DOI:10.11918/202509025

泰勒展开与复合注意力引导的红外与可见光图像融合

杨艳春,李 毅

(兰州交通大学 电子与信息工程学院,兰州 730070)

摘 要:为解决深度学习融合算法中存在的忽略像素间相关性,导致融合结果丢失重要全局纹理,以及难以平衡目标突出与场景增强的问题,本文提出了一种泰勒展开与复合注意力机制引导的红外与可见光图像融合算法。首先,设计了一种泰勒展开网络,将输入图像分解为映射层与导数层,从而实现对图像多层次特征信息的有效提取;其次,采用双分支特征提取网络,其中平行卷积网络负责捕获局部细节特征,SwinTransformer模块则专注于提取全局上下文信息,确保局部与全局特征的高效保留;再次,引入复合注意力机制来进一步提升特征融合的精度,该机制通过轴向注意力融合空间维度特征,同时利用通道注意力强化通道间的特征响应,以实现更精细的特征选择与融合。最后,通过图像重建得到融合图像。在公开数据集MSRS和RoadScene进行了相关实验,结果表明,本文方法融合图像不仅在纹理细节保持与全局信息保留方面更完整,而且在客观指标中取得显著优势。该研究结果可为深度学习图像融合领域提供新的思路。

关键词:红外与可见光图像融合;泰勒展开网络;SwinTransformer;双分支特征提取;复合注意力机制

中图分类号:TP391;TN29 **文献标志码**:A **文章编号**:0367-6234(2026)05-0054-09

Infrared and visible image fusion guided by Taylor expansion and composite attention

YANG Yanchun, LI Yi

(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: In order to solve the problems of ignoring the correlation between pixels in the deep learning fusion algorithm, which leads to the loss of important global texture in the fusion results, and the difficulty of balancing target highlight and scene enhancement, this paper proposed an infrared and visible image fusion algorithm guided by Taylor expansion and composite attention mechanism. Firstly, a Taylor expansion network was designed to decomposition the input image into a mapping layer and a derivative layer, so as to effectively extract the multi-level feature information of the image. Secondly, a dual-branch feature extraction network was used, in which the parallel convolutional network was responsible for capturing local detail features, and the SwinTransformer module focused on extracting global context information to ensure the efficient retention of local and global features. Then, the composite attention mechanism is introduced to further improve the accuracy of feature fusion. This mechanism fuses spatial dimensional features through axial attention, and uses channel attention to strengthen the feature response between channels, so as to achieve more refined feature selection and fusion. Finally, the fused image was obtained by image reconstruction. Experiments are carried out on the public datasets MSRS and RoadScene. The results show that the proposed method is not only more complete in maintaining texture details and global information, but also achieves significant advantages in objective indicators. The research results can provide new ideas for the field of deep learning image fusion.

Keywords: infrared and visible image fusion; Taylor expansion network; SwinTransformer; dual-branch feature extraction; compound attention mechanism

近年来,红外与可见光图像融合技术在计算机视觉领域受到了广泛关注,其目标是通过整合同一场景下多模态图像的信息,生成表达更全面、准确的场景表示。红外传感器通过捕捉场景中的热辐射信息,能够有效反映目标物体的热特征,即使在低光照、遮挡或目标隐藏等复杂条件下,仍可稳定检测出人、车辆等红外目标。然而,红外图像通常空间分辨率较低,缺乏丰富的纹理细节[1]。相比之下,可见

收稿日期:2025-09-07;**录用日期**:2025-10-13;**网络首发日期**:2025-11-14
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20251113.1517.004
基金项目:国家自然科学基金(62462043,62067006);甘肃省重点研发计划(25YFGA047);甘肃省自然科学基金(23JRRA847,21JR7RA300)
作者简介:杨艳春(1979—),女,副教授,硕士生导师
通信作者:李 毅,1544726016@qq.com

光图像依靠物体反射光成像，具有较高的空间分辨率和清晰的纹理信息，更符合人眼的视觉感知习惯，但其成像质量容易受到雨、雾等环境因素的干扰，稳定性较差。红外与可见光图像融合通过集成来自同一场景中多个传感器的互补信息，生成更符合人类感知或计算机处理需求的融合结果，提升图像信息的可用性与可理解性，目前，这种融合技术被广泛应用于军事侦察、遥感与目标检测等领域[2-4]。

当前的图像融合技术主要分为传统方法和基于深度学习的融合方法两大类。传统方法依赖于图像处理工具提取源图像的多类特征，例如多尺度变换[5]、稀疏表示[6]、子空间法[7]、显著性检测[8]等，随后通过对这些特征进行逆变换生成融合图像。尽管传统方法在一定程度上是有效的，但也存在一些局限性，其像素活动水平的计算和融合规则的设计通常依赖于人工经验，难以应对复杂多变的图像场景。

深度学习方法具有强大的特征提取能力和解决非线性过拟合的能力，能够提高网络模型的泛化能力。常见的深度学习的方法包括基于自动编码器（auto-encoder，AE）、基于卷积神经网络（convolutional neural network，CNN）、基于生成对抗网络（generative adversarial network，GAN）和基于 Transformer 方法。在 AE 方法中，Li 等[9]通过注意力模型突出空间和通道的重要性，增强了多尺度深度特征融合；Zhao 等[10]提出一种自监督特征自适应融合方法，通过注意力机制解码重建源图像以保留关键信息，并引入图像增强模块提升低质场景下的融合鲁棒性；Zhao 等[11]通过算法展开将优化模型转化为可训练网络，实现两尺度特征分离与融合，在测试阶段通过专用融合层与解码器生成优越的融合图像。但 AE 方法通常需要手工设计融合策略，而 CNN 方法通过其层次化的卷积结构，能够从数据中自动学习出最优的融合规则。在 CNN 方法中，Li 等[12]创新性地融合了 CNN 与图神经网络，首先通过 CNN 提取多尺度特征，再借助图交互模块将特征转换为图结构，以实现有效的跨模态融合；Yue 等[13]通过多通道扩散特征与专用损失函数，在提升纹理与强度保真度的同时，显著提高了色彩还原能力；Yang 等[14]利用交叉卷积模块提取多向细节，结合局部显著性注意力网络生成权重图优化特征融合，并通过自适应像素损失提升训练效果。但在 CNN 方法中由于网络模型深度的增加，源图像的细节不可避免地会丢失。GAN 方法通过鉴别器与生成器对抗博弈，有效加强了对图像细节的保留能力。在 GAN 方法中，Zhou 等[15]提出语义监督双鉴别器生成对抗网络，通过信息量判别模块和双鉴别器机制，实现红外与可见光图像的高层语义感知融合，有效保留热辐射与纹理细节特征；Yin 等[16]在生成器中采用金字塔分解路径与剩余注意力融合规则整合多尺度特征，结合跨尺度金字塔注意力模块及双鉴别器优化分布逼近，实现多级特征高效解码与重构；Wang 等[17]通过小波变换分解多频带特征，结合混合频率聚合模块和双重鉴别器，有效解决传统方法忽略频域信息的问题，显著提升多模态图像融合质量。但 GAN 方法忽略了图像中远程相关性的提取，导致融合结果中丢失了重要的全局纹理。Transformer 方法通过自注意力机制有效建模长程依赖，弥补了这一缺陷。在 Transformer 方法中，Yi 等[18]开发语义文本网络框架，集成文本语义编码器与融合解码器，利用自然语言描述引导多模态特征对齐，实现交互式可控融合；Liu 等[19]创新性引入视觉语言模型的提示学习范式，通过语义提示向量动态调节融合权重，同步优化目标识别置信度与视觉质量指标；Tang 等[20]提出 Y 形动态 Transformer 融合方法，通过动态转换器模块同步提取局部特征与上下文信息，利用 Y 形网络结构协同保持红外热辐射与可见光纹理细节，显著提升融合效果。

上述深度学习融合算法虽然取得一定的融合效果，但这些方法忽略了像素之间的相关性，致使融合结果中丢失了重要的全局纹理，难以兼顾目标凸显与场景整体感知的平衡。针对上述问题，本文提出了一种泰勒展开与复合注意力的红外与可见光图像融合算法，引入泰勒展开网络与损失函数，以建模像素间深层次相关性；设计双路径特征提取网络，确保全局重要纹理不丢失；构建复合注意力融合网络，实现特征的精细筛选与保留。

1　本文方法

1.1　网络框架

本文提出的算法总体结构如图 1 所示。该网络架构由泰勒展开网络、特征提取、特征融合和特征重建 4 部分组成。首先，输入红外图像和可见光图像，通过泰勒展开网络将输入图像分解成为基础映射层与导数层，从而得到多层次的深度特征图。其次，将得到的特征图输入到由 PCN 和 SwinTransformer 的双分支特征提取网络来增强特征的表征能力，提高图像特征的局部与全局信息。最后，利用复合注意力进行特征融合，经过图像重建得到融合图像。

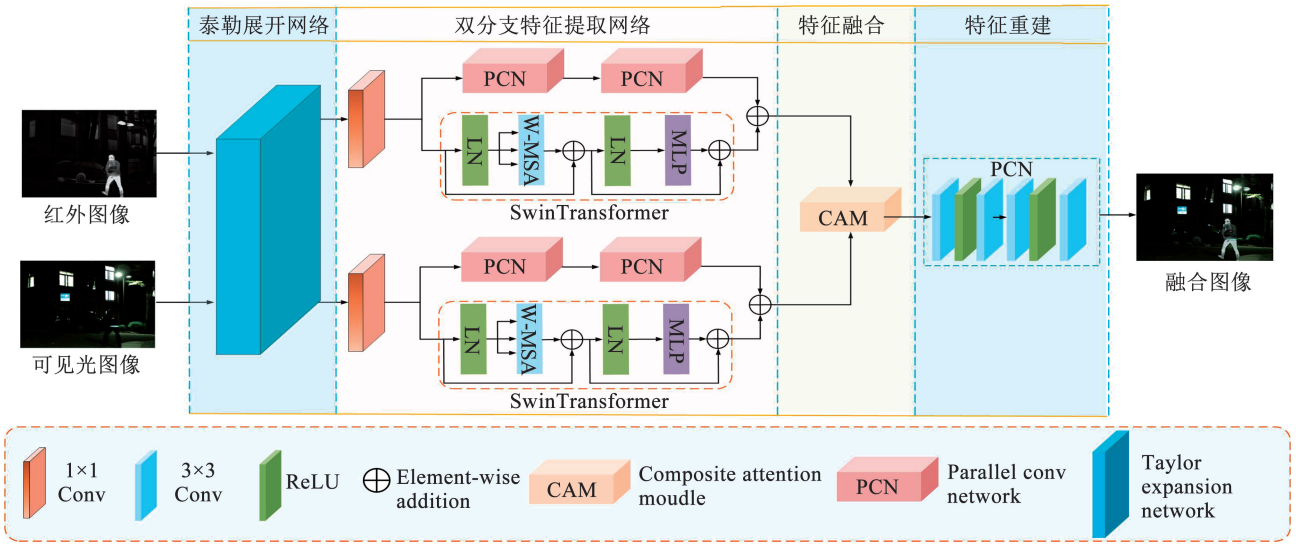


图 1 整体网络框架

Fig. 1 Overall network framework

1.2 泰勒展开网络

泰勒展开作为一种重要的数学工具，已广泛应用于计算机视觉任务中，根据泰勒展开定理，当$f(x)$ $(x\in(a,b))$具有 n 阶导数时，其可以在 $x_0(x_0\in(a,b))$处展开如下：

$$f(x)=f(x_0)+\frac{f'(x_0)}{1!}(x-x_0)+\cdots+\frac{f^{(n)}(x_0)}{n!}(x-x_0)^n=f(x_0)+\sum_{k=1}^{n}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k \tag{1}$$

式中：$f(x)$、$f(x_0)$和 $\sum_{k=1}^{n}\frac{f^{(k)}(x_0)}{k!}(x-x_0)^k$ 分别为源可见光图像或红外图像、基本大尺度特征图和导数特征。从式(1)可以看出，其各层次之间的特征映射具有密切的相关性，这将有助于网络将全局多结构从原始图像转移到融合结果。为了充分提取图像特征图，使用一个映射网络来近似基础层，同时使用一个共享的导数网络来处理导数特征。

在公式(1)中，泰勒展开网络(Taylor expansion network，TEN)需要估计的主要有两个部分，包括基础层$f(x_0)$及其导数层如$(f^k(x_0),k\in[1,n])$。为了充分提取其特征，采用映射网络 $\phi_b(\cdot)$来计算基础层(可以表示为$f(x_0)=\phi_b(f(x))$和一个导数网络 $\phi_d(\cdot)$来处理每个导数分量。由式(1)可知，其邻域导数分量具有密切的相关性。除后者是前者的导数外，还有附加项$(x-x_0)$，其与$f(x)$和$f(x_0)$有关。将$f(x)$和$f^k(x_0)$，$k\in[1,n]$连接到 ϕ_d 中作为推理的输入，可以表示为

$$f(x)=\phi_b(f(x))+\sum_{k=1}^{n}\frac{\phi_d(\mathrm{Concat}(f^{(k-1)}(x_0),f(x)))}{k!} \tag{2}$$

1.2.1 映射网络

映射网络主要用于从图像中提取全局信息，以防止大规模细节的丢失，可以看出，其是由 1 个核大小为 5 × 5 的卷积层和 3 个 ResBlock 组成的。ResBlock 的具体框架如图 2(a)所示，包括 3 个核大小为 1 × 1 的卷积层、1 个核大小为 3 × 3 的卷积层，以及每个卷积层后的 4 个 LReLU 层以增强网络稳定性，同时保持梯度活性与非线性。

1.2.2 导数网络

导数网络由两个卷积层组成，包括 5 × 5 的卷积层和 3 个 ResDBlock。ResDBlock 的具体框架如图 2(b)所示。为了在特征提取过程中保留尽可能多的精细特征，ResDBlock 中加入了密集连接和 Mish 激活函数。

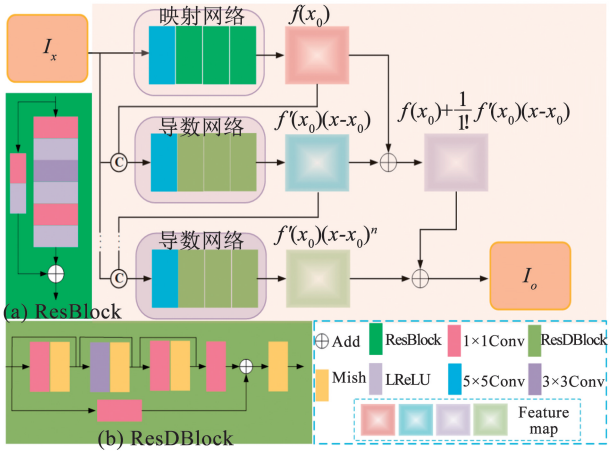


图 2 泰勒展开网络

Fig. 2 Taylor expansion network

1.3 双分支特征提取网络

如图 1 所示，双分支特征提取网络包括 1 × 1 卷积层、SwinTransformer(ST)和平行卷积网络(parallel

convolutional network，PCN）。ST 主要包含两个阶段：基于窗口的多头自注意（window-based multi-head self-attention，W-MSA）和多层感知器（multi-layer perceptron，MLP）。ST 模块通过其独特的双阶段结构有效整合多模态图像特征，W-MSA 阶段用于捕捉输入图像序列中的依赖关系，使模型能够关注图像序列中的不同位置，实现模态间特征的动态交互与增强。MLP 阶段则通过非线性变换对融合后的特征进行深度加工，学习跨模态的高阶特征表示，从而提升融合图像对原始双模态信息的保留能力。为优化特征融合过程，每个阶段前均引入 LayerNorm（LN）层进行特征分布标准化，确保多模态输入在统一特征空间中进行对齐。PCN 网络由 3×3 卷积层和 ReLU 构成，对于红外图像分支，输入特征被送到 ST 中以捕获长范围全局特征依赖性，而由 PCN 处理以学习局部特征依赖性。可见光图像分支以类似的方式生成。

1.4　复合注意力融合机制

为了从双分支特征提取网络中融合特征，本文设计了一种新的复合注意力融合策略（compound attention，CA），以进一步提高通用多模态图像融合模型的融合精度。具体的 CA 模块如图 3 所示。复合模块主要由两个自注意模块、一个通道注意模块和辅助连接模块组成。轴向注意模块可以提取特征的空间信息。通道注意模块进一步实现特征沿通道方向的适应性，平衡 CNN 支路特征和 Transformer 支路特征。辅助连接模块融合轴向注意信息和通道注意信息，最终实现全局信息和局部信息的整合。轴向注意模块具有全局性特点，其体现在高度和宽度轴同时使用上。对于多模态特征映射，在特征的宽度轴和高度轴上定义了具有位置敏感和可调参数的轴向注意层。轴向注意层采用多头注意机制，利用轴向注意模块，通过输入多模态表征 $\boldsymbol{r}_1$ 和 $\boldsymbol{r}_2$，得到空间增强的中间表征 $\boldsymbol{f}_1$ 和 $\boldsymbol{f}_2$。通道注意力模块的计算过程如式(3)～式(5)所示。在图 3 中，通过两个中间层特征的输入，通道注意力机制的最终输出结果可由式(5)计算。

$$\boldsymbol{a}_1 = o[\mathrm{pool}(\boldsymbol{f}_1)] \tag{3}$$

$$\boldsymbol{a}_2 = o[\mathrm{pool}(\boldsymbol{f}_2)] \tag{4}$$

$$\tilde{\boldsymbol{f}}_{\mathrm{f}} = \xi(\boldsymbol{a}_1)\cdot\boldsymbol{f}_1 + \xi(\boldsymbol{a}_2)\cdot\boldsymbol{f}_2 \tag{5}$$

式中：$\boldsymbol{a}_1$，$\boldsymbol{a}_2$ 为归一化特征结果；$\tilde{\boldsymbol{f}}_{\mathrm{f}}$ 为通道注意力机制的最终输出特征；pool 为全局池化操作，本文使用最大池化；o 为归一化算子；ξ 为维数恢复操作。复合注意力融合模块的最终总体策略可由式(6)计算得出。

$$\boldsymbol{r} = 0.5\times[(\boldsymbol{f}_1+\boldsymbol{f}_2)+\tilde{\boldsymbol{f}}_{\mathrm{f}}] \tag{6}$$

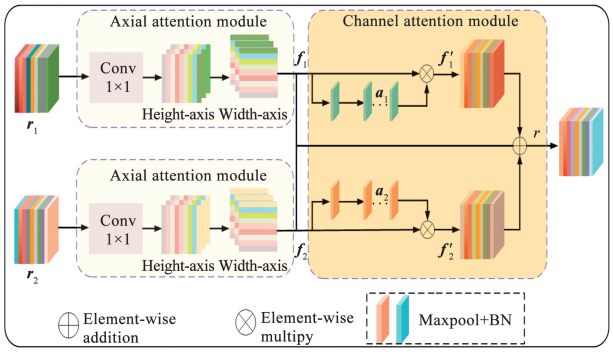


图 3　复合注意力

Fig. 3　Compound attention

1.5　损失函数

1.5.1　泰勒损失

为了得到期望的泰勒展开特征映射，本文采用泰勒损失对泰勒展开网络进行优化，定义为：

$$L_{\mathrm{T}} = L_{\mathrm{pixel}} + \lambda L_{\mathrm{grad}} \tag{7}$$

式中：L_{pixel} 和 L_{grad} 分别为像素损失和梯度细节损失；λ 为约束其幅度差的超参数。像素损失 L_{pixel} 在像素级度量两幅图像之间的差异，定义为：

$$L_{\mathrm{pixel}} = \frac{1}{HW}\|\boldsymbol{I}_x - \boldsymbol{I}_o\|_1 \tag{8}$$

式中：$\boldsymbol{I}_x$ 和 $\boldsymbol{I}_o$ 分别表示泰勒展开网络的输入图像和输出图像。H 为图像的高度，W 为图像的宽度，$\|\cdot\|_1$ 表示 L_1 范数。梯度细节损失定义为：

$$L_{\mathrm{grad}} = \frac{1}{HW}\|\,|\nabla\boldsymbol{I}_o| - \max(|\nabla\boldsymbol{I}_x|,|\nabla\boldsymbol{I}_n|)\,\|_1 \tag{9}$$

式中：∇、$|\cdot|$ 和 $\max(\cdot)$ 分别表示 Sobel 操作、绝对值操作和最大值操作，$\boldsymbol{I}_n$ 表示由导数网络分解的特征图。

1.5.2　融合损失

本文利用 MS-SSIM 损失函数 $L_{\mathrm{ssim}}^{\mathrm{ms}}$ 来保持融合图像更清晰的强度分布，定义为：

$$L_{\mathrm{ssim}}^{\mathrm{ms}} = (1-\mathrm{SSIM}(\boldsymbol{I}_{\mathrm{fus}},\boldsymbol{I}_{\mathrm{ir}})) + (1-\mathrm{SSIM}(\boldsymbol{I}_{\mathrm{fus}},\boldsymbol{I}_{\mathrm{vis}})) \tag{10}$$

式中，$\boldsymbol{I}_{\mathrm{ir}}$、$\boldsymbol{I}_{\mathrm{vis}}$、$\boldsymbol{I}_{\mathrm{fus}}$ 分别为红外图像、可见光图像和融合图像，为了促进纹理细节的恢复，本文建立了梯度分布损失，定义为：

$$L_{\mathrm{JGrad}} = \|\max(\nabla\boldsymbol{I}_{\mathrm{ir}},\nabla\boldsymbol{I}_{\mathrm{vis}}),\nabla\boldsymbol{I}_{\mathrm{fus}}\|_1 \tag{11}$$

为使融合图像具有更清晰的纹理细节，融合图像的梯度被强制接近红外图像和可见光图像梯度之间的最大值。此外，为了保留两幅图像的显著性目标，使用自适应视觉显著性图来构建新的损失 L_{svs}：

$$\boldsymbol{\omega}_{\mathrm{ir}} = \boldsymbol{S}_{\mathrm{fir}}/(\boldsymbol{S}_{\mathrm{fir}}-\boldsymbol{S}_{\mathrm{fvis}}) \tag{12}$$

$$\boldsymbol{\omega}_{\mathrm{vis}} = 1-\boldsymbol{\omega}_{\mathrm{ir}} \tag{13}$$

$$L_{\mathrm{svs}} = \|(\boldsymbol{\omega}_{\mathrm{ir}}\otimes\boldsymbol{I}_{\mathrm{ir}} + \boldsymbol{\omega}_{\mathrm{vis}}\otimes\boldsymbol{I}_{\mathrm{vis}}),\boldsymbol{I}_{\mathrm{fus}}\|_1 \tag{14}$$

式中：$\boldsymbol{S}$ 为显著性矩阵，$\boldsymbol{\omega}_{ir}$ 和 $\boldsymbol{\omega}_{vis}$ 分别为红外图像和可见光图像的权重图。总融合损失计算公式为：

$$L_{fus}=\lambda_{ssim}L_{ssim}^{ms}+\lambda_{JG}L_{JGrad}+\lambda_{svs}L_{svs} \tag{15}$$

式中，λ_{ssim}、λ_{JG} 和 λ_{svs} 为权重系数，分别设置为 1.0、20.0 和 5.0 来计算总融合损失。

2 实验结果与分析

2.1 实验设置

本实验框架基于 PyTorch 实现。在训练阶段，使用 MSRS 数据集中的训练集进行模型训练，并采用其验证集进行模型验证与调参。为提升数据多样性，对训练图像进行随机裁剪，生成 256 × 256 的图像块，每个训练批次包含 8 个图像块。测试阶段使用 RoadScene 数据集中的 121 幅图像进行性能评估。实验在 NVIDIA GeForce RTX 4060 GPU 上进行。优化器选用 Adam，初始学习率设为 0.001，训练共进行 300 个周期，且学习率保持不变。

2.2 对比实验与分析

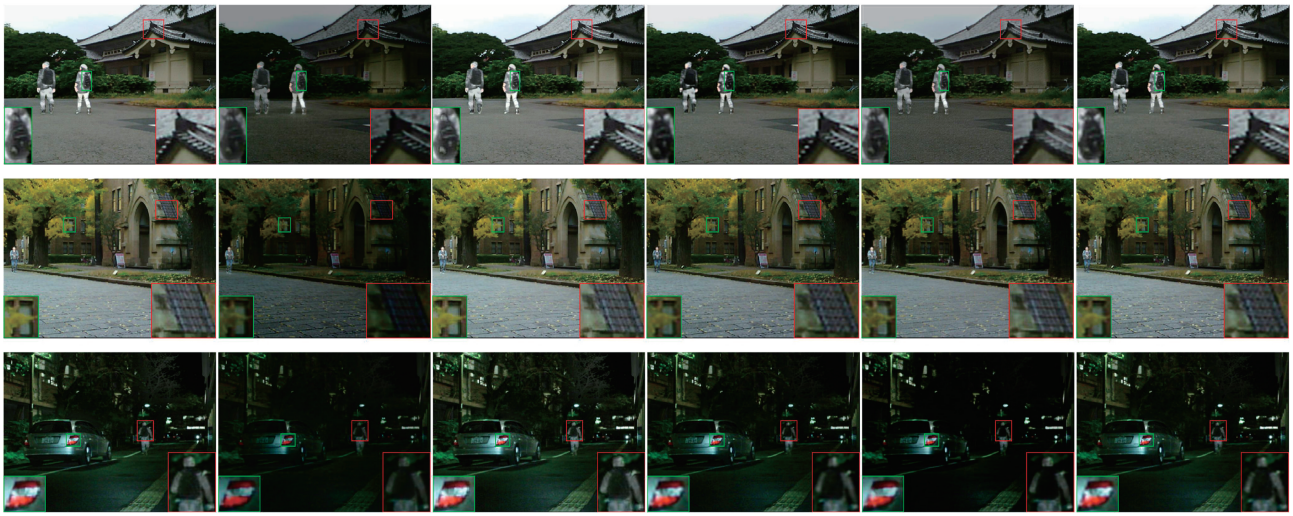
本文选取了 9 种具有代表性的先进融合方法进行对比研究，包括 DenseFuse[21]、FusionGAN[22]、GANMcC[23]、PMGI[24]、SeAFusion[25]、U2Fusion[26]、SDCFusion[27]、IMF[28] 和 T2EA[29]。为客观评价融合效果，实验采用了 6 个质量评价指标：空间频率 F_S、平均梯度 G_A、互信息 I_M、视觉信息保真度 F_{VI}、峰值信噪比 P_S、结构相似性 S_S。上述评价指标的数值与融合质量呈正相关关系，即指标值越大，表明融合效果越好。

2.3 MSRS 数据集对比实验

2.3.1 主观评价

本文在 MSRS 数据集中选取了 3 组图像进行主观评价，结果如图 4 所示，其中，红框和绿框标示了图像的局部放大区域。

(a) IR (b) VI (c) DenseFuse (d) FusionGAN (e) GANMcC (f) PMGI



(g) SeAFusion (h) U2Fusion (i) SDCFusion (j) IMF (k) T2EA (l) Ours

图 4 MSRS 数据集不同算法融合结果

Fig. 4 The fusion results of different algorithms in the MSRS dataset

如图 4 所示可以看到，在第 1 组和第 2 组的白天场景中，DenseFuse、FusionGAN、GANMcC、U2Fusion 受光谱污染影响，导致融合图像的颜色偏暗且均偏向于红外图像，缺失了源可见光图像的细节信息。PMGI 出现了颜色失真，导致融合质量的视觉效果较差，SeAFusion 与 IMF 虽然亮度尚可，但出现了树木边缘模糊。SDCFusion 较好地保留了可见光的纹理，但红外背景细节略有丢失。T2EA 的颜色亮度较为偏暗，在第 3 组黑夜场景中，DenseFuse、FusionGAN、GANMcC 在黑暗背景中引入可见光噪声，导致目标边缘出现重影，PMGI 与 T2EA 对人物红外辐射强度估计不足，目标暗淡。SeAFusion、IMF 与 SDCFusion 虽提升了红外对比度，却在车灯高光区域产生过曝。与其他方法相比，本文方法获得的图像纹理清晰且无伪影，整体亮度均衡、无颜色偏移。

2.3.2　客观评价

表 1 给出了在 MSRS 数据集 20 组融合图像的平均值，其中，粗体表示最优值，通过对 MSRS 数据集的 6 项关键指标评估，本文方法展现出较好的综合性能。具体而言，在 F_S、G_A、F_{VI}、P_S、S_S的 5 项核心指标上，本文方法均位列第一，比其他方法分别平均提高 72.18%、66.68%、22.15 %、0.79%、3.36%，在 I_M上排名第 3，但仍优于多数对比方法。F_S最优，说明融合图像具有更高的整体清晰度和更丰富的纹理细节，这得益于本文设计的双分支特征提取网络保留了局部信息与全局信息；G_A最优，说明融合图像细节更突出；F_{VI}最优，说明融合图像保留了最多的来自源图像的可感知信息；P_S最优，说明融合图像具有非常高的保真度，这是因为本文设计的泰勒展开网络提取多层次特征信息，同时保留更多精细特征；S_S最优，说明融合图像更好地保留了源图像的结构信息。I_M排名仅次于 SeAFusion 与 SDCFusion，说明融合图像对于源图像信息也有不错的保留。

表 1　MSRS 数据集 20 组融合图像的平均值

Tab. 1　The average of 20 groups of fused images in the MSRS dataset

Method	F_S	G_A	I_M	F_{VI}	P_S	S_S
DenseFuse	0.023 5	2.004 3	2.662 5	0.676 7	66.677 4	0.931 0
FusionGAN	0.017 7	1.520 2	2.053 8	0.500 0	64.708 6	0.831 9
GANMcC	0.024 8	2.229 0	2.547 1	0.675 9	65.244 8	0.915 9
PMGI	0.032 4	2.935 2	2.170 6	0.674 6	60.299 3	0.931 5
SeAFusion	0.043 6	3.575 4	**4.102 7**	0.960 9	63.763 6	0.975 5
U2Fusion	0.024 3	1.888 4	2.292 6	0.562 4	65.972 3	0.899 1
SDCFusion	0.046 0	3.844 6	3.679 0	0.971 6	63.571 7	0.940 0
IMF	0.035 2	2.888 4	2.931 3	0.769 8	65.455 4	0.926 9
T2EA	0.033 0	2.844 4	2.604 6	0.729 5	65.797 0	0.939 8
Ours	**0.052 5**	**4.580 5**	2.994 4	**0.972 9**	**68.224 7**	**0.981 6**

2.4　RoadScene 数据集对比试验

2.4.1　主观评价

本文在 RoadScene 数据集中选取 3 组图像进行主观评价，以验证算法的泛化性能。其中，红框和绿框标示了图像的局部放大区域，如图 5 所示，在第 1 组的白天场景中，DenseFuse 整体对比度不足，车辆轮廓边缘模糊；FusionGAN 与 GANMcC 在引擎盖区域出现明显光谱伪影，颜色失真；PMGI 对路面标识的保持较弱，细节被平滑；SeAFusion 亮度尚可，但后车窗纹理被过度增强，产生亮斑；U2Fusion 与 SDCFusion 在树木边缘处均出现轻微重影；IMF 与 T2EA 整体偏灰，缺乏饱和度。在第 2 组、第 3 组的夜晚场景中，DenseFuse 整体亮度不足，道路标线几乎不可辨；FusionGAN 与 GANMcC 边缘出现伪影，灯光细节信息丢失；PMGI 整体色调偏冷，绿色植被失去原有饱和度；SeAFusion 与 SDCFusion 均出现过度曝光；U2Fusion 与 T2EA 缺失了路灯细节；与其他方法相比，本文方法主观观感最接近可见光源图像，同时保留了红外目标的显著性。

2.4.2　客观评价

表 2 给出了在 RoadScene 数据集 20 组融合图像的平均值，其中，粗体表示最优值，通过对 RoadScene 数据集的 6 项关键指标评估，本文方法展现出较好的综合性能。具体而言，在 F_S、G_A、I_M、F_{VI}四项核心指标上，本文方法均位列第 1，说明本文方法在纹理细节保留与视觉显著性方面取得较好的效果，这得益于本文设计的泰勒展开网络与双分支特征提取网络充分提取到多层级特征并保留了全局特征信息。在 P_S、S_S指标排名第 3，这是因为本文使用 MSRS 数据集作为训练集，而 RoadScene 数据集作为测试集，导致指标出现下降。整体上，本文在 RoadScene 数据集取得较好的效果。

图 5 RoadScene 数据集不同算法融合结果

Fig. 5 The fusion results of different algorithms in the RoadScene dataset

表 2 RoadScene 数据集 20 组融合图像的平均值

Tab. 2 The average of 20 groups of fused images in the RoadScene dataset

Method	F_S	G_A	I_M	F_{VI}	P_S	S_S
DenseFuse	0.034 6	3.439 3	3.058 1	0.651 8	**62.029 7**	0.879 7
FusionGAN	0.038 5	3.814 3	2.973 1	0.579 2	59.015 4	0.751 9
GANMcC	0.040 1	4.277 3	2.795 3	0.723 8	59.229 8	0.862 4
PMGI	0.046 6	4.863 6	3.329 4	0.773 7	59.977 9	**0.901 4**
SeAFusion	0.049 2	4.416 1	3.163 0	0.832 0	59.646 0	0.876 3
U2Fusion	0.035 2	3.743 2	2.277 2	0.723 8	61.529 0	0.864 0
SDCFusion	0.071 8	7.280 3	2.680 3	0.722 6	59.883 3	0.899 7
IMF	0.043 3	4.572 9	3.009 4	0.732 9	60.536 2	0.836 0
T2EA	0.052 9	5.565 1	2.566 6	0.673 8	59.485 5	0.887 4
Ours	**0.075 2**	**7.420 3**	**3.878 0**	**0.862 2**	60.609 3	0.888 6

2.5 消融实验

本文对完整模型(Ours)进行了消融实验,分别移除 TEN、PCN、ST 以及 CA 模块,分析了对应的模型变体(Model1、Model2、Model3、Model4)。如表 3 与图 6 所示,完整模型在 MSRS 数据集上的融合性能优于各消融策略下的模型,证明了其融合效果的优越性。Model1 因缺少 TEN 模块导致图像细节丢失,各项指标下降。Model2 移除 PCN 模块后,人物边缘与场景信息保持欠佳,G_A、I_M等指标显著降低。Model3 与 Model4 分别移除 ST 与 CA 模块,均表现出在边缘、纹理等细节融合上的不足,并导致 F_{VI}、P_S、G_A 等关键指标下滑。上述实验结果验证了 TEN、PCN、ST 和 CA 模块对于提升融合图像的细节质量与客观指标均具有不可替代的重要作用。

表 3　消融实验结果

Tab. 3　Ablation experimental results

模型	TAN	PCN	ST	CA	F_S	G_A	I_M	F_{VI}	P_S	S_S
Model1		√	√	√	0.035 2	3.672 3	2.368 4	0.901 7	66.459 8	0.914 8
Model2	√		√	√	0.045 1	3.845 5	2.528 7	0.917 6	66.741 8	0.925 4
Model3	√	√		√	0.046 9	3.989 2	2.635 9	0.954 8	66.588 2	0.899 7
Model4	√	√	√		0.047 6	3.865 1	2.954 8	0.971 3	68.015 7	0.971 6
Ours	√	√	√	√	**0.052 5**	**4.580 5**	**2.994 4**	**0.972 9**	**68.224 7**	**0.981 6**

(a) IR　(b) VI　(c) Model1　(d) Model2　(e) Model3　(f) Model4　(g)Ours

图 6　MSRS 数据集消融实验结果

Fig. 6　Ablation experiment results of the MSRS dataset

2.6　计算复杂度分析

本文对算法的计算复杂度进行分析，核心指标包括参数量与计算效率。由表 4 可知，本文方法的参数量排名第 4，模型结构相对轻量。在融合效率方面，本文方法在 MSRS 和 RoadScene 数据集上的单位融合时间分别位列第 2 和第 3，表现出优异的计算效率。实验结果表明，本文方法在参数量和平均运行时间上取得较好的效果。

表 4　参数量和平均运行时间

Tab. 4　Parameter quantity and average running time

Method	Params/M	Time/s	
		MSRS	RoadScene
DenseFuse	**0.075 9**	0.035 4	0.021 5
FusionGAN	0.925 6	1.893 4	1.265 4
GANMcC	1.865 6	2.364 7	1.935 1
PMGI	0.651 8	0.579 3	0.268 7
SeAFusion	0.167 7	0.167 9	0.043 4
U2Fusion	0.653 8	1.617 9	0.827 9
SDCFusion	0.565 4	0.196 7	0.056 4
IMF	0.736 3	**0.015 1**	**0.015 2**
T2EA	0.296 4	0.546 6	0.154 6
Ours	0.449 4	0.025 4	0.025 0

3　结　论

本文提出了一种泰勒展开与复合注意力的红外与可见光图像融合网络，通过大量实验验证，得出以下主要结论。

1）通过泰勒展开网络将源图像分解为映射层与导数层，以充分提取多层级特征，有效避免了细节丢失并保留细微结构。在此基础上，构建了双分支特征提取结构，保留了局部特征与全局上下文信息。

2）为进一步提升融合性能，引入复合注意力机制，从空间与通道两个维度协同优化，显著提升了多模态图像融合的精度与鲁棒性。

3）与其他融合方法相比，在 MSRS 数据集测试中，本文方法在 F_S、G_A、F_{VI}、P_S、S_S 五项核心指标方面，均位列第 1，比其他方法分别平均提高 72.18%、66.68%、22.15%、0.79%、3.36%。同时，在 RoadScene 数据集中也取得了较好的结果。证明了本文方法能够保障融合图像的整体清晰度、全局纹理的丰富度，并在目标与背景的平衡上表现出卓越的细节保持能力。

参考文献

[1] LIU Jinyuan, WU Guanyao, LIU Zhu, et al. Infrared and visible image fusion: From data compatibility to task adaption[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, 47(4): 2349. DOI: 10.1109/TPAMI.2024.3521416

[2] 张洲宇，曹云峰，丁萌，等. 采用多层卷积稀疏表示的红外与可见光图像融合[J]. 哈尔滨工业大学学报，2021，53(12)：51
ZHANG Zhouyu, CAO Yunfeng, DING Meng, et al. Infrared and

visible image fusion via multi-layer convolutional sparse representation[J]. Journal of Harbin Institute of Technology, 2021, 53(12): 51. DOI:10.11918/202005038

[3]史文云, 任晓明, 颜楠楠. 基于 Schatten-p LatLRR 的电力设备红外与可见光图像融合[J]. 激光技术, 2025, 49(1): 67
SHI Wenyun, REN Xiaoming, YAN Nannan. Fusion of infrared and visible light images of power equipment based on Schatten-p LatLRR[J]. Laser Technology, 2025, 49(1): 67. DOI:10.7510/jgjs.issn.1001 - 3806.2025.01.011

[4]李秋恒, 邓豪, 刘桂华, 等. 基于多尺度及多头注意力的红外与可见光图像融合[J]. 红外技术, 2024, 46(7): 765
LI Qiuheng, DENG Hao, LIU Guihua, et al. Infrared and visible images fusion method based on multi-scale features and multi-head attention[J]. Infrared Technology, 2024, 46(7): 765

[5]CHEN Jun, LI Xuejiao, LUO Linbo, et al. Multi-focus image fusion based on multi-scale gradients and image matting [J]. IEEE Transactions on Multimedia, 2021, 24: 655. DOI:10.1109/TMM.2021.3057493

[6]WANG Haozhe, SHU Chang, LI Xiaofeng, et al. Two-stream edge-aware network for infrared and visible image fusion with multi-level wavelet decomposition[J]. IEEE Access, 2024, 12: 22190. DOI: 10.1109/ACCESS.2024.3364050

[7]LI Yonghua, LIU Gang, BAVIRISETTI D P, et al. Infrared-visible image fusion method based on sparse and prior joint saliency detection and LatLRR-FPDE[J]. Digital Signal Processing, 2023, 134: 103910. DOI:10.1016/j.dsp.2023.103910

[8]LI Jun, SONG Minghui, PENG Yuanxi. Infrared and visible image fusion based on robust principal component analysis and compressed sensing[J]. Infrared Physics & Technology, 2018, 89: 129. DOI: 10.1016/j.infrared.2018.01.003

[9]LI Hui, WU Xiaojun, 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. DOI:10.1109/TIM.2020.3005230

[10]ZHAO Fan, ZHAO Wenda, YAO Libo, et al. Self-supervised feature adaption for infrared and visible image fusion [J]. Information Fusion, 2021, 76: 189. DOI:10.1016/j.inffus.2021.06.002

[11]ZHAO Zixiang, XU Shuang, ZHANG Jiangshe, et al. Efficient and model-based infrared and visible image fusion via algorithm unrolling[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(3): 1186. DOI:10.1109/TCSVT.2021.3075745

[12]LI Jiawei, CHEN Jiansheng, LIU Jinyuan, et al. Learning a graph neural network with cross modality interaction for image fusion[C]// Proceedings of the 31st ACM International Conference on Multimedia. Ottawa: ACM, 2023: 4471. DOI:10.1145/3581783.3612135

[13]YUE Jun, FANG Leyuan, XIA Shaobo, et al. Dif-Fusion: Toward high color fidelity in infrared and visible image fusion with diffusion models[J]. IEEE Transactions on Image Processing, 2023, 32: 5705. DOI:10.1109/TIP.2023.3322046

[14]YANG Yong, ZHOU Na, WAN Weiguo, et al. MACCNet: Multiscale attention and cross-convolutional network for infrared and visible image fusion[J]. IEEE Sensors Journal, 2024, 24(10): 16587. DOI:10.1109/JSEN.2024.3385638

[15]ZHOU Huabing, WU Wei, ZHANG Yanduo, et al. Semantic-supervised infrared and visible image fusion via a dual-discriminator generative adversarial network [J]. IEEE Transactions on Multimedia, 2021, 25: 635. DOI:10.1109/TMM.2021.3129609

[16]YIN Haitao, XIAO Jinghu, CHEN Hao. CSPA-GAN: A cross-scale pyramid attention GAN for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1. DOI:10.1109/TIM.2023.3317932

[17]WANG Zhishe, ZHANG Zhouqun, QI Wuqiang, et al. FreqGAN: Infrared and visible image fusion via unified frequency adversarial learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(1): 728. DOI:10.1109/TCSVT.2024.3460172

[18]YI Xupeng, XU Han, ZHANG Hao, et al. Text-IF: Leveraging semantic text guidance for degradation-aware and interactive image fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024: 27026. DOI:10.48550/arxiv.2403.16387

[19]LIU Jinyuan, LI Xingyuan, WANG Zirui, et al. PromptFusion: Harmonized semantic prompt learning for infrared and visible image fusion [J]. IEEE/CAA Journal of Automatica Sinica, 2024, 12(3): 502. DOI:10.1109/JAS.2024.124878

[20]TANG Wei, HE Fazhi, LIU Yu. YDTR: Infrared and visible image fusion via Y-shape dynamic transformer[J]. IEEE Transactions on Multimedia, 2022, 25: 5413. DOI: 10.1109/TMM.2022.3192661

[21]LI Hui, WU Xiaojun. DenseFuse: A fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614. DOI:10.1109/TIP.2018.2887342

[22]MA Jiayi, YU Wei, LIANG Pengwei, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion [J]. Information Fusion, 2019, 48: 11. DOI:10.1016/j.inffus.2018.09.004

[23]MA Jiayi, ZHANG Hao, SHAO Zhenfeng, et al. GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1. DOI:10.1109/TIM.2020.3038013

[24]ZHANG HAO, XU Han, XIAO Yang, et al. Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity[C]//34th AAAI Conference on Artificial Intelligence. New York: AAAI, 2020: 12797. DOI: 10.1609/AAAI.V34i07.6975

[25]TANG Linfeng, YUAN Jiteng, MA Jiayi. Image fusion in the loop of high-level vision tasks: A semantic-aware real-time infrared and visible image fusion network[J]. Information Fusion, 2022, 82: 28. DOI:10.1016/j.inffus.2021.12.004

[26]XU Han, MA Jiayi, JIANG Junjun, et al. U2Fusion: A unified unsupervised image fusion network [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(1): 502. DOI:10.1109/TPAMI.2020.3012548

[27]LIU Xiaowen, HUO Hongtao, LI Jing, et al. A semantic-driven coupled network for infrared and visible image fusion [J]. Information Fusion, 2024, 108: 102352. DOI:10.1016/j.inffus.2024.102352

[28]WANG Di, LIU Jinyuan, MA Long, et al. Improving misaligned multi-modality image fusion with one-stage progressive dense registration[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2024, 34(11): 10944. DOI: 10.1109/TCSVT.2024.3412743

[29]HUANG Zhenghua, LIN Cheng, XU B, et al. T2EA: Target-aware Taylor expansion approximation network for infrared and visible image fusion[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025, 35(5): 4831. DOI:10.1109/TCSVT.2024.3524794

(编辑 吕雪梅)