文章编号　1004-924X(2024)10-1567-15

基于二次图像分解的红外图像与可见光图像融合

马　鑫，喻春雨*，童亦新，张　俊
(南京邮电大学 电子与光学工程学院、柔性电子(未来技术)学院，江苏 南京 210023)

摘要：针对红外图像与可见光图像融合中细节丢失严重，红外图像的特征信息未能突出显示以及源图像的语义信息被忽视的问题，提出一种基于二次图像分解的红外图像与可见光图像融合网络(Secondary Image Decomposition For Infrared And Visible Image Fusion，SIDFuse)。利用编码器对源图像进行二次分解以提取不同尺度的特征信息，然后利用双元素注意力为不同尺度的特征信息分配权重、引入全局语义支路，再采用像素相加法作为融合策略，最后通过解码器重建融合图像。实验选择FLIR数据集用于训练，采用TNO和RoadScene两个数据集进行测试，并选取八种图像融合客观评价参数进行实验对比分析。由TNO数据集的图像融合实验表明，在信息熵、标准差、空间频率、视觉保真度、平均梯度、差异相关系数、多层级结构相似性、梯度融合性能评价指标上，SIDFuse比基于卷积网络中经典融合算法DenseFuse分别平均提高12.2%，9.0%，90.2%，13.9%，85.1%，16.8%，6.7%，30.7%，比最新的融合网络LRRNet分别平均提高2.5%，5.6%，31.5%，5.4%，25.2%，17.9%，7.5%，20.7%。可见本文所提算法融合的图像对比度较高，可以同时更有效保留可见光图像的细节纹理和红外图像的特征信息，在同类方法中占有明显优势。

关　键　词：图像融合；图像二次分解；全局语义支路；双元素注意力；图像对比度
中图分类号：TP394.1；TH691.9　　**文献标识码**：A　　**doi**：10.37188/OPE.20243210.1567

Infrared image and visible image fusion algorithm based on secondary image decomposition

MA Xin，YU Chunyu*，TONG Yixin，ZHANG Jun

(*College of Electronic and Optical Engineering & College of Flexible Electronics (Future Technology)*，*Nanjing University of Posts and Telecommunications*，*Nanjing* 210023，*China*)
* *Corresponding author*，*E-mail*：*yucy@njupt.edu.cn*

Abstract：In view of the serious detail loss，the feature information of infrared image is not highlighted and the semantic information of source image is ignored in the fusion of infrared image and visible image，a fusion network of infrared image and visible image based on secondary image decomposition was proposed. The encoder was used to decompose the source image twice to extract the feature information of different scales，then the two-element attention was used to assign weights to the feature information of different scales，the global semantic branch is introduced，the pixel addition method was used as the fusion strategy，and the fusion image was reconstructed by the decoder. In the experiment，FLIR data set was selected for training，TNO and RoadScene data sets were used for testing，and eight objective evaluation param-

收稿日期：2023-09-26；**修订日期**：2023-11-14.
基金项目：国家自然科学基金资助项目(No. 61801239)；中央高校基本科研业务费专项资金资助项目(No. 30918014106)；南京邮电大学校企合作项目(No. 2018外002，No. 2019外157)

eters of image fusion were selected for comparative analysis. The image fusion experiment of TNO data set shows that in terms of information entropy, standard deviation, spatial frequency, visual fidelity, average gradient and difference correlation coefficient, SIDFuse is 12.2%, 9.0%, 90.2%, 13.9%, 85.1% , 16.8%,6.7%,30.7% higher than DenseFuse, the classical fusion algorithm based on convolutional networks, respectively. Compared with the latest fusion network LRRNet, the average increase is 2.5%, 5.6%, 31.5%, 5.4%, 25.2% , 17.9%,7.5%,20.7 respectively. It can be seen that the image fusion algorithm proposed in this paper has a high contrast, and can retain the detail texture of visible image and the feature information of infrared image more effectively at the same time, which has obvious advantages in similar methods.

Key words: image fusion; image secondary decomposition; global semantic branch; two-element attention; image contrast

1 引　言

图像融合是将由不同传感器获得的两张或更多张图像的显著特征融合到一张图像上以增强感知性的过程。红外图像与可见光图像的融合是图像融合的典型代表,二者融合的结果既可以保留目标红外图像的显著特征又可以保留可见光图像中目标的结构特征和纹理细节,因此红外图像与可见光图像融合技术已经被广泛应用于自动驾驶,视觉追踪和视频监控[1]。

传统图像融合方法主要为多尺度变换方法[2]、稀疏表示方法[3]、子空间方法[4]和基于显著性[5]的方法,随着深度学习技术的发展,许多基于神经网络的图像融合方法被提出。2019年Li等提出了DenseFuse[6],它是首次使用经过预训练的自编码器进行图像融合网络模型;然后Li等于2020年又提出了融合网络NestFuse[7],该网络利用巢状连接的编码器网络以及引入空间、通道注意力可以保留可见光图像中更详细的背景信息并增强红外图像的显著特征;接下来,Li等于2021年将人工设计的融合规则改为设计了一种新的可学习的融合策略,提出了RFN-Nest[8]以提取源图像中更全面信息。2020年Zhang等提出了一种通用的融合模型IFCNN[9],该网络利用卷积层从多个输入图像中获取显著特征,IFCNN可用于多聚焦融合、红外与可见光图像融合、医学图像融合等。2021年Ma等人提出了一种压缩分解网络SDNet[10],将融合问题转化为重构梯度和强度信息,这样使融合结果中包含更多场景细节。2022年Liu等人提出了一种基于神经搜索的融合网络SMOANet[11],它可以自动学习不同种类图像特征。2023年Li等人提出了一种端到端的模型LRRNet[12],它可以在参数量减少的情况下确保融合图像的细节特征和显著特征不丢失。

基于深度学习的图像融合方法虽然取得了较大突破,但是这些方法仍存在对可见光图像细节提取不足、红外图像的显著性区域不突出、融合过程中忽视源图像的语义信息。由此,本文提出了一种用于可见光与红外图像融合的图像二次分解模型(Secondary Image Decomposition For Infrared And Visible Image Fusion, SIDFuse),图像分解和重建通过编码器和解码器完成。二次图像分解是分别对源图像的特征信息进行粗尺度提取和细尺度提取。同时采用改进的双元素注意力机制模块(Dual Element-wise Attention Mechanism, DEAM),该模块自适应地对粗尺度特征和细尺度特征进行权重分配。还设计了全局语义信息提取模块避免深层密集残差网络对源图像语义信息的丢失。

2 SIDFuse工作原理

2.1 SIDFuse网络构成

如图1所示,SIDFuse训练网络框架由编码器模块、双元素注意力模块、全局语义支路模块以及解码器模块组成。其中特征提取编码器由Conv1、Conv2构成的公共卷积层、背景提取模块(Background Extraction Residual Module, BERM)和细节提取模块(Detail Extraction Residual Module, DERM)组成。SIDFuse训练网络的

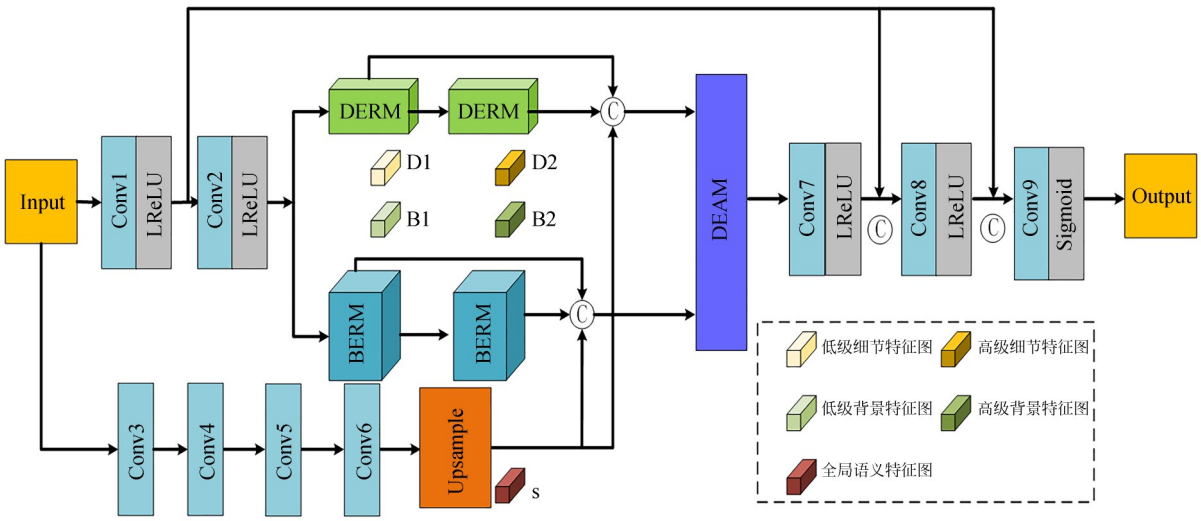


图 1　训练网络框架图

Fig. 1　Training network frame diagram

目的是获得一个对源图像进行多次分解的编码器，同时获得一个保留源图像信息的解码器，训练网络过程为：首先，将源图像分别输入公共卷积层，公共卷积层由一个 5×5 卷积层和一个 3×3 卷积层组成，激活函数使用 LReLU；然后，在上下支路分别经过细节提取模块 DERM 和背景提取模块 BERM 生成低级特征的背景和细节特征向量 B1 和 D1，并将第一次提取的低级特征向量分别送入 DERM、BERM，相应生成高级特征的背景特征向量 B2、细节特征向量 D2；再由四个 3×3 卷积层构成的全局语义特征分支对源图像进行快速下采样，并将通道数经重复卷积后依次扩展为 32，64 和 128，将图像通道扩展后再将通道数降回 64、经上采样恢复图像原始尺寸；最后将编码器提取的 B1，D1，B2，D2，S 进行通道拼接后输入 DEAM 进行特征信息自适应权重分配，再送入解码器模块经过三个 3×3 卷积层解码后输出重建图像。

SIDFuse 测试网络框架如图 2 所示，它的工

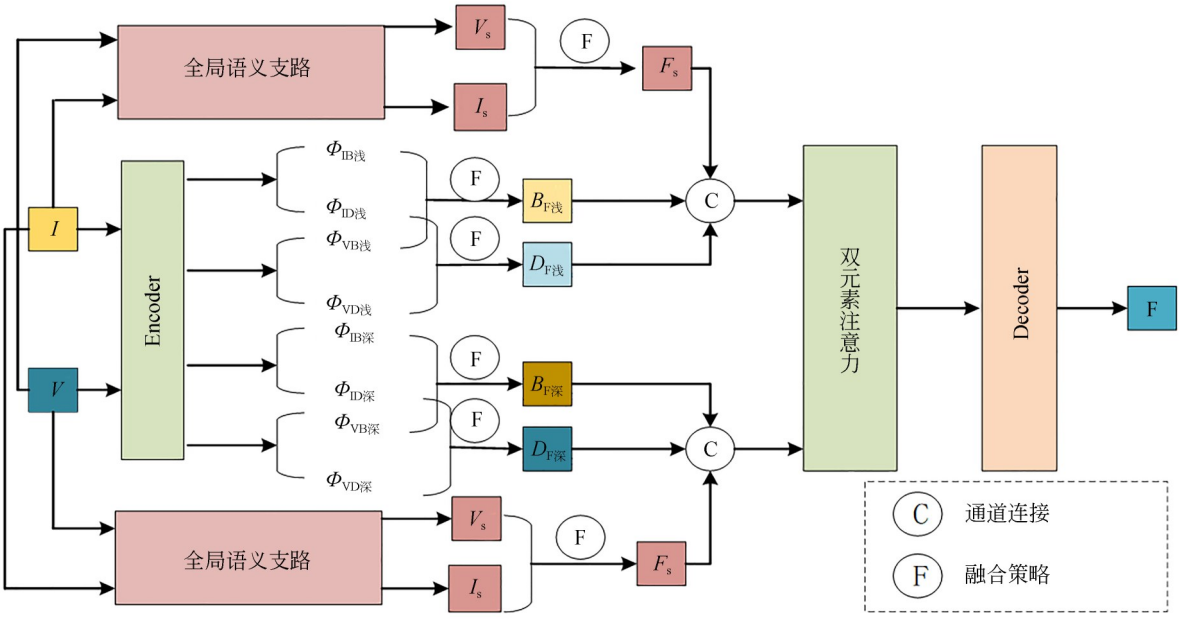


图 2　测试网络框架图

Fig. 2　Test network frame diagram

作原理为：首先将红外图像和可见光图像输入到编码器中，先由浅层编码网络在后文定义的损失函数的指导训练下生成浅层的背景特征图 $\Phi_{\mathrm{IB浅}}$，$\Phi_{\mathrm{VB浅}}$ 和细节特征图 $\Phi_{\mathrm{ID浅}}$，$\Phi_{\mathrm{VD浅}}$，后经过深层编码网络后生成深层特征的背景特征图 $\Phi_{\mathrm{IB深}}$，$\Phi_{\mathrm{VB深}}$ 和细节特征图 $\Phi_{\mathrm{ID深}}$，$\Phi_{\mathrm{VD深}}$；将提取到的浅、深层对应特征图利用特定融合规则进行融合，分别得到浅层背景、细节特征图 $B_{\mathrm{F浅}}$，$D_{\mathrm{F浅}}$ 和深层背景、细节特征图 $B_{\mathrm{F深}}$，$D_{\mathrm{F深}}$；再通过全局语义分支分别提取可见光、红外图像的语义信息 V_s，I_s，将二者融合生成 F_s；最后将浅层、深层、全局语义的三种融合信息经过通道拼接后送入 DEAM，将经过注意力机制的特征信息通过解码器得到融合图像。

2.2　背景提取残差模块和细节提取残差模块

背景提取残差模块（Background Extraction Residual Module，BERM）[13] 和细节提取模块（Detail Extraction Residual Module，DERM）是编码器中进行信息提取的主要模块，BERM 结构如图 3 所示，每个 BERM 由四个卷积层 Conv1，Conv2，Conv3，Conv4 和一个跳连的普通卷积层 Conv 组成。Conv1，Conv4 卷积核大小为 1×1，Conv2，Conv3 卷积核大小为 3×3。Conv1，Conv2，Conv3 均使用 LReLu 作为激活函数，经 Conv1 生成低维度特征映射，Conv4 输出与 Conv 输出相加。DERM 模块如图 4 所示，用来提取可见光图像和红外图像中的细节细粒度特征，其主流支路采用 2 个 3×3 卷积层和 2 个 1×1 卷积层来提取特征，采用 LReLU 为激活函数。残差支路中通过 Sobel 梯度运算来计算特征的梯度幅度，采用 1×1 卷积层消除输入输出通道维度的不同。最后将主流支路和残差支路通过元素相加保留深度信息和细粒度细节特征。

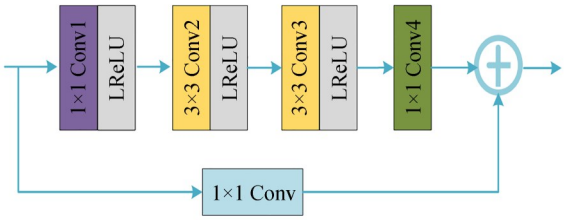


图 3　BERM 模块

Fig. 3　BERM module

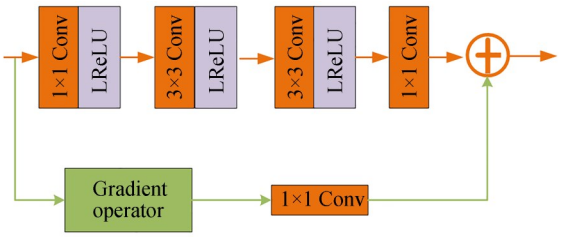


图 4　DERM 模块

Fig. 4　DERM module

2.3　双元素注意力机制

改进的双元素注意力机制 DEAM[14] 模块如图 5 所示，它拥有浅层特征模块 b 和深层特征模块 f 两个输入。首先利用上采样调整输入 f，通过 Concat 层将双元素拼接为 c，然后先将 c 送入权重映射模块通过一个 1×1 卷积层进行降维，再通过层间使用 ReLU 激活函数的两个 3×3 卷积层，最后通过 Sigmoid 激活层将权重归一为权重张量 α。之后在双值权重发生器中得到权重 α 和 β，$\beta=1-\alpha$。最终 DEAM 模块的输出可以表示为：$g=\alpha\otimes b+\beta\otimes f_0$。

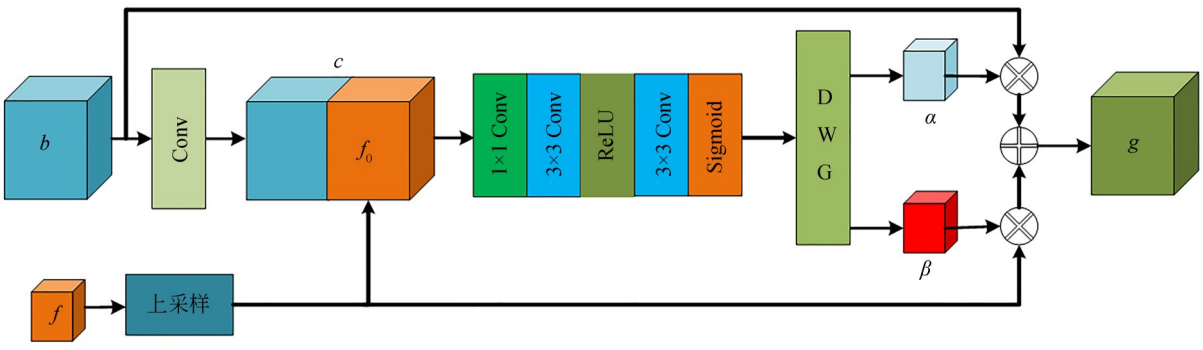


图 5　双元素注意力模块

Fig. 5　Dual element-wise attention mechanism module

2.4　损失函数

在训练阶段通过信息提取支路，获得一个能将图像多次分解的编码器，以及能使融合图像很好保留源图像信息的解码器，因此在训练阶段忽略融合策略的影响。在对源图像进行图像分解时，背景特征图是红外图像与可见光图像的共同特征，细节特征图则用来捕捉红外图像与可见光图像的不同特征[15]。为了减小背景特征图的差距以及加大细节特征图的差距，定义图像分解的损失函数为：

$$L=\Phi(\|B_V-B_I\|_2^2)-\alpha\Phi(\|D_V-D_I\|_2^2),\quad(1)$$

其中：B_V，D_V分别为可见光图像的背景、细节特征图；B_I，D_I分别为红外图像的背景、细节特征图；$\Phi(\cdot)$是tanh函数，目的是将范围限制在(−1,1)。

SIDFuse中对于输入的源图像进行二次图像分离，定义对于浅层特征细节背景分离和深层特征细节背景分离的损失函数分别为：

$$L_{浅}=\Phi(\|B_{V浅}-B_{I浅}\|_2^2)-\alpha_1\Phi(\|D_{V浅}-D_{I浅}\|_2^2),\quad(2)$$

$$L_{深}=\Phi(\|B_{V深}-B_{I深}\|_2^2)-\alpha_2\Phi(\|D_{V深}-D_{I深}\|_2^2).\quad(3)$$

在图像重建时，定义重建图像的损失函数为：

$$L_R=\alpha_3 f(I,\hat{I})+\alpha_4 f(V,\hat{V})+\alpha_5\|\nabla V-\nabla\hat{V}\|_1,\quad(4)$$

其中：I，$\hat{I}$为输入的红外图像、重建后的红外图像；V，$\hat{V}$为输入的可见光图像、重建后的可见光图像；∇代表梯度操作，在L_R的最后一项表示对可见光图像进行梯度操作并进行正则化，以确保重建图像与源图像纹理上的一致。在L_R中：

$$f(X,\hat{X})=\|X-\hat{X}\|_2^2+\lambda L_{SSIM}(X,\hat{X}),\quad(5)$$

其中：X，$\hat{X}$分别表示在L_R中的输入图像和重建图像；λ是一个超参数。SSIM是结构相似性(Structural similarity)指数，表示两张图像在亮度、对比度和结构上的相似程度。在公式(5)中，L_{SSLM}用来衡量原始像素与重建图像像素的一致性。L_{SSLM}可以表示为：

$$L_{SSIM}(X,\hat{X})=\frac{1-SSIM(X,\hat{X})}{2},\quad(6)$$

$$L_{SSIM}(X,\hat{X})=[l(X,\hat{X})]^{\alpha}\cdot[l(X,\hat{X})]^{\beta}\cdot[l(X,\hat{X})]^{\gamma}.\quad(7)$$

根据公式(1)～式(4)，设计的总体损失函数为：

$$L_{total}=L_{浅}+L_{深}+L_R=\Phi(\|B_{V浅}-B_{I浅}\|_2^2)-\alpha_1\Phi(\|D_{V浅}-D_{I浅}\|_2^2)+\Phi(\|B_{V深}-B_{I深}\|_2^2)-\alpha_2\Phi(\|D_{V深}-D_{I深}\|_2^2)+\alpha_3 f(I,\hat{I})+\alpha_4 f(V,\hat{V})+\alpha_5\|\nabla V-\nabla\hat{V}\|_1.\quad(8)$$

2.5　融合策略

测试阶段的网络框架如图2所示。在此网络中加入融合策略的目的是将编码器从可见光图像和红外图像中提取的特征在融合中最大程度地保留，因此定义融合方法为：

$$B_F=Fusion(B_I,B_V)\,D_F=Fusion(D_I,D_V),\quad(9)$$

其中，B_F，D_F为融合后的背景特征图和细节特征图。实验中对比了以下3种融合策略。

(1)像素相加法：将从可见光和红外图像中提取的背景特征图和细节背景图直接相加构成融合图像的细节特征图与背景细节图。由于在编码器部分对源图像进行了二次细节、背景分解，所以融合策略表示为：

$$B_{F浅}=B_{I浅}+B_{V浅}\,B_{F深}=B_{I深}+B_{V深},\quad(10)$$

$$D_{F浅}=D_{I浅}+D_{V浅}\,D_{F深}=D_{I深}+D_{V深}.\quad(11)$$

(2)加权相加法：将分别从可见光图像和红外图像提取到的特征图进行加权后相加。所以加权平均法的公式为：

$$B_F=\mu_1 B_I+\mu_2 B_V,\,D_F=\mu_3 D_I+\mu_4 D_V.\quad(12)$$

由于在编码器中采用二次分解，所以：

$$B_{F浅}=\lambda_1 B_{I浅}+\lambda_2 B_{V浅}\,B_{F深}=\lambda_3 B_{I深}+\lambda_4 B_{V深},\quad(13)$$

$$D_{F浅}=\lambda_5 D_{I浅}+\lambda_6 D_{V浅}\,D_{F深}=\lambda_7 D_{I深}+\lambda_8 D_{V深}.\quad(14)$$

为了使得融合图像均匀地从可见光图像和红外图像获取特征，避免因为权重分配而使得融合特征偏向可见光或者红外图像，规定$\lambda_i=0.5$ ($i=1,2,\cdots,8$)。

(3)通道注意力：如图6所示，将对于从红外图像和可见光图像提取的背景特征图以及细节特征图分别利用通道注意力进行融合，用B_I，B_V分别表示红外图像和可见光图像的背景特征图。细节特征图的融合策略与此类似。

将红外特征图、可见光特征图采用全局平均

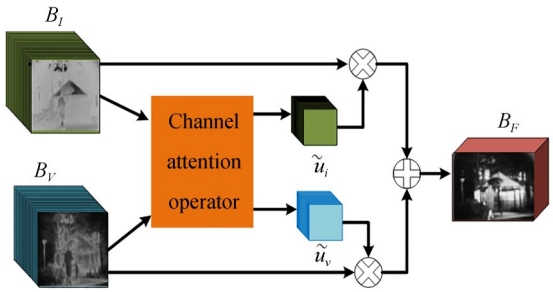


图 6　通道注意力示意图

Fig. 6　Channel attention diagram

池化操作 Ω(•)分别提取 B_I,B_V 各通道的初始活动度 u_i,u_v,然后归一化得到 $\tilde{u}_i$ 和 $\tilde{u}_v$,并将 B_I 与 $\tilde{u}_i$ 相乘、B_V 与 $\tilde{u}_v$ 相乘最终得到增强的信道特征 B_F 如公式(16)所示：

$$u_i=\Omega(B_I),u_v=\Omega(B_V), \tag{15}$$

$$\tilde{u}_i=\frac{u_i}{u_i+u_v},\tilde{u}_v=1-\tilde{u}_i, \tag{16}$$

$$B_F=\tilde{u}_iB_I+\tilde{u}_vB_V. \tag{17}$$

3　实验结果与分析

3.1　实验设置

软件环境为 Windows10，python3.6 及 PyTorch 深度学习框架；硬件环境为 GeForce RTX 3070 8 G 显卡。选取 FLIR 数据集中 200 对图像作为训练集图像，并通过翻转、裁剪等方法进行数据增强后训练集为 800 对图像。在训练前将所有图像转化为灰度图像，并进行居中裁剪为 128×128 大小。在训练时将批量数设置为 4，优化器选用 Adma[15] 优化器，训练阶段设置 epoch 为 260，初始学习率为 1×10^{-3}，并设置每 50 个 epoch 后学习率衰减 10 倍。采用 TNO[16]，RoadScene[17] 两个经典数据集作为测试集。在测试阶段分别从 TNO 数据集选取 25 对红外与可见光图像作为第一组测试集，在 RoadScene 数据集中选用 40 对数据作为第二组测试集。

3.2　融合图像评价指标

在本文中使用信息熵(Entropy，EN)、标准差(Standard Deviation，SD)，空间频率(Spatial Frequency，SF)、视觉保真度(Visual Fidelity，VIF)、平均梯度(Average Gradient，AG)、差异相关系数(Sum of Correlations of Differences，SCD)、多层级结构相似性(Multi-Scale Structure Similarity，MS-SSIM)、梯度融合性能评价指标 Q_{abf} 八种客观评价指标对融合图像进行评价。EN 和 SD 用来衡量图像中包含的信息量；SF 用于测量图像的梯度分布；VIF 表示融合图像中保留源图像结构信息的能力；AG 用于衡量融合图像中纹理和细节表征能力；SCD 表示从源图像传输到融合图像的信息量；MS-SSIM 能衡量不同尺度下的结构相似程度；Q_{abf} 用于表示来自输入的显著信息在融合图像中的表现程度。

3.3　融合策略比较实验

在含有 25 张图像的 TNO 数据集和含有 40 张图像的 RoadScene 两个不同的数据集上对于 2.5 节提出的三种融合策略进行比较。

表 1 和表 2 分别显示了 TNO 和 RoadScene 数据集在三种融合策略下六个客观指标上的数

表 1　TNO 数据集三种融合策略的融合结果对比

Tab. 1　Comparison of fusion results of three fusion strategies in TNO dataset

	EN	SD	SF	VIF	AG	SCD
像素相加法	**7.338 9**	**9.843 0**	0.046 8	**0.828 0**	**4.529 4**	**1.927 8**
加权相加法	6.742 4	9.116 2	0.027 7	0.698 0	2.725 8	1.695 2
通道注意力	7.337 1	9.816 4	**0.047 0**	0.812 6	4.445 7	1.864 4

表 2　RoadScene 数据集三种融合策略的融合结果对比

Tab. 2　Comparison of fusion results of three fusion strategies in RoadScene dataset

	EN	SD	SF	VIF	AG	SCD
像素相加法	**7.377 8**	10.929 6	0.055 6	**0.715 4**	**5.452 3**	**1.809 2**
加权相加法	6.920 3	9.715 2	0.035 6	0.651 9	3.615 5	1.459 2
通道注意力	7.357 3	**11.247 6**	**0.056 0**	0.690 7	5.303 4	1.796 3

值对比，均取平均值。由两张表的对比可以发现，像素相加法融合策略明显优于加权相加法和通道注意力这两种融合策略。所以本文采用像素相加法作为融合策略。

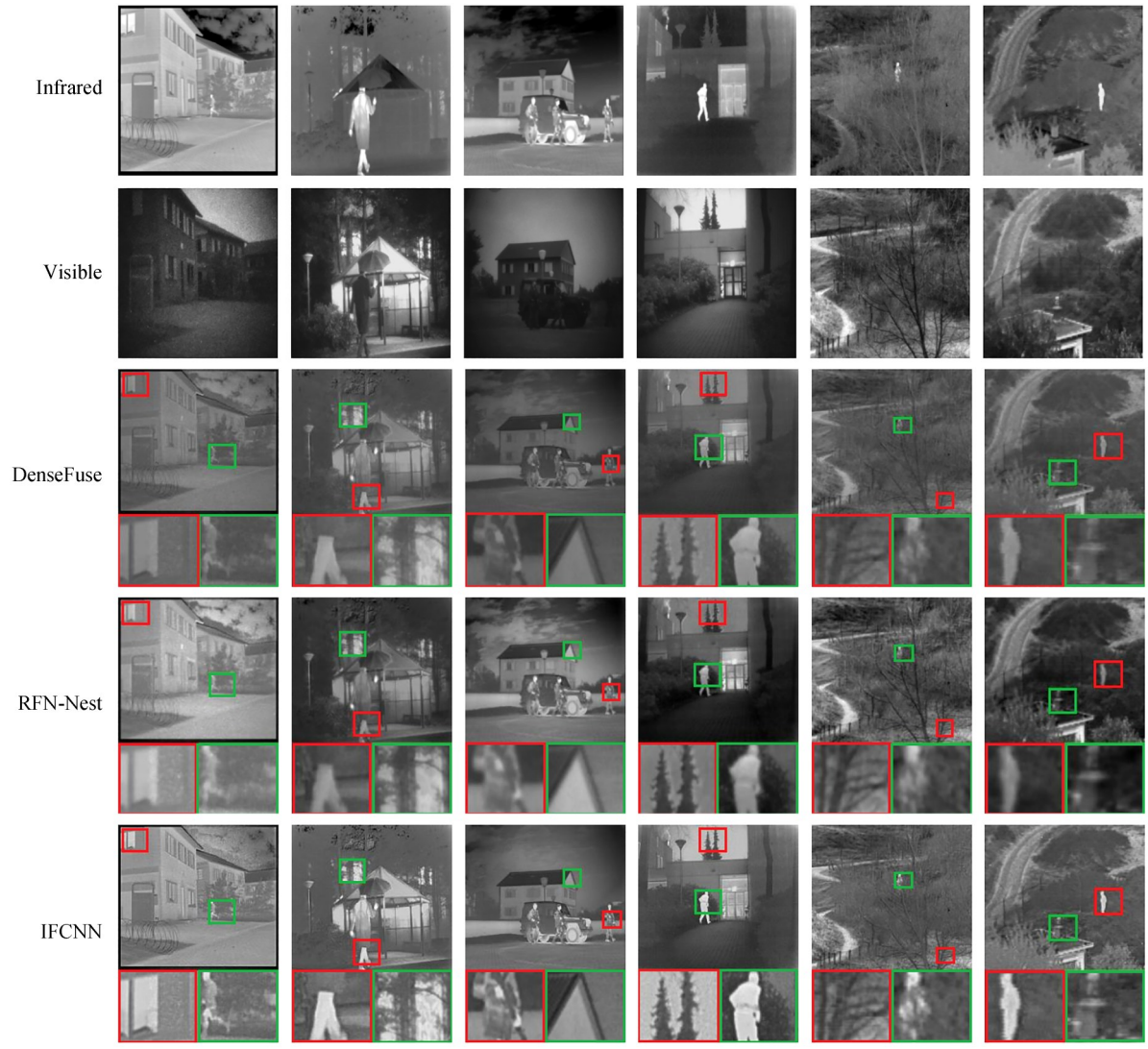
3.4　与目前主流融合方法比较实验

选取 9 种目前主流的红外与可见光融合方法在 TNO，RoadScence 这两个数据集上进行比较。在 TNO，RoadScene 分别选取 25 对、40 对图像进行比较实验。采用 DenseFuse，RFN-Nest，IFCNN，NestFuse，GTF[18]，Dualbranch[19]，LRRNet，SMOA，SDNet 等九种具有代表性的融合方法与所提出方法比较图像融合能力。

3.4.1.　TNO 数据集结果分析

在图 7 中选取了 TNO 数据集中具有代表性的 6 对不同场景下的可见光图像和红外图像。利用目前主流的 9 种融合方法生成融合图像与本文提出的融合方法生成的图像进行比较。每张融合图像利用红色框和绿色框选取景物细节和红外显著目标并将放大的感兴趣区（Region Of Interest，ROI）拼接在原图下方（彩图见期刊电子版）。

通过 ROI 放大可以发现，DenseFuse 的融合图像对比度较低且细节信息模糊，红外显著区域没有突出显示，例如在场景 3 中，天上的云朵细节模糊不清，场景 5 和场景 6 中红外显著区域“人”不突出显示；RFN-Nest 保留了细节信息但是显著的红外目标信息缺失，例如场景 5 和场景 6 的“人”同样不能突出显示；SDNet 引入了大量噪声，导致一部分融合图像模糊不清，例如场景 3 和场景 4，噪声过大整体融合图像清晰度过低；SMOA 提供的细节纹理过少，例如场景 4 和场景



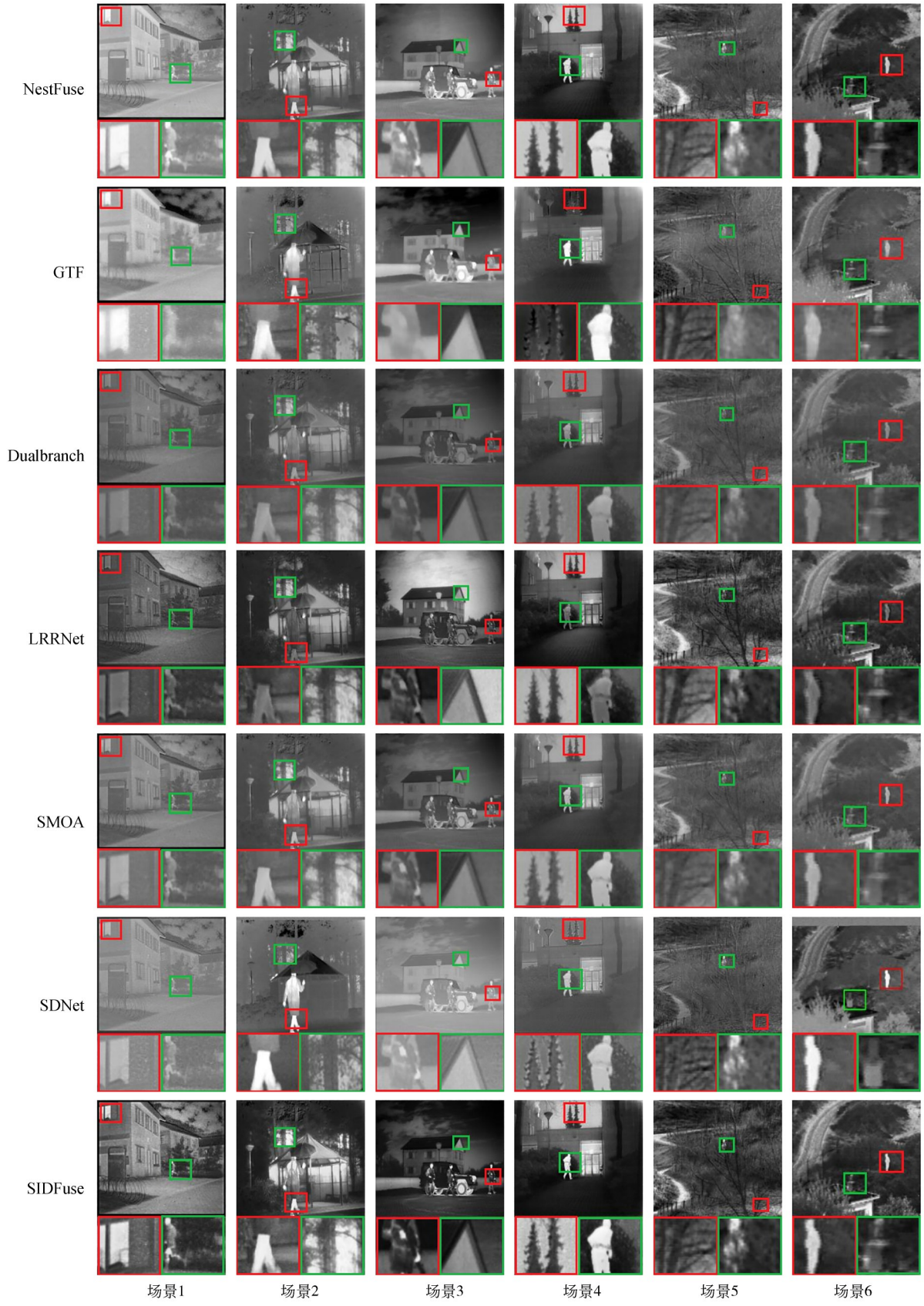


图 7　TNO 数据集上各种方法融合图像

Fig. 7　Various methods fuse images on TNO datasets

5融合图像中的松树和树枝的细节纹理均缺失。相比较，本文提出的融合方法图像对比度较高，可以保留可见光的细节纹理和红外图像的显著信息，例如场景2、场景4、场景5中“树”的细节纹理均能清晰呈现，例如场景2、场景5、场景6中红外显著区域的“人”均突出显示。

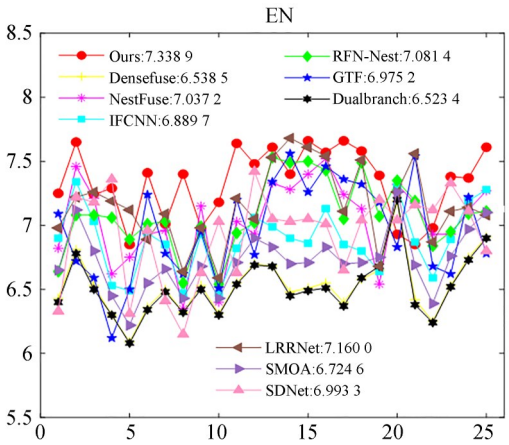
从客观评价指标分析，表3是10种融合方法在八种评价指标上的平均值，每个评价指标中最好的方法用加粗表示，次优用红色表示，第三名用蓝色表示（彩图见期刊电子版）。可以清晰地看出本文融合方法在EN，SD，SF，AG，SCD，MS-SSIM这6个客观评价指标上在这10种融合方法中排在第一名，仅评价指标VIF，Q_{abf}排在第三名。相较于DenseFuse在EN，SD，SF，VIF，AG，SCD，MS-SSIM，Q_{abf}指标上分别平均提高12.2%，9.0%，90.2%，13.9%，85.1%，16.8%，6.7%，30.7%；相较于最新的融合算法LRRNet在EN，SD，SF，VIF，AG，SCD，MS-SSIM，Q_{abf}指标上分别平均提高2.5%，5.6%，31.5%，5.4%，25.2%，17.9%，7.5%，20.7%。综上所述，本文提出的融合方法在主观评价和客观评价上均远超其他融合方法，在TNO数据集中本文的方法具有显著的优势。

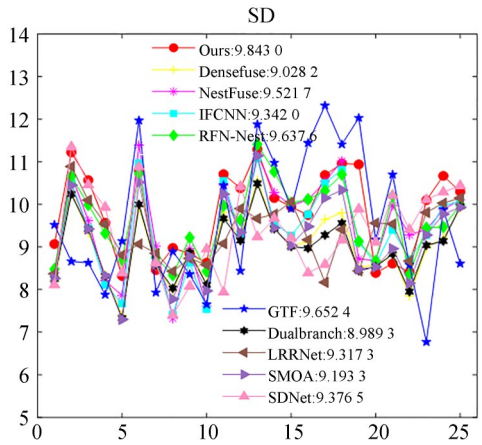
上述表3通过10种融合方法在八种评价指标上的平均值进行分析比较，但是由于融合图像可能被噪音影响或者融合结果产生畸变导致某几张图像的客观评价指标过高而使平均值过高，所以图8利用折线图绘制出每一张图上客观评价指标数值，横坐标表示第几张图，纵坐标表示方法在该评价指标上的数值，本文提出方法用红色表示。通过图8可以清晰地观察出在EN，AG，SCD，MS_SSIM这四个评价指标中本文提出方法的每一张融合图像的数值均位于第一名，在SD，SF，VIF，Q_{abf}中本文所提方法的数据均位于前三名。

表3　在TNO数据集上10种融合方法客观比较

Tab. 3　Ten fusion methods were objectively compared on the TNO dataset

	EN	SD	SF	VIF	AG	SCD	MS_SSIM	Q_{abf}
DenseFuse	6.538 5	9.028 2	0.024 6	0.667 9	2.446 7	1.633 6	0.872 4	0.331 3
RFN-Nest	7.081 4	9.637 6	0.021 7	0.792 4	2.495 8	1.817 6	0.906 1	0.327 9
IFCNN	6.889 7	9.342 0	0.046 6	0.784 7	4.497 8	1.699 3	0.909 3	0.504 2
NestFuse	7.037 2	9.521 7	0.036 3	0.863 3	3.482 2	1.729 8	0.880 1	**0.531 7**
GTF	6.975 2	9.652 4	0.032 9	0.606 8	3.116 4	0.980 9	0.809 8	0.319 4
Dualbranch	6.523 4	8.989 3	0.023 0	0.661 9	2.377 4	1.624 4	0.867 6	0.321 2
LRRNet	7.160 0	9.317 3	0.035 6	0.785 9	3.617 0	1.618 8	0.866 4	0.358 6
SMOA	6.724 6	9.193 3	0.025 1	0.720 8	2.461 7	1.715 0	0.885 7	0.306 1
SDNet	6.722 8	9.505 8	0.035 6	**0.966 2**	3.393 6	0.880 7	0.726 7	0.428 9
SIDFuse	**7.338 9**	**9.843 0**	**0.046 8**	0.828 0	**4.529 4**	**1.907 8**	**0.931 2**	0.432 9





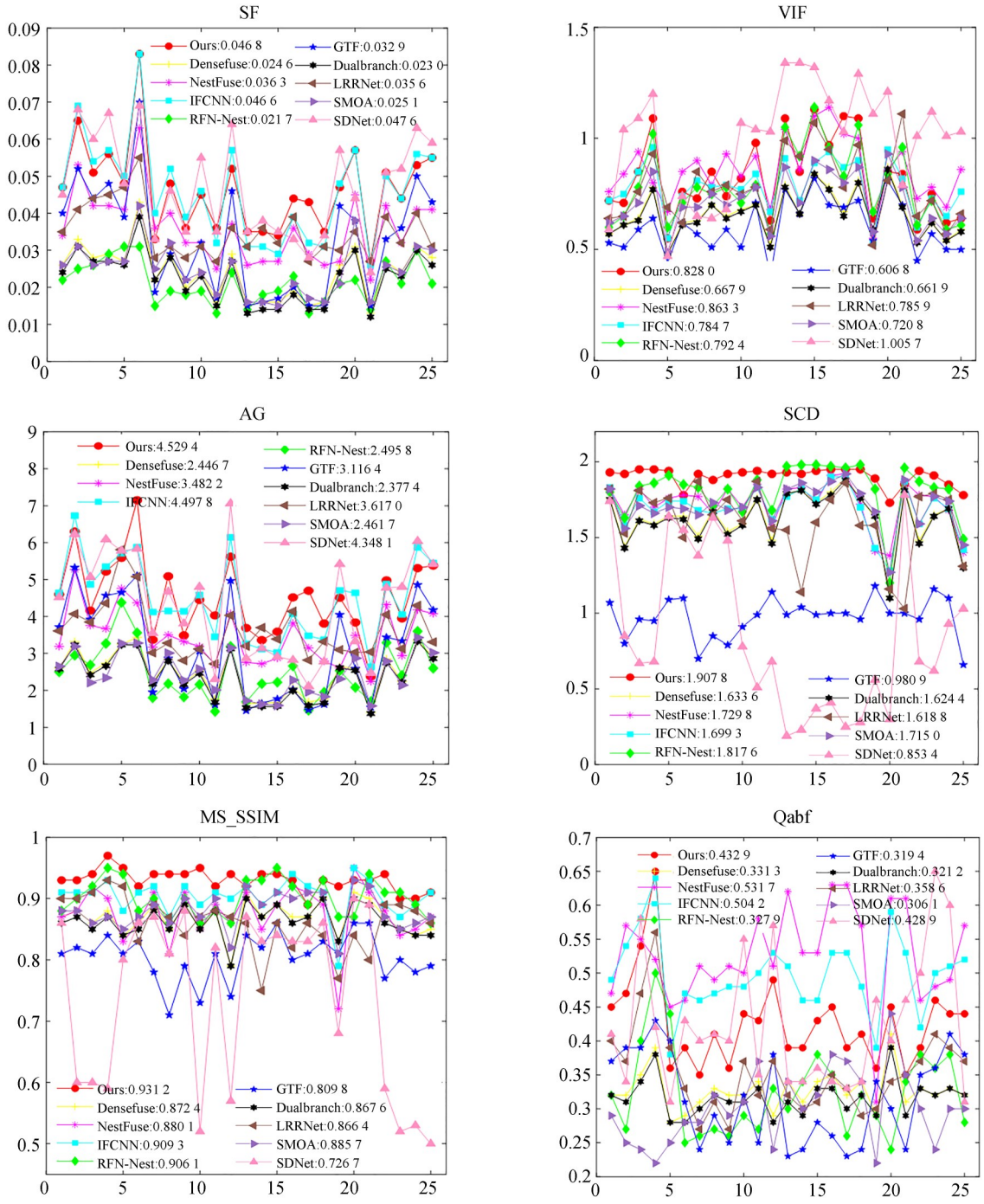


图 8　TNO 数据集客观评价指标折线图

Fig. 8　TNO data set objective evaluation indicators line chart

3.4.2　RoadScene 数据集结果分析

从 RoadScene 数据集中选取 40 对数据进行实验验证，图 9 是我们挑选的一对图像利用不同方法进行融合展示如下（彩图见期刊电子版）。

从图 9 可以明显看出，本文所提的融合方法中红外显著信息对比其他方法更加高亮清晰，红色框经过放大后的树枝可以清楚地看出纹理，DenseFuse，RFN-Nest 和 GTF 在图 9 中人的脚和滑板等红外显著信息显示模糊，而 SMOA 和 SDNet 对于可见光细节信息融合能力较差，对于 LRRNet 也可以清晰地表示树枝的纹理细节但是红外显著区域不能高亮显示，同时 LRRNet 的融

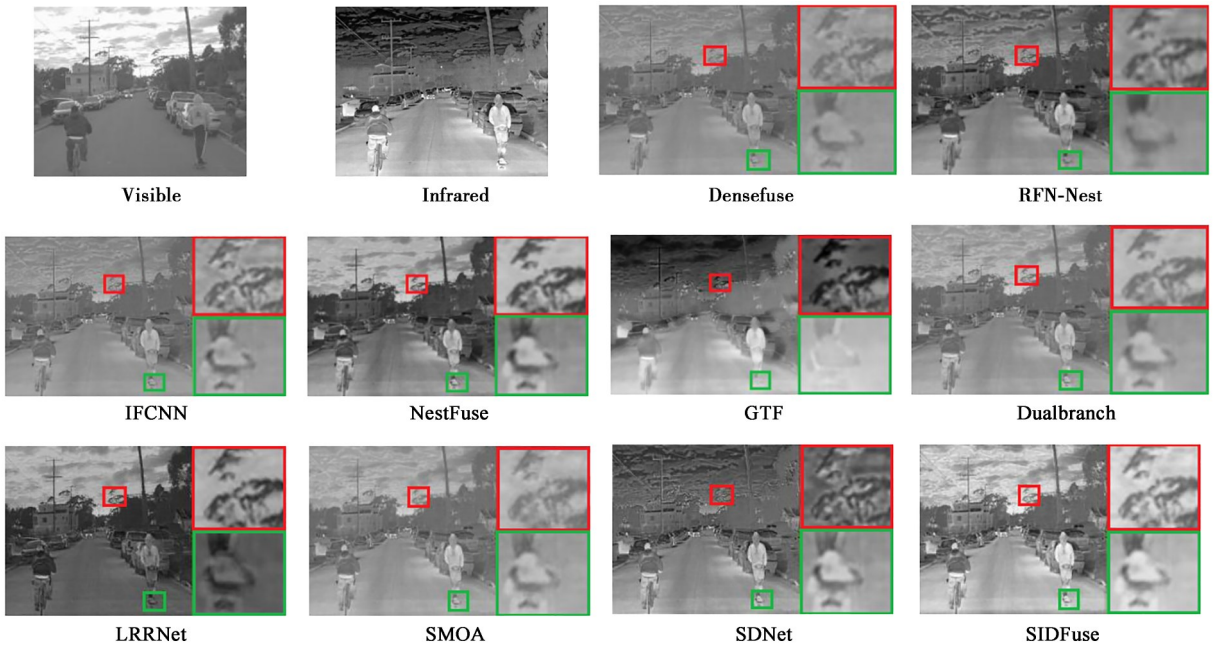


图 9　RoadScene 数据集融合结果分析

Fig. 9　Analysis of RoadScene data set fusion results

合图像偏向于可见光图像,不能很好地平衡可见光细节信息和红外显著信息。本文所提出的融合算法 SIDFuse 在细节信息的提取上优势明显,通过红色 ROI 框“树梢”可以明显观察到相较于其他经典融合算法,SIDFuse 细节提取能力更强,细节纹理更清晰。通过绿色 ROI 框观察红外显著信息“脚”可以看出,SIDFuse 融合后红外显著区域高亮且突出。

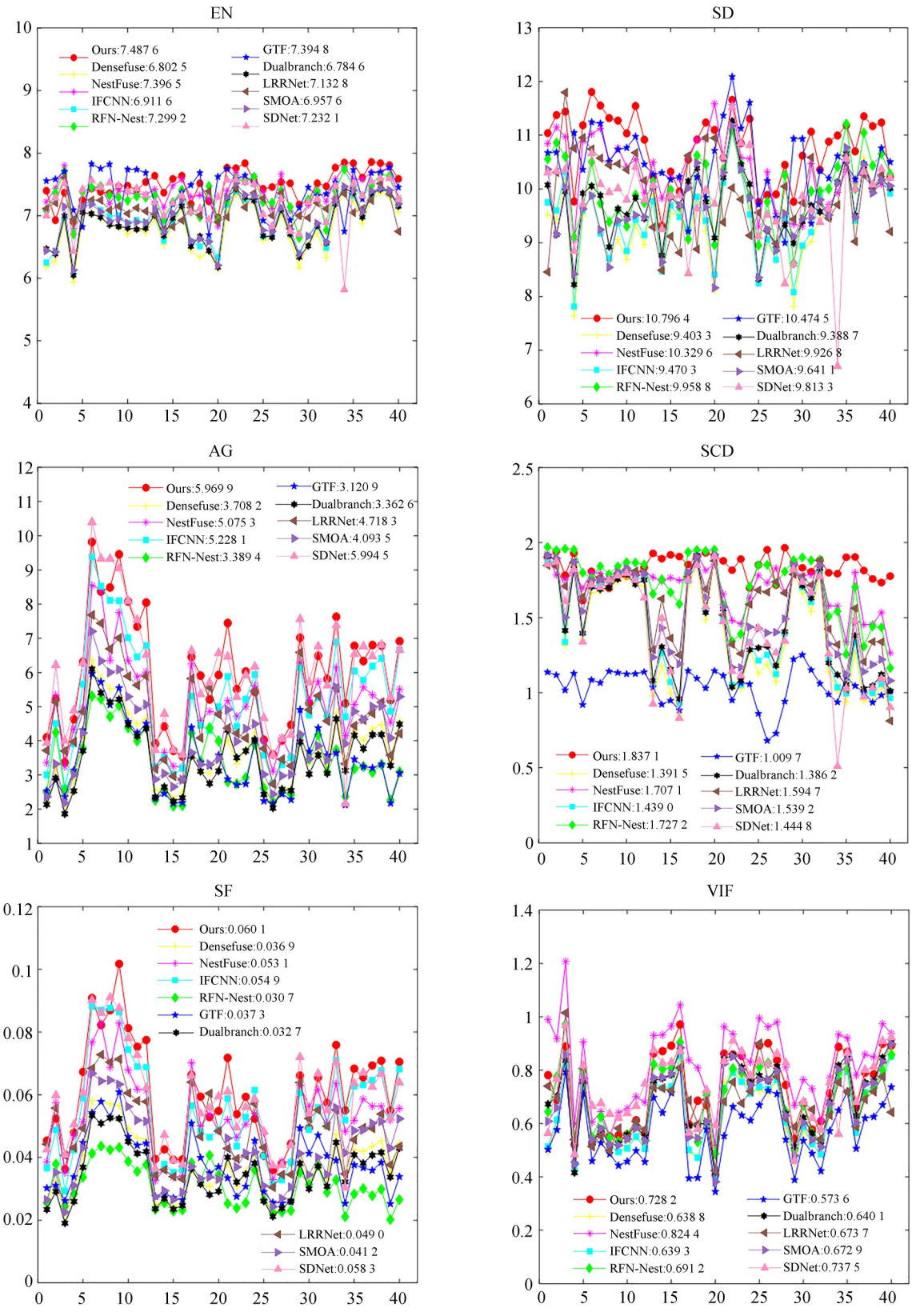
表 4 是对于 10 种不同融合方法在 RoadScene 数据集上的客观评价参数值,通过对比可见本文提出的 SIDFuse 融合方法在八个客观评价指标中,5 个参数位于第一名,AG 位于第二名,VIF 位于第三名、Q_{abf} 位于第四名。相较于经典算法 DenseFuse 在 EN,SD,SF,VIF,AG,SCD,MS-SSIM,Q_{abf} 指标上分别平均提高 10.1%,14.8%,62.9%,13.9%,60.9%,32.0%,7.5%,38.3%;相较于最新的融合算法 LRRNet 在 EN,SD,SF,VIF,AG,SCD,MS-SSIM,Q_{abf} 指标上分别平均提高 4.9%,8.8%,22.7%,8.1%,26.5%,15.2%,15.6%,40.3%。

表 4　在 RoadScene 数据集上 10 种融合方法客观比较

Tab. 4　Ten fusion methods were objectively compared on RoadScene dataset

	EN	SD	SF	VIF	AG	SCD	MS_SSIM	Q_{abf}
GTF	7.394 8	10.474 5	0.037 3	0.573 6	3.120 9	1.009 7	0.732 3	0.337 6
DenseFuse	6.802 5	9.403 3	0.036 9	0.638 8	3.708 2	1.391 5	0.848 1	0.353 1
RFN-Nest	7.299 2	9.958 8	0.030 7	0.691 2	3.389 4	1.727 2	0.863 9	0.292 6
IFCNN	6.911 6	9.470 3	0.054 9	0.639 3	5.228 1	1.439 0	0.900 1	**0.521 8**
NestFuse	7.396 5	10.329 6	0.053 1	**0.824 4**	5.075 3	1.707 1	0.865 0	0.502 2
Dualbranch	6.784 6	9.388 7	0.032 7	0.640 1	3.362 6	1.386 2	0.856 5	0.418 0
LRRnet	7.132 8	9.926 8	0.049 0	0.673 7	4.718 3	1.594 7	0.788 8	0.348 1
SMOA	6.957 6	9.641 1	0.041 2	0.672 9	4.093 5	1.539 2	0.877 0	0.432 2
SDNet	7.232 1	9.813 3	0.058 3	0.737 5	**5.994 5**	1.444 8	0.893 1	0.495 3
SIDFuse	**7.487 6**	**10.796 4**	**0.060 1**	0.728 2	5.969 9	**1.837 1**	**0.911 7**	0.488 4

图 10 是 RoadScene 数据集 40 张图像八种客观评价指标的折线图，每个点代表一张图像，我们的方法用红色表示(彩图见期刊电子版)。从图 10 中可以看出，本文方法在 SD，SF，SCD，MS-SSIM 这四个评价指标中每张融合图像基本均位于第一，EN，VIF，AG 这三个指标的每张融合图像指标均位于前三。



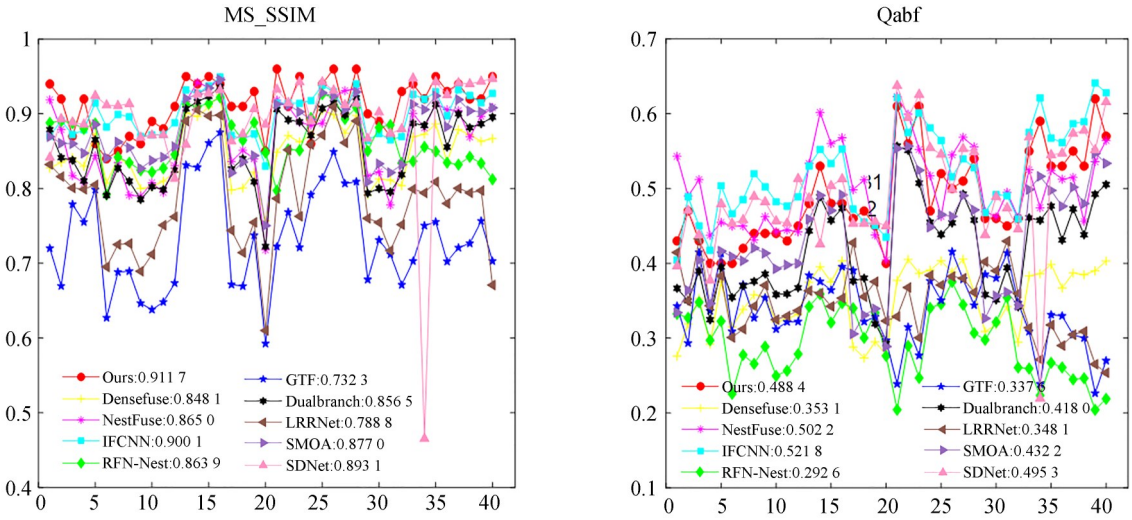


图 10　RoadScene 数据集客观评价指标折线图

Fig. 10　RoadScene data set objective evaluation index line chart

由此可以说明,本文所提出的融合方法每张图像的客观评价指标均优于其他经典算法,不是由于个别融合图像评价指标过高而导致平均值高。同时通过对于 RoadScene 数据集的融合实验分析,可以说明本文提出的 SIDFuse 鲁棒性较好,在不同的数据集上均具有优秀表现。

3.5　消融实验

为了验证本文提出方法中各模块的功能及是否有效,共设置 5 组消融实验,详情如下:

(a)编码器解码器均为普通卷积层且只进行一次图像分解;

(b)进行二次图像分解且编码器解码器均使用普通卷积层;

(c)进行二次图像分解且编码器使用背景提取残差模块 BERM 和细节提取残差模块 DERM;

(d)进行二次图像分解,编码器为背景提取残差模块 BERM 和细节提取残差模块 DERM,加入双元素注意力机制;

(e)本文所提完整融合模型结构。

表 5　TNO 数据集消融实验

Tab. 5　TNO data set ablation experiment

	EN	SD	SF	VIF	AG	SCD
实验 a	7.004 3	8.876 3	0.040 7	0.766 4	4.407 4	1.797 6
实验 b	7.032 9	8.964 1	0.044 7	0.776 2	4.175 1	1.815 0
实验 c	7.220 4	9.472 5	0.045 4	0.816 0	4.362 0	1.875 8
实验 d	7.256 2	9.715 2	**0.047 4**	0.821 6	**4.801 9**	1.901 4
实验 e	**7.338 9**	**9.843 0**	0.046 8	**0.828 0**	4.529 4	**1.907 8**

选取 TNO 数据集中的 25 对图像的融合结果进行消融实验的客观评价指标。通过表 5 的客观评价指标对比可以发现,实验(b)进行二次图像分解分离的融合结果明显比实验(a)只进行一次细节背景分离的客观评价指标更好,除了 AG 其他客观评价指标均有明显提升;实验(c)引入密集残差模块各项指标均有了较大的提高,其中 EN,SD 提升最为明显,表明从源图像提取的信息更加丰富;实验(d)和实验(e)分别依次加入双元素注意力模块和第三条全局语义模块相较于实验(c)六个客观评价指标均提升明显,说明双元素注意力模块和第三条语义支路会使本文提出的融合算法在性能上有重大提升。

图 11 是对 TNO 数据集中一种典型场景对

(a)～(e)五种消融实验的结果,消融实验结果可以从主观视觉角度发现实验(a)采用普通卷积层提取信息且只进行一次图像分解,编码器提取的特征信息过少,实验(c)在引入细节提取残差模块 DERM 且进行两次图像分解后,融合结果表现出更加丰富的细节纹理,对比实验(a)和实验(b),实验(c)在加入 DERM 后细节特征更加丰富如“树枝”、“铁丝网”,实验(d)融合图对比实验(b)引入双元素注意力机制,使得实验(d)能够自适应的平衡二维图像分解得到的特征信息,因此在融合图中的红外信息“人”的轮廓更加清晰,实验(e)是采用本文完整模型的融合结果,对比实验(a)～(d)在细节纹理,红外显著信息和图像对比度上均有较大提升。

图 11　消融实验结果

Fig. 11　Ablation results

4　结　论

本文提出了一种将红外图像和可见光图像进行二次图像分解的融合方法,在网络模型的编码和解码部分,通过 DERM、BERM 残差模块,解决了现存算法对于源图像细节信息提取不足的问题,同时通过双元素注意力机制自适应地平衡二次提取的粗尺度信息和细尺度信息,并且为了防止在残差网络信息提取时产生信息丢失,引入第三条全局语义信息支路,避免了对语义信息的忽略。我们提出改进的像素相加法融合策略,该融合策略可以使解码后的图像最大程度地保留可见光和红外信息。在 TNO 数据集中融合图像的客观评价指标 EN, SD, SF, VIF, AG, SCD 均值分别为 7.338 9, 9.843 0, 0.046 8, 0.828 0, 4.529 4, 1.907 8;在 RoadScene 数据集中 EN、SD, SF, VIF、AG, SCD 均值分别为 7.487 6, 10.796 4, 0.060 1, 0.728 2, 5.969 9, 1.837 1,同时鉴于对融合效率的考虑,本文选择了最佳的二次图像分解,由此本文模型在上述硬件平台对 TNO 数据集中单张运行时间为 0.25 s,可以实现快速融合处理。与其他几种主流算法相比,经 SIDFuse 融合的融合图像在视觉效果和评价指标均具有优势,能够有效保留可见光与红外图像信息。此外,由于融合由于光照不均匀的微光图像会影响融合效果,后续工作将针对微光图像的光照均匀性在此模型的基础上继续完善。

参考文献:

[1] MA J Y, MA Y, LI C. Infrared and visible image fusion methods and applications: a survey[J]. *Information Fusion*, 2019, 45: 153-178.

[2] LI S T, YANG B, HU J W. Performance comparison of different multi-resolution transforms for image fusion [J]. *Information Fusion*, 2011, 12 (2): 74-84.

[3] ZONG J J, QIU T. Medical image fusion based on sparse representation of classified image patches[J]. *Biomed Signal Process Control*, 2017, 34: 195-205.

[4] PATIL U, MUDENGUDI U. Image fusion using hierarchical PCA [C]. 2011 *International Conference on Image Information Processing. Shimla, India.* IEEE, 2011: 1-6.

[5] ZHANG X Y, MA Y, FAN F, *et al*. Infrared and visible image fusion via saliency analysis and local edge-preserving multi-scale decomposition[J]. *Journal of the Optical Society of America A, Optics, Image Science, and Vision*, 2017, 34(8): 1400-1410.

[6] LI H, WU X J. DenseFuse: a fusion approach to infrared and visible images[J]. *IEEE Transactions on Image Processing*, 2018: 2614-2623.

[7] LI H, WU X J, DURRANI T. NestFuse: an infrared and visible image fusion architecture based on nest connection and spatial/channel attention models [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(12): 9645-9656.

[8] LI H, WU X J, KITTLER J. RFN-Nest: an end-to-end residual fusion network for infrared and visible images[J]. *Information Fusion*, 2021, 73(C): 72-86.

[9] ZHANG Y, LIU Y, SUN P, *et al*. IFCNN: a general image fusion framework based on convolutional neural network[J]. *Information Fusion*, 2020, 54: 99-118.

[10] ZHANG H, MA J Y. SDNet: a versatile squeeze-and-decomposition network for real-time image fusion [J]. *International Journal of Computer Vision*, 2021, 129(10): 2761-2785.

[11] LIU J Y, WU Y H, HUANG Z B, *et al*. SMoA: searching a modality-oriented architecture for infrared and visible image fusion[J]. *IEEE Signal Processing Letters*, 2021, 28: 1818-1822.

[12] LI H, XU T Y, WU X J, *et al*. LRRNet: a novel representation learning guided fusion network for infrared and visible images [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(9): 11040-11052.

[13] MA J Y, TANG L F, XU M L, *et al*. STDFusionNet: an infrared and visible image fusion network based on salient target detection [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5009513.

[14] REN C, HE X H, WANG C C, *et al*. Adaptive consistency prior based deep network for image denoising[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). *Nashville*, *TN*, *USA*. IEEE, 2021: 8592-8602.

[15] KINGMA DP, BA J. Adam: A method for stochastic optimization [C]. *International Conference on Learning Representations* (*ICLR*), 2015: 1-15.

[16] TOET A, HOGERVORST MA. Progress in color night vision[J]. *Optical Engineering*, 2012, 51 (1): 010901.

[17] XU H, MA J Y, JIANG J J, *et al*. U2Fusion: a unified unsupervised image fusion network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 502-518.

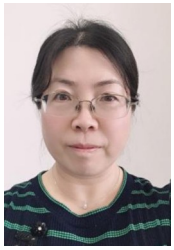
[18] MA J Y, CHEN C, LI C, *et al*. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. *Information Fusion*, 2016, 31(C): 100-109.

[19] FU Y, WU X J. A dual-branch network for infrared and visible image fusion[C]. 2020 25*th International Conference on Pattern Recognition* (*ICPR*). *Milan*, *Italy*. IEEE, 2021: 10675-10680.

作者简介：

马　鑫(1999－)，男，陕西西安人，硕士研究生，2021年于西安邮电大学获得学士学位，研究方向是光电成像与图像处理。E-mail：mx18229015396@163.com

通讯作者：



喻春雨(1976－)，女，辽宁沈阳人，副教授，2006年于南京理工大学获光学工程博士学位，2006～2011年在北京大学深圳研究生院信息工程学院工作，2011年至今在南京邮电大学电子与光学工程学院、柔性电子(未来技术)学院工作，其中2014～2015年在宾夕法尼亚大学医学院访学。E-mail：yucy@njupt.edu.cn