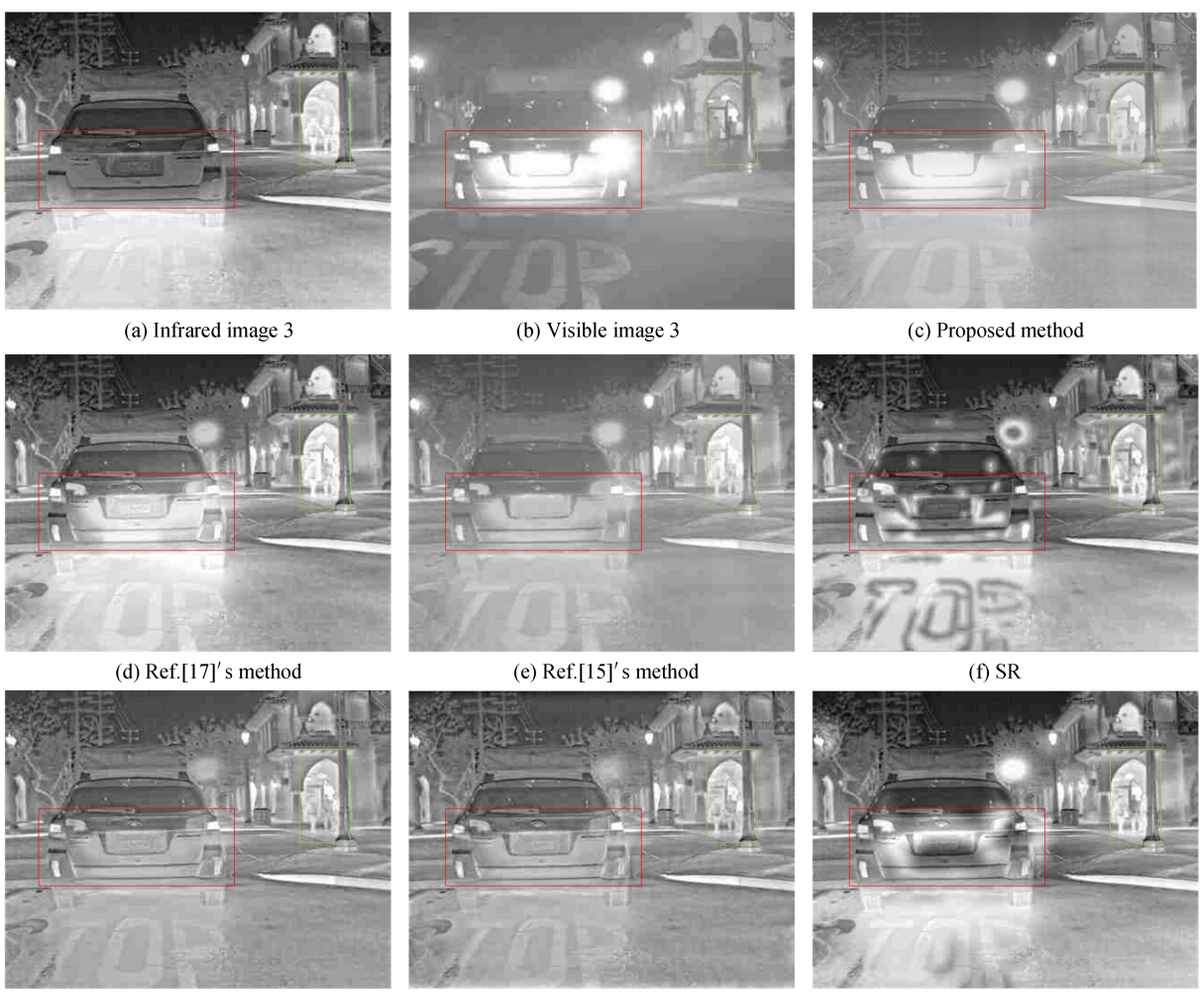
(g) DTCWT　(h) CVT　(i) NSCT & SR

图 9　第三组红外与可见光图像融合结果对比

Fig.9　Comparison of the third set of infrared and visible image fusion results

影；SR融合方法噪声程度非常严重，细节显示效果很差；而NSCT结合SR融合方法噪声程度较SR有明显进步，但建筑物周围出现了明显的伪影。总体来看，本文方法清晰度和对比度都明显好于其他几种图像融合方法，目标的辨识度也在合理范围内。

图10为第四组红外与可见光图像融合的结果。图10中的源图像图10(a)和图10(b)取自车载FLIR热传感器和CCD传感器在高级驾驶辅助系统中的实际应用图像[28]。图10(c)至图10(i)分别为不同方法的融合结果。从图中可以看到，在红框处的车辆尾部，本文算法无论是在细节或是对比度上都是最接近于可见



(a) Infrared image 3　(b) Visible image 3　(c) Proposed method

(d) Ref.[17]′s method　(e) Ref.[15]′s method　(f) SR

(g) DTCWT　(h) CVT　(i) NSCT & SR

图 10　第四组红外与可见光图像融合结果对比

Fig.10　Comparison of the fourth set of infrared and visible image fusion results

光图像的，路面字体也比较清晰。这说明本文算法在保留源可见光图像的背景信息上有较好的效果。整体来看，文献[17]的融合算法在灰度的过渡上显得太过强烈，比如车辆右上方的照明和周围的植物就显示出过多的层次；文献[15]的融合算法对比度过低，尤其在车辆尾部该问题特别突出；SR算法的噪声和伪影都比较严重；DTCWT和CVT算法的图像不够清晰，对比度较低，路面字体存在比较明显的灰度差；而NSCT结合SR算法的噪声和伪影程度较SR有明显进步，目标辨识度也较好，但车辆和路面的字体出现了不应有的灰度差。所以，本文的算法在清晰度和背景信息的显示上较其它几种算法有比较明显的优势，目标的辨识度也在合理的范围内。

对上述的融合图像采用4种常用的融合图像评价指标[29]进行客观质量评价，分别为信息论指标Q_{MI}，梯度特征指标Q_P，图像结构相似度指标Q_S和人类视觉敏感度指标Q_{CB}。其中，互信息指标Q_{MI}用于度量两幅图像相关程度，也就是融合图像中包含源图像的信息量；梯度指标Q_P用于度量源图像至融合图像中的梯度信息；结构相似度指标Q_S用于度量融合图像的结构信息保存程度；视觉敏感度指标Q_{CB}取全局质量图均值。具体的计算方法参考文献[29]。

根据表1的数据，在第一组图像中，本文方法在互信息指标Q_{MI}、梯度特征指标Q_P和人类视觉敏感度指标Q_{CB}都有不同程度的领先，结构相似度指标Q_S也与排名最高的文献[15]融合方法差别不大；在第二组图像中，本文方法在互信息指标Q_{MI}和人类视觉敏感度指标Q_{CB}领先。由于文献[17]融合方法在图像对比度

表1 不同融合方法客观评价结果

Table 1 Objective evaluation results of different fusion methods

Image	Fusion method	Q_{MI}	Q_P	Q_S	Q_{CB}
First set of fusion results	Proposed method	**0.374 1**	**0.416 8**	0.789 4	**0.589 6**
	Ref. [17]	0.285 0	0.324 4	0.796 1	0.582 4
	Ref. [15]	0.257 0	0.250 6	**0.803 8**	0.519 1
	SR	0.298 7	0.167 7	0.726 2	0.552 0
	DTCWT	0.223 1	0.273 1	0.790 8	0.549 8
	CVT	0.210 4	0.220 7	0.780 4	0.528 6
	NSCT_SR	0.232 2	0.251 7	0.784 2	0.564 0
Second set of fusion results	Proposed method	**0.460 3**	0.435 8	0.784 6	**0.550 5**
	Ref. [17]	0.347 5	**0.477 0**	**0.810 7**	0.522 6
	Ref. [15]	0.280 6	0.336 1	0.783 8	0.466 9
	SR	0.392 6	0.234 2	0.704 6	0.503 2
	DTCWT	0.237 8	0.428 7	0.804 0	0.456 9
	CVT	0.224 5	0.359 1	0.796 3	0.455 5
	NSCT_SR	0.307 1	0.383 5	0.781 3	0.462 2
Third set of fusion results	Proposed method	0.399 5	**0.263 1**	**0.844 9**	0.466 5
	Ref. [17]	0.335 2	0.255 7	0.824 6	0.486 2
	Ref. [15]	0.218 2	0.230 6	0.801 7	0.461 1
	SR	0.444 4	0.117 3	0.780 9	0.494 1
	DTCWT	0.185 6	0.212 7	0.845 5	0.487 8
	CVT	0.163 3	0.184 6	0.830 6	0.477 1
	NSCT_SR	**0.498 3**	0.256 6	0.842 2	**0.520 6**
Fourth set of fusion results	Proposed method	0.358 1	0.414 5	**0.882 1**	0.474 4
	Ref. [17]	0.378 5	**0.509 6**	0.764 2	0.553 1
	Ref. [15]	0.333 9	0.389 0	0.801 8	0.511 6
	SR	**0.512 9**	0.382 4	0.796 3	**0.579 4**
	DTCWT	0.502 4	0.331 5	0.788 0	0.548 3
	CVT	0.283 3	0.414 8	0.847 3	0.508 5
	NSCT_SR	0.466 4	0.435 5	0.870 9	0.564 1

上较其它算法明显，如红框部分的图像有非常明显的像素差，比较接近于红外图像，所以该算法在Q_P和Q_S指标上领先；在第三组图像中，本文方法在梯度特征指标Q_P和结构相似度指标Q_S领先。NSCT结合SR融合方法在目标辨识度上有比较明显的优势，最接近红外光图像，而且具有非常严重的伪影，因而对比度也比较高。所以，该算法在Q_{MI}和Q_{CB}指标上领先。但考虑到SR和NSCT结合SR融合方法在第三组图像中糟糕的主观评价，在互信息指标Q_{MI}上排名第三的本文方法同样比其他融合方法有明显的优势；在第四组图像中，本文方法在结构相似度指标Q_S领先。由于SR算法存在严重的噪音和伪影，与NSCT结合SR融合方法在第三组图像中的情况类似，该算法在Q_{MI}和Q_{CB}指标上领先。文献[17]融合方法的图像对比度是最高的，所以该算法在Q_P指标上领先。综上所述，除去一些主观评价比较差的融合方法，本文方法在大部分客观指标上是具有优势的。基于以上分析，本文方法具有有效性

7 结论

本文提出一种在潜在低秩表示框架下基于卷积神经网络结合引导滤波的红外与可见光图像融合方法，与其他融合方法比较本文方法在保留可见光图像细节信息上有较为明显的优势。实验结果表明，本文方法在大部分客观评价指标上具有一定优势，并且主观评价也具有良好的视觉效果，能够增强观察者对场景的识别能力，有助于后期开展的探测或识别等实际应用。但同时本方法在目标的显示上不是特别突出，如何在尽可能保留可见光图像细节信息的同时，增强目标的突出显示是下一阶段研究的重点。

参考文献

[1] KUMAR W K, NONGMEIKAPAM K, SINGH A D, et al. Enhancing scene perception using a multispectral fusion of visible-near-infrared image pair[J]. IET Image Processing, 2019, 13(13): 2467-2479.

[2] AGRAWAL D, KARAR V. Bispectral image fusion using multi-resolution transform for enhanced target detection in low ambient light conditions[J]. Indian Journal of Pure and Applied Physics, 2019, 57(1): 33-41.

[3] DOGRA A, KADRY S, GOYAL B, et al. An efficient image integration algorithm for night mode vision applications[J]. Multimedia Tools and Application, 2020, 79(4): 10995-11012.

[4] ZHAN Lingchao, ZHUANG Yi, HUANG Longda. Infrared and visible images fusion method based on discrete wavelet transform[J]. Journal of Computers, 2017, 28(2): 57-71.

[5] SEETHALAKSHMI K, VALLI S. A fuzzy approach to recognize face using contourlet transform[J]. International Journal of Fuzzy Systems, 2019, 21(7): 2204-2211.

[6] WU Wenfu, GUO Songjing, CHENG Qimin. Fusing optical and synthetic aperture radar images based on shearlet transform to improve urban impervious surface extraction[J]. Journal of Applied Remote Sensing, 2020, 14(2): 024506.

[7] FENG Xin, HU Kaiqun, YUAN Yi, et al. Multi-focus image fusion based on super-resolution and group sparse representation[J]. Acta Photonica Sinica, 2019, 48(7): 0710003.
冯鑫, 胡开群, 袁毅, 等. 基于超分辨率和组稀疏表示的多聚焦图像融合[J]. 光子学报, 2019, 48(7): 0710003

[8] RAGHUWANSHI G, TYAGI V. Feed-forward content based image retrieval using adaptive tetrolet transforms[J]. Multimedia Tools and Applications, 2018, 77(6): 23389-23410.

[9] LI Wei, LIN Qinyong, WANG Keqiang, et al. Improving medical image fusion method using fuzzy entropy and nonsubsampling contourlet transform[J]. International Journal of Imaging Systems and Technology, 2020: 1-11.

[10] CHU Tianyong, TAN Yumin, LIU Qiang, et al. Novel fusion method for SAR and optical images based on nonsubsampled shearlet transform[J]. International Journal of Remote Sensing, 2020, 41(12): 4588-4602.

[11] YANG Bin, LI Shutao. Multifocus Image fusion and restoration with sparse representation[J]. IEEE Transactions on Instrumentation and Measurement, 2010, 59(4): 884-892.

[12] GOYAL S, SINGH V, RANI A, et al. FPRSGF denoised non-subsampled shearlet transform-based image fusion using sparse representation[J]. Signal Image and Video Processing, 2020, 14(4): 719-726.

[13] LIU Feiqiang, CHEN Lihui, LU lu, et al. Medical image fusion method by using Laplacian pyramid and convolutional sparse representation[J]. Concurency and Computation Practice and Experience, 2019, 32(17): e5632.

[14] LIU Guangcan, YAN Shuicheng. Latent low-rank representation for subspace segmentation and feature extraction[C]. 2011 International Conference on Computer Vision, IEEE, 2011: 1615-1622.

[15] LI Hui, WU Xiaojun. Infrared and visible image fusion using latent low-rank representation[OE]. 2018, arXiv: 1804.08992, https://arxiv.org/abs/1804.08992.

[16] LI Shutao, KANG Xudong, HU Jianwen. Image fusion with guided filtering[J]. IEEE Transactions on Image Processing, 2013, 22(7): 2864-2875.

[17] LIU Yu, CHEN Xun, CHENG Juan, et al. Infrared and visible image fusion with convolutional neural networks[J].

International Journal of Wavelets Multiresolution and Information Processing, 2017, 16(3): 1850018.
[18] LIU Guangcan, LIN Zhouchen, YU Yong. Robust subspace segmentation by low-rank representation[C]. Proceedings of the 27th International Conference on Machine Learning, 2010.
[19] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep learning[M]. MIT Press, 2016.
[20] GLOROT X, BORDES A, BENGIO Y. Deep sparse rectifier neural networks[J]. Journal of Machine Learning Research, 2011, 15: 315-323.
[21] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing System, 2012, 25(2): 1097-1105.
[22] KOCH G. Siamese neural networks for one-shot image recognition[C]. ICML Deep Learning Workshop, 2015.
[23] http://www.image-net.org/.
[24] JIA Yangqing, SHELHAMER E, DONAHUE J, et al. Caffe: convolutional architecture for fast feature embedding[C]. Proceedings of the ACM International Conference on Multimedia, 2014.
[25] HE Kaiming, SUN Jian, TANG Xiaoou. Guided image filtering[J]. IEEE Transactions on Software Engineering, 2013, 35(6): 1397-1409.
[26] LEWIS J J, O'CALLAGHAN R J, NIKOLOV S G. Pixel- and region- based image fusion with complex wavelets[J]. Information Fusion, 2007, 8(2): 119-130.
[27] NENCINI F, GARZELLI A, BARONTI S, et al. Remote sensing image fusion using the curvelet transform[J]. Information Fusion, 2007, 8(2): 143-156.
[28] https://www.flir.com/oem/adas/adas-dataset-form/.
[29] LIU Zheng, ERIK B, XUE Zhiyun, et al. Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: acomparative study[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(1): 94-109.

Foundation item: The National Natural Science Fund of China (Nos. 31501229, 61861025), Chongqing Nature Science Foundation for Fundamental Science and Frontier Technologies (No. cstc2018jcyjAX0483), Science and Technology Research Program of Chongqing Education Commission of China (Nos. KJQN201900821, KJQN202000803)