DOI:10.11918/202406056

基于语义驱动的红外与可见光图像交互融合

王瑾春[1],马 萍[2],张宏立[2],王 聪[2],苑 茹[1]

(1.新疆大学 电气工程学院,乌鲁木齐 830017;2.新疆大学 智能科学与技术学院,乌鲁木齐 830017)

摘 要:为解决现有的红外与可见光图像融合算法存在像素信息保留和语义特征提取不足的问题,提出一种基于语义驱动的红外与可见光图像交互融合算法。首先,通过联合操作图像融合网络和图像分割网络,形成语义驱动效果,更好地保留图像在像素域和语义域的信息特征;然后,构建跨域交互整合模块,捕捉红外与可见光图像特征,允许特征在不同空间和独立通道之间交互传递,实现特征从局部到全局的映射,增强两类图像的互补特性;最后,引入语义损失函数约束网络训练以保留源图像的内在语义特征。在多波段图像数据集和多光谱道路场景数据集上进行图像融合和分割实验,并与其他6种先进的融合算法进行比较。融合实验结果表明,本文算法在基于梯度的相似性度量、信息熵、峰值信噪比、空间频率、标准差、视觉保真度6个客观评价指标上分别平均提高了47.92%、6.15%、0.87%、44.31%、35.99%、36.88%;分割实验结果表明,本文算法在所有评价指标中,结果均为最优。所提算法在主观视觉效果的定性分析与客观质量评价的定量指标方面整体效果优于现有融合算法,融合图像可以兼顾视觉质量和高级语义任务,能更好地服务于人类视觉观察和机器视觉感知。

关键词:交互融合;红外与可见光图像;语义驱动;语义分割

中图分类号:TN911.73 **文献标志码**:A **文章编号**:0367-6234(2025)09-0056-09

Semantic-driven interactive fusion of infrared and visible images

WANG Jinchun[1], MA Ping[2], ZHANG Hongli[2], WANG Cong[2], YUAN Ru[1]

(1. School of Electrical Engineering, Xinjiang University, Urumqi 830017, China;
2. School of Intelligence Science and Technology, Xinjiang University, Urumqi 830017, China)

Abstract: In order to solve the limitations of existing infrared and visible images fusion algorithms in preserving pixel-level information and extracting semantic features, an infrared and visible image interactive fusion method based on semantic driven was proposed. First, the image fusion network and the image segmentation network were jointly operated to form a semantic-driven effect, enhancing the retention of information features of the image in both pixel domain and semantic domain. Then, a cross-domain interactive integration module was constructed to capture features of infrared and visible images, allowing for the interactive transfer of features across different spatial locations and independent channels, thereby mapping features from local to global, and enhancing the complementary characteristics of the two types of images. Finally, a semantic loss function was introduced to constrain the network training, preserving the intrinsic semantic features of the source images. Pixel-level fusion experiments and semantic-level segmentation experiments were conducted on multi-band data sets and multi-spectral road scene data sets. These experiment results were then compared with six other advanced fusion algorithms. The results of fusion experiments show that the proposed algorithm achieves improvements of 47.92%, 6.15%, 0.87%, 44.31%, 35.99% and 36.88% across six objective evaluation metrics, including gradient-based similarity measures, information entropy, peak signal-to-noise ratio, spatial frequency, standard deviation and visual fidelity. The results of segmentation experiments indicate that the proposed algorithm outperforms all other evaluation metrics. Therefore, the proposed method exhibits superior performance in both qualitative analysis of subjective visual effects and quantitative indicators of quality evaluation compared to existing algorithms. The fusion images effectively balance both visual quality and high-level semantic tasks, thereby enhancing utility for human visual observation and machine vision perception.

Keywords: interactive fusion; infrared and visible images; semantic-driven; semantic segmentation

收稿日期:2024-06-24;**录用日期**:2024-07-26;**网络首发日期**:2025-04-01
网络首发地址:https://link.cnki.net/urlid/23.1235.T.20250401.1405.002
基金项目:新疆维吾尔自治区自然科学基金(2022D01C367,2023D01C187);"天山英才"培养计划(2023TSYCQNTJ0020,2023TSYCCX0037)
作者简介:王瑾春(2000—),女,硕士研究生;马 萍(1994—),女,副教授,博士生导师
通信作者:马 萍,maping@xju.edu.cn

在图像视觉处理领域中，单一模态的图像通常只具备某一方面的信息，无法完整地表征复杂场景的全部信息[1]，因此具有信息集成特性的图像融合技术应时而生。图像融合技术通过整合不同传感器或多种成像模式的图像特征，生成更全面、更准确、更丰富的图像。其中，红外与可见光图像因其信息互补性强、模态差异性明显、环境适应性好的特性广泛应用于多传感器信息融合领域，涉及目标检测[2]、汽车无人驾驶[3]、军事监控识别[4]等。

基于深度学习的图像融合算法因其能自适应不同的数据，完成复杂场景的融合任务而受到广泛关注。主要包括自编码器模型（autoencoder，AE）、卷积神经网络模型（convolutional neural network，CNN）和生成对抗网络模型（generative adversarial network，GAN）3 类。AE 对图像进行编码和解码操作，实现对图像的压缩和重建，最终得到融合图像。Li 等[5]提出一种由编码器和解码器构成的密集连接图像融合网络，在编码器中加入密集连接块以增强特征提取。Jian 等[6]提出一种具有残差块对称编码器和解码器的融合网络框架，在降低网络复杂度的同时减少冗余信息生成。但这些算法仍需手工设计融合策略，而 CNN 通过卷积操作实现特征提取和融合。Tang 等[7]引入光照感知指导融合网络（progressive infrared and visible image fusion network based on illumination aware，PIAFusion），保持图像稳定融合。但使用 CNN 会产生过拟合问题，导致对新样本的泛化能力不佳，且参数调优耗费大量精力。GAN 通过生成器和鉴别器构建博弈指导图像融合，可以自动捕捉数据分布，获得最优融合结果。Ma 等[8]首次将 GAN 用于图像融合任务中，提出了基于红外与可见光图像融合的生成对抗网络（generative adversarial network for infrared and visible image fusion，FusionGAN），由于其特征提取机制局限于单一鉴别器，导致融合结果过度依赖红外图像的亮度特征，而未能充分保留可见光图像的纹理细节信息。因此，Li 等[9]将两种注意力机制分别用于提取和融合局部区域和整个图像的特征，增强了融合图像的效果。但现有的图像融合算法主要关注像素级信息，表现在是否具有良好的视觉效果，如红外特征是否显著、可见光纹理是否清晰等，忽略了场景的高级语义信息。

为实现良好的融合效果，并促进后续高级视觉任务，Tang 等[10]提出了一种语义感知实时红外与可见光图像融合网络（semantic-aware real-time infrared and visible image fusion network，SeAFusion）、级联融合网络与分割网络，将融合结果输入分割网络得到语义损失，通过内容损失和语义损失共同优化融合网络。Liu 等[11]采用双层优化形式进行图像融合和目标检测任务，实现高精度的检测和高质量的视觉融合效果。Liu 等[12]提出一种耦合融合-分割网络（semantic-driven coupled network for infrared and visible image fusion，SDCFusion），增强了融合图像的语义性能。语义驱动网络的本质是将像素级融合结果送入分割网络，获得语义表示，并通过语义驱动动作引导融合网络关注更多的语义信息。但是，语义信息的局部注入忽略了像素信息与语义信息的关系，难以平衡融合图像的像素表示和语义表示。

针对以上问题，本文提出一种基于语义驱动的红外与可见光图像交互融合算法，通过级联融合网络和分割网络，并引入语义损失函数，共同解决融合和分割任务的特征异质性问题。设计了跨域交互整合模块，通过局部通道和空间注意力模块与全局交叉注意力模块捕捉红外与可见光图像的局部关键信息和全局信息，以增强融合任务和分割任务的鲁棒性，为红外与可见光图像融合研究提供了新思路。

1　红外与可见光图像的语义驱动交互融合网络结构

图 1 展示了本文设计的红外与可见光图像的语义驱动交互融合网络结构，由 4 个核心模块组成。

1）共享编码器模块：基于双分网络架构分别提取红外与可见光图像的像素级特征表示。

2）跨域交互整合模块：整合编码器提取的特征，构建特征间的权值，生成跨域交互整合特征。

3）融合解码器模块：重建融合特征，得到融合图像。

4）分割解码器模块：生成语义分割结果。

共享编码器由两条独立的编码路径组成，分别提取多尺度红外与可见光特征。每条编码路径包括 4 个卷积块，其中，第 1 个卷积块的步长为 1，第 2 ~ 4 个卷积块的步长为 2。每个卷积块的卷积核尺寸为 3 × 3，由 Padding 层、Conv2D 层和 PReLU 层组成。将编码器提取的多尺度图像特征汇入跨域交互整合模块，为图像特征分配权重，集成跨域交互整合特征。将特征依次分配至融合解码器和分割解码器中，最终得到融合图像和分割图像。

此外，为了更好地满足高级视觉任务，引入了语义损失、内容损失和梯度损失共同约束网络，使融合后的图像具有更丰富的语义信息。

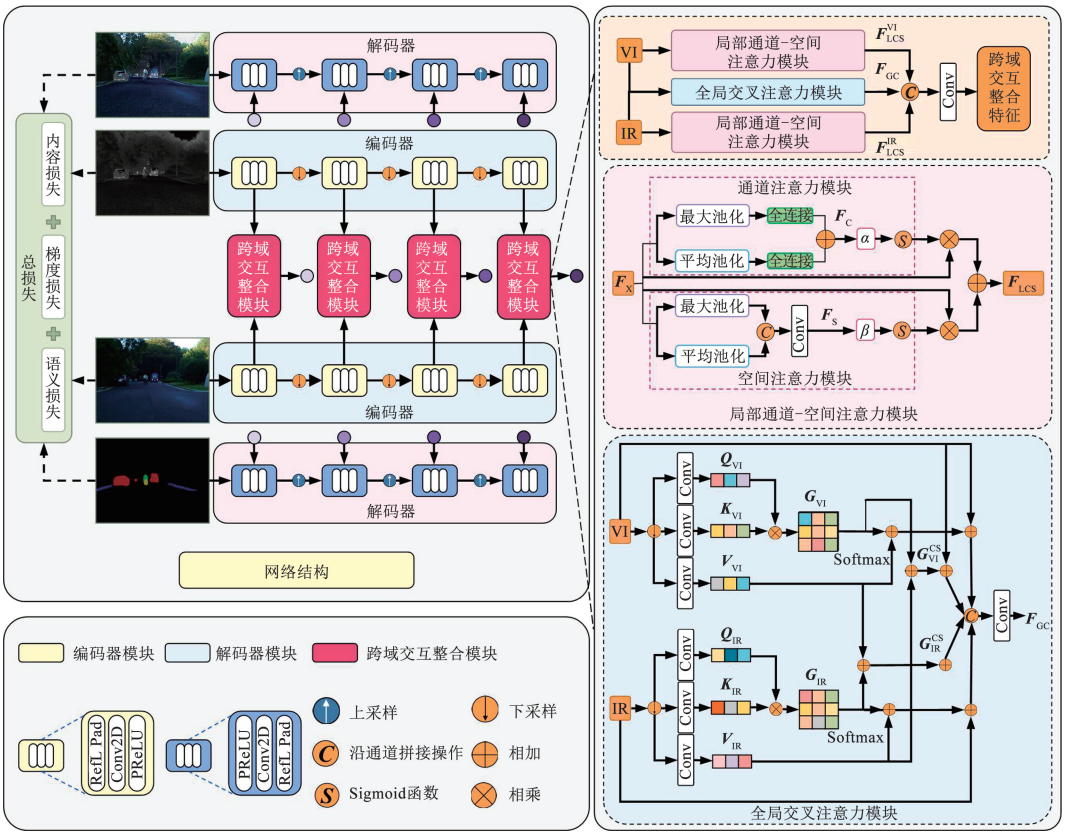


图 1　基于语义驱动的红外与可见光图像交互融合网络结构

Fig. 1　Semantic-driven interactive fusion network structure of infrared and visible images

2　红外与可见光图像的语义驱动交互融合算法

2.1　跨域交互整合模块

为充分融合跨域图像像素和语义信息，本文设计了一种跨域交互整合模块，该模块有两个局部通道-空间注意力模块和 1 个全局交叉注意力模块组成，以此获取单模态图像的关键特征和多模态图像的全局特征。

由于红外与可见光图像的模态差异，图像中的关键信息和非重要信息往往不一致，为提升模型对图像重要特征的关注和提取能力，采用局部通道-空间注意力模块，结合通道注意力和空间注意力，自适应地调整特征图在通道维度和空间维度上的重要性，获取不同模态的加权特征表示。首先，将通道维度上的红外与可见光特征送入通道注意力模块中，分别利用平均池化和最大池化对特征 $\boldsymbol{F}_{\mathrm{x}}$ 进行处理，之后利用全连接层处理特征图，学习每个通道的注意力权重，最后将全连接层的两个输出特征进行逐像素相加，引入可学习权重 α，对特征加权，生成通道特征 $\boldsymbol{F}_{\mathrm{c}}$，这一过程表示为

$$\boldsymbol{F}_{\mathrm{c}}=F_{\mathrm{Mlp}}(F_{\mathrm{Avg}}(\boldsymbol{F}_{\mathrm{x}}))\oplus F_{\mathrm{Mlp}}(F_{\mathrm{Max}}(\boldsymbol{F}_{\mathrm{x}})) \tag{1}$$

式中：$F_{\mathrm{Avg}}(\cdot)$ 为平均池化操作，$F_{\mathrm{Max}}(\cdot)$ 为最大池化操作，$F_{\mathrm{Mlp}}(\cdot)$ 为确定全连接层各通道的权值操作，$\oplus$ 为元素相加。

$$\boldsymbol{F}_{\alpha}=S(\alpha\boldsymbol{F}_{\mathrm{c}}) \tag{2}$$

式中：$\boldsymbol{F}_{\alpha}$ 为引入 α 权重后的通道特征；$S(\cdot)$ 为 Sigmoid 函数，将输入数值转换在 0 ~ 1 内，实现数据归一化处理。

同样地，将空间维度上的红外与可见光特征送入空间注意力模块中，分别利用平均池化和最大池化对 $\boldsymbol{F}_{\mathrm{x}}$ 进行通道拼接，利用卷积降维后，引入可学习权重 β 对特征加权。融合权重生成过程可表示为

$$\boldsymbol{F}_{\mathrm{s}}=F_{\mathrm{Conv}}(C(F_{\mathrm{Avg}}(\boldsymbol{F}_{\mathrm{x}}))\oplus C(F_{\mathrm{Max}}(\boldsymbol{F}_{\mathrm{x}}))) \tag{3}$$

式中：$F_{\mathrm{Conv}}(\cdot)$ 为卷积操作；$C(\cdot)$ 为 Concat 操作，表示沿通道拼接操作。

$$\boldsymbol{F}_{\beta}=S(\beta\boldsymbol{F}_{\mathrm{s}}) \tag{4}$$

式中 $\boldsymbol{F}_{\beta}$ 为引入 β 权重后的空间特征。

最终得到局部通道-空间注意力融合特征，其表达式为

$$\boldsymbol{F}_{\mathrm{LCS}}=(\boldsymbol{F}_{\mathrm{x}}\otimes\boldsymbol{F}_{\alpha})\oplus(\boldsymbol{F}_{\mathrm{x}}\otimes\boldsymbol{F}_{\beta}) \tag{5}$$

式中 $\otimes$ 为元素相乘。

针对红外与可见光图像间存在的语义相关性和差异性，为得到丰富的全局上下文信息，采用全局交叉注意力模块对两种模态图像的语义关系进行建模，对输入特征进行 3 次线性变换，分别得到查询

($\boldsymbol{Q}_{\mathrm{IR}}$,$\boldsymbol{Q}_{\mathrm{VI}}$)、键($\boldsymbol{K}_{\mathrm{IR}}$,$\boldsymbol{K}_{\mathrm{VI}}$)和值($\boldsymbol{V}_{\mathrm{IR}}$,$\boldsymbol{V}_{\mathrm{VI}}$)。首先，计算单模态语义关系，得到可见光全局关系图谱 $\boldsymbol{G}_{\mathrm{VI}}$ 和红外全局关系图谱 $\boldsymbol{G}_{\mathrm{IR}}$。计算过程为

$$\begin{cases}\boldsymbol{G}_{\mathrm{VI}}=\sigma(\boldsymbol{Q}_{\mathrm{VI}}\boldsymbol{K}_{\mathrm{VI}}^{\mathrm{T}})\\ \boldsymbol{G}_{\mathrm{IR}}=\sigma(\boldsymbol{Q}_{\mathrm{IR}}\boldsymbol{K}_{\mathrm{IR}}^{\mathrm{T}})\end{cases}\tag{6}$$

式中 $\sigma(\cdot)$ 为 Softmax 函数，表示对每个节点的注意力权重进行归一化处理。

然后，计算跨模态语义信息，得到可见光跨域关系图谱 $\boldsymbol{G}_{\mathrm{VI}}^{\mathrm{CS}}$ 和红外跨域关系图谱 $\boldsymbol{G}_{\mathrm{IR}}^{\mathrm{CS}}$。计算过程为

$$\begin{cases}\boldsymbol{G}_{\mathrm{VI}}^{\mathrm{CS}}=\sigma(\boldsymbol{Q}_{\mathrm{IR}}\boldsymbol{K}_{\mathrm{VI}}^{\mathrm{T}})\\ \boldsymbol{G}_{\mathrm{IR}}^{\mathrm{CS}}=\sigma(\boldsymbol{Q}_{\mathrm{VI}}\boldsymbol{K}_{\mathrm{IR}}^{\mathrm{T}})\end{cases}\tag{7}$$

最后，将原始红外与可见光特征、红外与可见光全局关系图谱、红外与可见光跨域关系图谱沿通道方向进行连接后送到 1×1 的卷积层，得到全局交叉注意力融合特征 $\boldsymbol{F}_{\mathrm{GC}}$。

将局部通道–空间注意力融合特征和全局交叉注意力融合特征沿通道方向进行连接，结果汇集至收缩卷积层，最终获得跨域交互整合特征。

2.2　损失函数

为约束融合结果与源图像之间的差异，引导融合网络保留像素的内容和梯度，引入了内容损失 L_{con} 和梯度损失 L_{grad}；为使图像包含更多的语义信息，获得具有语义跨域耦合特征，引入了语义损失 L_{se} 以约束融合网络。因此，本文的损失函数有内容损失 L_{con}、梯度损失 L_{grad} 和语义损失 L_{se}。

为使融合图像中包含可见光图像中的完整光谱和背景信息，减少红外冗余信息的干扰，引入了语义分割标签中的目标掩模 M，计算式为

$$M(i,j)=\begin{cases}1, & \boldsymbol{I}_{\mathrm{IR}}(i,j)>\boldsymbol{I}_{\mathrm{VI}}(i,j)\\ 0, & \text{其他}\end{cases}\tag{8}$$

式中：(i,j) 为图像的像素位置，$\boldsymbol{I}_{\mathrm{IR}}$ 为红外图像信息，$\boldsymbol{I}_{\mathrm{VI}}$ 为可见光图像信息。

$$L_{\mathrm{con}}=\frac{1}{HW}(\|\boldsymbol{I}_{\mathrm{F}}-\boldsymbol{I}_{\mathrm{VI}}\|_2^2+M\|\boldsymbol{I}_{\mathrm{F}}-\boldsymbol{I}_{\mathrm{IR}}\|_1)\tag{9}$$

式中：$\|\cdot\|_n(n=1,2)$ 为 L_n 范数，$\boldsymbol{I}_{\mathrm{F}}$ 为融合图像信息，H 为图像高度，W 为图像宽度。

为提升融合图像中可见光纹理特征与红外有效信息的保留效果，引入 L_{grad} 最小化融合图像与源图像之间的梯度差异，L_{grad} 表达式为

$$L_{\mathrm{grad}}=\frac{1}{HW}\|\,|\nabla\boldsymbol{I}_{\mathrm{F}}|-\max(|\nabla\boldsymbol{I}_{\mathrm{IR}}|,|\nabla\boldsymbol{I}_{\mathrm{VI}}|)\,\|_1\tag{10}$$

式中 ∇ 为 Sobel 梯度算子。

利用交叉熵损失函数(online hard example mining cross entropy loss，L_{ohem})计算 L_{se}，计算式为

$$L_{\mathrm{se}}=L_{\mathrm{ohem}}(I_{\mathrm{seg}},I_{\mathrm{label}})\tag{11}$$

式中：I_{seg} 为语义分割结果，I_{label} 为语义分割标签。

网络训练的总损失函数为

$$L_{\mathrm{total}}=\lambda_1L_{\mathrm{con}}+\lambda_2L_{\mathrm{grad}}+\lambda_3L_{\mathrm{se}}\tag{12}$$

式中 λ_1、λ_2 和 λ_3 为权重系数，分别用于平衡 L_{con}、L_{grad} 和 L_{se} 的关系。

3　实验结果与分析

3.1　实验设置

3.1.1　实验平台及参数设置

实验硬件平台的操作系统为 Windows11，CPU 为 AMD EPYC Processor，主频 2.45 GHz，内存 32 GB，GPU 为 NVIDIA GeForce RTX 4090；软件平台为 MATLAB2023a，Pycharm2023。损失函数的权重系数 λ_1、λ_2、λ_3 分别为 1、10、1，利用 Adam 优化器进行参数更新，训练批次设置为 30，迭代数设定为 8，学习率为 2×10^{-5}。

3.1.2　数据集

在训练阶段，采用多光谱道路场景 MSRS 数据集[13]中的 1 083 对已配准训练数据作为训练集。在测试阶段，分别从多波段图像数据集 TNO[14] 和 MSRS 中选择 42、361 对图像进行实验验证。

3.1.3　对比算法及评价指标

为评估所提算法的性能优势，选择了 6 种典型的深度学习算法与其进行对比，即 DenseFuse[5]、PIAFusion[7]、FusionGAN[8]、SeAFusion[10]、SDCFusion[12] 和基于 SuperFusion 融合算法[15]。

选取基于梯度的相似性度量(gradient-based similarity measurement，$Q^{\mathrm{AB/F}}$)、信息熵(entropy，E_{N})、峰值信噪比(peak signal-to-noise ratio，R_{PSN})、空间频率(spatial frequency，F_{S})、标准差(standard deviation，D_{S})、视觉保真度(visual information fidelity，F_{VI})作为图像融合的客观评价指标。其中，$Q^{\mathrm{AB/F}}$ 通过衡量图像之间的相互影响进行评价；E_{N} 和 R_{PSN} 用于衡量图像的相似性或失真程度；F_{S} 和 D_{S} 反映图像的纹理丰富程度和对比度变化；F_{VI} 侧重人眼对图像质量的直观感受。选取分割结果与分割标签之间的交并比(intersection over union，I_{oU})和平均交并比(mean intersection over union，M_{IoU})作为图像分割客观评价指标，用于直观地衡量分割结果与真实标签之间的重叠程度。

3.2　消融实验

为验证融合模型中设计组件的有效性，分别对设计的跨域交互整合模块和损失函数进行消融实验。其中，损失函数的消融实验旨在探究内容损失掩模和语义损失对融合结果的影响。实验采用 MSRS 测试集的 361 对图像进行评价，主观定性和客观定量结果见图 2 和表 1。

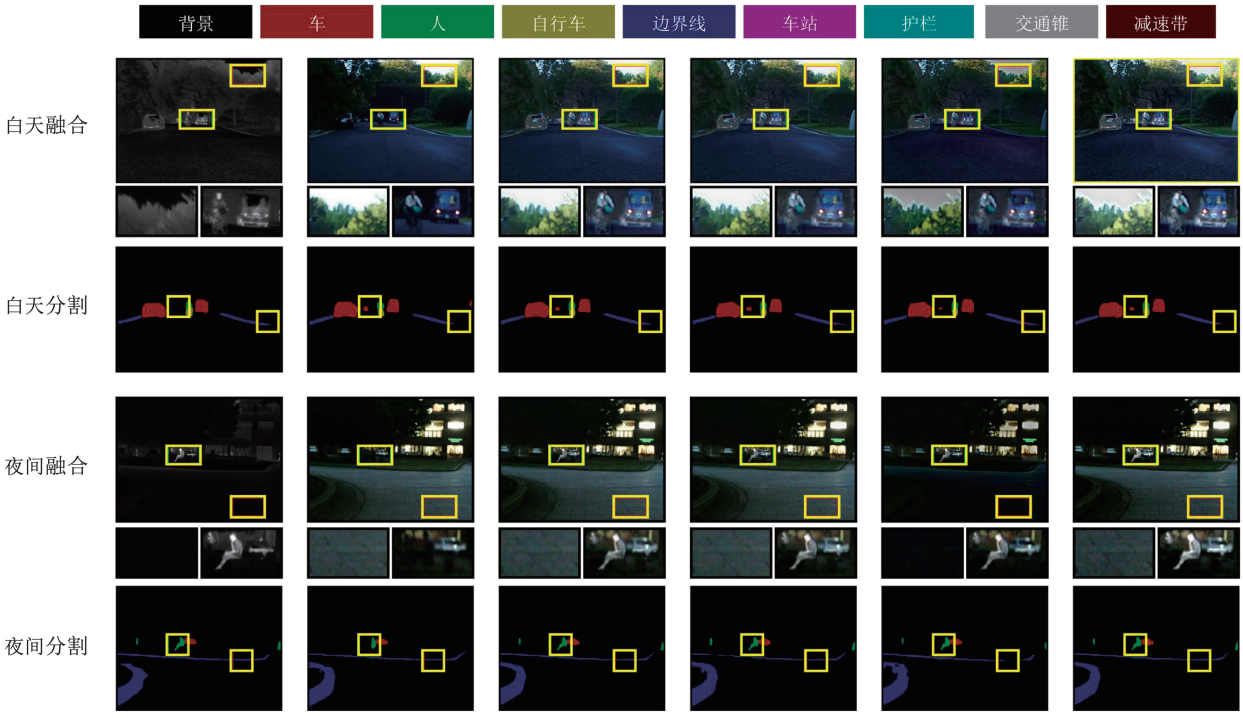


图 2 不同融合策略和不同损失函数在 MSRS 数据集上的定性结果

Fig. 2 Qualitative results of different fusion strategies and different loss functions on MSRS dataset

表 1 不同融合策略和不同损失函数在 MSRS 数据集上的定量结果

Tab. 1 Quantitative results of different fusion strategies and different loss functions on MSRS dataset

算法	$Q^{AB/F}$	E_N	R_{PSN}/dB	F_S	D_S	F_{VI}
去除跨域交互整合模块	**0.726**	6.649	64.142	11.496	41.795	1.046
去除语义损失	0.718	6.663	64.120	11.495	41.912	**1.059**
去除内容损失掩模	0.649	6.018	**65.678**	11.416	31.290	0.862
本文算法	0.719	**6.664**	64.158	**11.574**	**41.984**	1.049

由图 2 可以看出:在白天和夜间场景中,去除跨域交互整合模块的融合结果丢失了背景内容,如“树叶”边缘信息不明显;去除语义损失后的视觉效果有所改善,但纹理信息仍然较弱;去除内容损失掩模后,由于冗余红外信息的干扰,无论是白天还是夜间,整体的融合效果均不佳,对比度差,细节纹理弱。由表 1 可以看出:去除跨域交互整合模块会强化关键区域的信息,导致 $Q^{AB/F}$ 增加,但却忽略了局部关键信息和全局信息;去除语义损失后,人类视觉性能提高,但融合信息减少;去除内容损失掩模后,受到噪声的影响,整体结果均变差;本文算法的结果不仅保留了背景纹理信息,也突出了红外目标。

对比语义分割的结果,在白天场景中,去除跨域交互整合模块和去除语义损失的分割结果检测到了较多的“边界线”,去除内容损失的分割结果中“车”的内容检测不完整;在夜间场景中,去除跨域交互整合模块的分割结果中只检测到部分“车”,去除语义损失和内容损失的分割结果中“边界线”内容断断续续,检测不完整。因此,验证了去除任何模块都将或多或少地降低融合性能,证明了跨域交互整合模块和语义驱动在增强融合图像的视觉及表达语义信息能力方面的有效性。

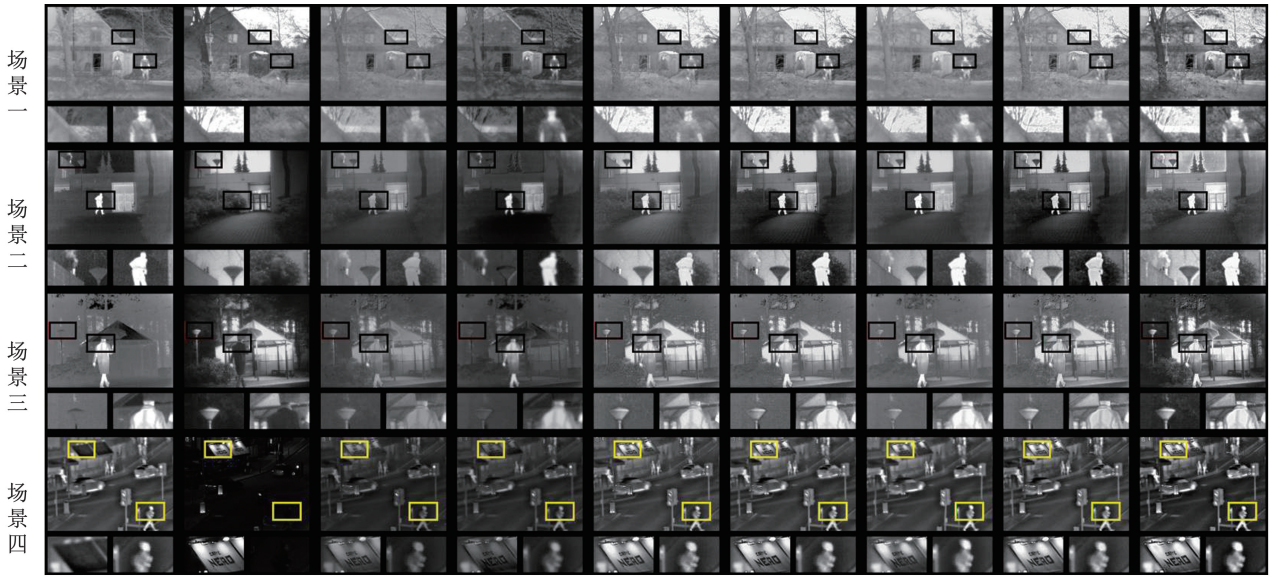
3.3 图像融合对比实验

3.3.1 TNO 数据集实验验证

选取 TNO 数据集中 42 对红外与可见光图像对 7 种方法进行主观和客观比较。

为深入分析融合效果,本文采用局部放大技术对关键区域进行可视化展示。图 3 为不同融合算法在 TNO 数据集上的融合定性结果中的 4 对结果。由图 3 中左侧和右侧矩形框标注的区域可以观察到以下现象:在场景一中,可见光图像因光照不足导致目标纹理模糊,且背景干扰严重;场景二中,复杂的背景信息对目标识别造成干扰;场景三中,阴雨天气导致行人轮廓模糊,细节信息丢失严重;场景四为夜

间图像，可见光中目标特征大部分丢失。由图 3 场景一的左侧矩形区域可以看出，在 DenseFuse、FusionGAN、PIAFusion、SeAFusion、SuperFusion、SDCFusion 算法的融合结果中，“树木”的纹理信息均有不同程度的丢失；由右侧矩形区域可以看出，所有算法的融合结果都保留了“人”的信息，但 DenseFuse、FusionGAN、PIAFusion、SuperFusion、SDCFusion 算法中“人”的细节边缘模糊，热红外目标的辐射亮度均不如本文算法显著。在场景二中，由左侧矩形区域可以看出，DenseFuse 和 FusionGAN 算法受噪声影响严重，“树枝”信息缺失，PIAFusion 和 SuperFusion 算法的可见光图像纹理缺失，无法清晰描述烟雾特征；由右侧矩形区域可以发现，DenseFuse、FusionGAN 和 PIAFusion 算法中“人”边缘模糊，SeAFusion 和 SDCFusion 算法的“人”的边缘出现伪影。在场景三中，由左侧矩形区域可以看出，DenseFuse、FusionGAN 和 SuperFusion 算法画面的对比度较差，“灯罩”边缘模糊；由右侧矩形区域可以看出，DenseFuse、FusionGAN 和 PIAFusion 算法“人”的边缘模糊，SeAFusion、SuperFusion 和 SDCFusion 算法的“行人衣服”纹理有不同程度的缺失。在场景四中，由左侧矩形区域可以看出，DenseFuse、FusionGAN 和 SuperFusion 算法的画面亮度对比不明显，“灯牌上的字母”显示不清；由右侧矩形区域可以看出，DenseFuse、FusionGAN、PIAFusion、SeAFusion、SuperFusion 和 SDCFusion 算法的融合结果中，“人”的边缘模糊，亮度较差。视觉对比结果表明，本文算法生成的融合图像在整体视觉效果上优于对比算法，特别是在复杂场景中，本文算法能够同时保持清晰的纹理细节和显著的热目标信息，避免了传统算法中常见的细节丢失或目标模糊问题。



(a) 红外图　(b) 可见光图　(c) DenseFuse　(d) FusionGAN　(e) PIAFusion　(f) SeAFusion　(g) SuperFusion　(h) SDCFusion　(i) 本文算法

图 3　不同融合算法在 TNO 数据集上的融合定性结果

Fig. 3　Qualitative fusion results of different fusion methods in TNO dataset

在 TNO 数据集上随机选取 40 对图像实验数据性能指标进行定量分析，结果见表 2。由表 2 可知，本文算法在 $Q^{AB/F}$、F_{VI} 上取得最优结果，E_N、R_{PSN} 和 D_S 取得次优结果。最优的 $Q^{AB/F}$ 体现了融合算法在边缘信息保存能力方面的优势，展现出优异的表达能力，使得融合图像呈现出了更丰富的场景信息；最优的 F_{VI} 说明融合算法能在特征保留与视觉效果上均表现出优势。因此，通过主观评估与客观指标分析，融合图像不仅完整保留了关键特征信息，同时在视觉质量上也优于对比方法，充分验证了所提算法的优越性。

3.3.2　MSRS 数据集实验验证

为全面评估所提出算法的融合效果，采用 MSRS 多光谱数据集开展系统的定性与定量实验验证。鉴于该数据集中的可见光图像采用 RGB 三通道色彩模式，为在融合过程中更好地保持源图像的色彩特征，首先对可见光图像进行色彩空间转换，即将 RGB 色彩空间转换为 YUV 色彩空间[16]，通过转换能够将图像信息分解为亮度、色度和浓度 3 个独立分量，不仅可以有效分离图像的亮度信息与色彩信息，还能在融合过程中更好地保留源图像的色彩特征，从而提高融合结果的视觉效果和信息完整性。

表 2 不同融合算法在 TNO 数据集上的客观指标定量结果比较

Tab. 2 Comparison of quantitative results of objective indicators of different fusion methods in TNO dataset

算法	$Q^{AB/F}$	E_N	R_{PSN}/dB	F_S	D_S	F_{VI}
DenseFuse	0.455	6.819	**62.574**	8.985	34.825	0.658
FusionGAN	0.234	6.558	60.979	6.275	30.633	0.422
PIAFusion	0.529	6.814	61.775	9.619	37.141	0.739
SeAFusion	0.487	**7.133**	61.391	**12.252**	**44.243**	0.704
SuperFusion	0.297	6.806	61.768	8.744	38.233	0.382
SDCFusion	0.548	7.058	61.367	12.124	39.944	0.706
本文算法	**0.567**	7.080	61.784	11.722	42.212	**0.768**

同样地,在 MSRS 数据集上采用局部放大技术对关键区域进行可视化展示,不同融合算法的融合定性结果中的 4 对结果如图 4 所示。其中,场景一、二为白天场景,场景三、四为夜间场景。由图 4(b)可以看出,受低光照条件影响,场景一、二中“行人”特征的纹理信息不明显;场景三、四中可见光中目标特征大部分丢失。由图 4 场景一的左侧矩形区域可以看出,DenseFuse、FusionGAN 和 SuperFusion 算法的“路面”纹理信息模糊,暗部失帧;由右侧矩形区域可以看出,DenseFuse、FusionGAN、PIAFusion、SuperFusion 和 SDCFusion 算法的“行人”边缘模糊,有伪影。由场景二左侧矩形区域可以看出,DenseFuse、FusionGAN、SuperFusion 和 SDCFusion 算法的“楼梯”细节丢失,SeAFusion 算法出现画面畸变;由右侧矩形区域可以看出,DenseFuse、FusionGAN 和 SuperFusion 算法的“行人”轮廓边缘模糊,PIAFusion 和 SeAFusion 算法中“树叶”出现伪影,画面过渡不够平滑。在场景三中,由于是夜间场景,DenseFuse、FusionGAN、SeAFusion 和 SuperFusion 算法的融合结果受噪声影响严重,“栏杆”边缘不清晰;在右侧矩形区域中,DenseFuse、FusionGAN、PIAFusion、SeAFusion、SuperFusion 和 SDCFusion 算法的“行人”目标有不同程度的边缘模糊,热红外信息保留不完全。在场景四中,由左侧矩形区域可以看出,DenseFuse 和 FusionGAN 算法的“交通锥”纹理丢失,PIAFusion、SeAFusion 和 SDCFusion 算法的“道路线”边缘信息丢失;由右侧矩形区域可以看出,DenseFuse、FusionGAN 和 SuperFusion 算法的“行人”边缘模糊,人像细节丢失,而 SeAFusion 和 SDCFusion 算法的亮部较暗,画面对比不明显。总体而言,本文算法生成的融合图像在目标显著性、细节清晰度和整体视觉效果方面均表现得更优,特别是在夜间场景和复杂背景条件下,本文算法能够同时保持清晰的热目标信息和丰富的场景细节,而且具有一定的场景恢复能力,整体融合效果优于对比算法。

(a) 红外图 (b) 可见光图 (c) DenseFuse (d) FusionGAN (e) PIAFusion (f) SeAFusion (g) SuperFusion (h) SDCFusion (i) 本文算法

图 4 不同融合算法在 MSRS 数据集上的融合定性结果

Fig. 4 Qualitative fusion results of different fusion methods in MSRS dataset

在 MSRS 数据集上，随机选取 40 对图像实验数据性能指标进行定量分析，结果见表 3。由表 3 可以看出，本文算法在 $Q^{AB/F}$、E_N 和 F_{VI} 指标上取得最优，在 F_S 指标上取得次优。说明本文算法能够有效保留红外与可见光图像的细节特征，实现了融合结果与源图像特征分布的高度一致性。此外，融合结果在视觉效果上更贴近真实场景，更符合人类视觉系统的感知特性。总体而言，本文算法较为完整地提取并增强了红外图像中的显著热目标，有效保留了可见光图像中的纹理和边缘信息，生成的融合图像具有良好的视觉效果和场景适应性。

表 3　不同融合算法在 MSRS 数据集上的客观指标定量结果比较

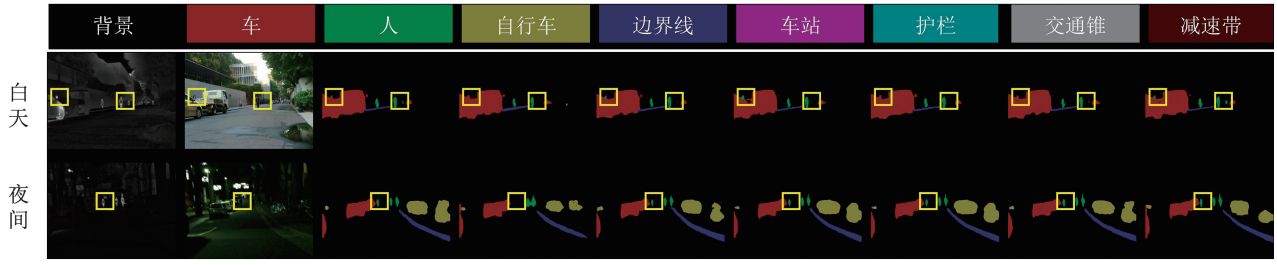
Tab. 3　Comparison of quantitative results of objective indicators of different fusion methods in MSRS dataset

算法	$Q^{AB/F}$	E_N	R_{PSN}/dB	F_S	D_S	F_{VI}
DenseFuse	0.050	5.936	61.759	6.025	23.567	0.047
FusionGAN	0.048	5.431	62.850	4.354	17.076	0.436
PIAFusion	0.666	6.637	64.127	**12.122**	**45.336**	1.016
SeAFusion	0.674	6.651	64.331	11.248	41.841	0.968
SuperFusion	0.558	6.582	**64.418**	10.639	42.315	0.808
SDCFusion	0.715	6.656	64.231	11.488	42.090	1.043
本文算法	**0.718**	**6.664**	64.157	11.474	41.984	**1.048**

在 TNO 和 MSRS 数据集上，本文算法在 $Q^{AB/F}$、E_N、R_{PSN}、F_S、D_S 和 F_{VI} 6 个客观评价指标上分别平均提高了 47.92%、6.15%、0.87%、44.31%、35.99%、36.88%，说明本文算法具有优越的视觉性能。

3.4　图像分割对比实验

为验证本文算法的语义信息保留和信息表达能力，通过语义分割任务验证其有效性。将 MSRS 数据集用于分割任务，利用 DeepLab v3 + 测试分割性能，并计算分割结果与分割标签之间的交并比。图 5为白天和夜间两个场景下的融合图像分割结果。可以看出：白天场景下，DenseFuse、FusionGAN、SuperFusion 和 SDCFusion 无法捕获车身附近“人”的信息，FusionGAN、SeAFusion 和 SuperFusion 无法捕获右侧边框中的“人”，DenseFuse、PIAFusion 和 SDCFusion 能检测到“人”的少量信息；夜间场景下，FusionGAN 丢失了“人”的信息特征。表 4 为分割性能指标比较。由表 4 可知，本文算法在所有类别中 I_{oU} 值均为第一，M_{IoU} 值最优，说明本文算法在语义分割任务上具有优越的性能。



(a) 红外图　(b) 可见光图　(c) DenseFuse　(d) FusionGAN　(e) PIAFusion　(f) SeAFusion　(g) SuperFusion　(h) SDCFusion　(i) 本文算法

图 5　不同融合算法在 MSRS 数据集上的分割结果

Fig. 5　Segmentation results for different fusion methods in MSRS dataset

表 4　不同融合算法在 MSRS 数据集上的图像分割性能指标比较

Tab. 4　Comparison of image segmentation performance metrics for different fusion methods in MSRS dataset　%

算法	I_{oU}									M_{IoU}
	背景	车	人	自行车	边界线	车站	护栏	交通锥	减速带	
DenseFuse	98.25	87.65	67.42	70.70	61.76	71.10	76.52	55.69	74.02	73.67
FusionGAN	98.25	87.35	66.92	70.09	61.99	65.66	73.85	55.30	69.49	71.47
PIAFusion	98.23	87.64	67.29	70.53	61.72	70.70	70.60	56.52	72.62	72.87
SeAFusion	98.23	87.77	67.54	70.59	61.13	70.60	73.38	56.49	71.63	73.04
SuperFusion	98.06	86.31	66.57	69.92	59.66	67.92	76.25	53.49	66.25	71.60
SDCFusion	98.23	87.88	67.55	70.28	61.31	70.68	70.40	57.20	72.99	72.95
本文算法	**98.45**	**89.10**	**69.67**	**73.05**	**65.04**	**74.66**	**76.99**	**61.01**	**78.35**	**76.26**

4 结 论

1)提出了一种基于语义驱动的红外与可见光图像交互融合算法,通过级联融合网络和分割网络,设计跨域交互整合模块,并引入语义损失函数,有效解决了图像融合中像素信息和语义信息不平衡问题,促进了融合图像在高级视觉任务中的鲁棒性。

2)通过联合操作,将融合网络和分割网络整合在一个联合框架下,有效联合优化了上下游任务。

3)设计了跨域交互整合模块,使得算法能够有效捕捉红外与可见光图像的局部关键信息和全局信息,从多个角度解决了跨模态全局信息的相关性问题。

4)在 TNO 和 MSRS 红外与可见光图像数据集上的融合实验表明,本文算法在 $Q^{AB/F}$、E_N、R_{PSN}、F_S、D_S 和 F_{VI} 6 个客观评价指标上分别平均提高了 47.92%、6.15%、0.87%、44.31%、35.99%、36.88%,融合结果在热目标特征保留与背景细节呈现方面均表现出显著优势,能够实现多源数据的有效互补与协同增强。分割实验表明,本文算法在所有评价指标中均取得了最优结果,在语义分割任务中也展现出了优越的性能,整体性能优于其他 6 种现有的图像融合算法。

参考文献

[1]李晓玲, 陈后金, 李艳凤, 等. 多重关系感知的红外与可见光图像融合网络[J]. 电子与信息学报, 2024, 46(5): 2217
LI Xiaoling, CHEN Houjin, LI Yanfeng, et al. Infrared and visible image fusion network with multi-relation perception[J]. Journal of Electronics & Information Technology, 2024, 46(5): 2217. DOI: 10.11999/JEIT231062

[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]吕品, 李凯, 许嘉, 等. 无人驾驶汽车协同感知信息传输负载优化技术[J]. 计算机学报, 2021, 44(10): 1984
LYU Pin, LI Kai, XU Jia, et al. Cooperative sensing information transmission load optimization for automated vehicles[J]. Chinese Journal of Computers, 2021, 44(10): 1984. DOI:10.11897/SP.J.1016.2021.01984

[4]陈咸志, 罗镇宝, 李艺强, 等. 自动目标识别在图像末制导中的应用[J]. 红外与激光工程, 2022, 51(8): 266
CHEN Xianzhi, LUO Zhenbao, LI Yiqiang, et al. Application of automatic target recognition in image terminal guidance[J]. Infrared and Laser Engineering, 2022, 51(8): 266. DOI: 10.3788/IRLA20220391

[5]LI Hui, WU Xiaojun. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614. DOI:10.1109/TIP.2018.2887342

[6]JIAN Lihua, YANG Xiaomin, LIU Zheng, et al. SEDRFuse: a symmetric encoder-decoder with residual block network for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1. DOI:10.1109/TIM.2020.3022438

[7]TANG Linfeng, YUAN Jiteng, ZHANG Hao, et al. PIAFusion: a progressive infrared and visible image fusion network based on illumination aware[J]. Information Fusion, 2022, 83: 79. DOI: 10.1016/j.inffus.2022.03.007

[8]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:110.1016/j.inffus.2018.09.004

[9]LI Kaixin, LIU Gang, GU Xinjie, et al. DANT-GAN: a dual attention-based of nested training network for infrared and visible image fusion[J]. Digital Signal Processing, 2024, 145: 104316. DOI:10.1016/j.dsp.2023.104316

[10]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

[11]LIU Jinyuan, FAN Xin, HUANG Zhanbo, et al. Target-aware dual adversarial learning and a multi-scenario multi-modality benchmark to fuse infrared and visible for object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orlean: IEEE, 2022: 5792. DOI:10.1109/CVPR52688.2022.00571

[12]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

[13]MA Jiayi, TANG Linfeng, FAN Fan, et al. SwinFusion: cross-domain long-range learning for general image fusion via swin transformer[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(7): 1200. DOI:10.1109/JAS.2022.105686

[14]TOET A. The TNO multiband image data collection[J]. Data in Brief, 2017, 15: 249. DOI:10.1016/j.dib.2017.09.038

[15]TANG Linfeng, DENG Yuxin, MA Yong, et al. SuperFusion: a versatile image registration and fusion network with semantic awareness[J]. IEEE/CAA Journal of Automatica Sinica, 2022, 9(12): 2121. DOI:10.1109/JAS.2022.106082

[16]杨帆, 王志社, 孙婧, 等. 红外与可见光图像交互自注意力融合方法[J/OL]. 光子学报. (2024-05-10)[2024-06-20]. http://kns.cnki.net/kcms/detail/61.1235.O4.20240509.0906.010.html
YANG Fan, WANG Zhishe, SUN Jing, et al. Infrared and visible image fusion method via interactive self-attention[J/OL]. Acta Photonica Sinica. (2024-05-10)[2024-06-20]. http://kns.cnki.net/kcms/detail/61.1235.O4.20240509.0906.010.html