液晶与显示
Chinese Journal of Liquid Crystals and Displays

文章编号:1007-2780(XXXX)XX-0001-12

目标语义分层驱动的红外和可见光图像融合

陶俣卓,罗玉婷,赵　凡*
(辽宁师范大学 物理与电子技术学院,辽宁 大连 116000)

摘要:红外和可见光图像融合旨在保留不同模态的互补特征以实现对复杂场景的鲁棒感知,在安防监控、军事侦察、自动驾驶等众多领域都有着至关重要的应用价值。但现有的图像融合算法侧重增强图像的视觉效果,导致关键语义信息在融合过程中未被有效保留,进而影响了融合图像在高级视觉任务中的应用效果。虽然现有方法尝试将融合任务与高级视觉任务(分割,检测等)级联,但这种顺序连接对语义信息的增强作用有限。为了兼顾视觉效果和下游任务,本文提出了一种目标语义分层驱动的红外和可见光图像融合网络SDFusion。首先采用共享特征编码器对红外与可见光图像进行多层级跨模态特征提取,然后通过图像融合解码器与语义分割解码器的并行协同优化,同时将编码特征分层注入到解码特征以增强特征表示,实现融合特征与语义特征的联合建模。在融合性能上,通过实验表明,与传统方法相比本方法在五个客观评价指标EN,SD,MI,VIFF,$Q^{AB/F}$上分别提升了3.7%,7.3%,45.3%,18.5%,7.2%。另外,本方法的融合结果较传统方法在下游分割任务上展示出了更好的性能。这些实验充分证明了SDFusion方法的有效性,其融合结果不仅实现了视觉效果的明显提升,还极大地促进了高级视觉任务的开展,为红外和可见光图像融合技术的发展提供了新的思路和方法。

关　键　词:红外和可见光图像融合;语义驱动;协同优化
中图分类号:TP394.1;TH691.9　**文献标识码**:A　**doi**:10.37188/CJLCD.2025-0141　**CSTR**:32172.14.CJLCD.2025-0141

Target semantic hierarchy driven infrared and visible image fusion

TAO Yuzhuo, LUO Yuting, ZHAO Fan*

(*College of Physics and Electronic Technology*, *Liaoning Normal University*, *Dalian* 116000, *China*)

Abstract: Infrared and visible image fusion aims to retain the complementary features of different modalities to achieve robust perception of complex scenes, and it plays a crucial role in numerous fields such as security monitoring, military reconnaissance, and autonomous driving. However, existing image fusion algorithms focus on enhancing the visual effects of images, leading to the ineffective preservation of key semantic information during the fusion process, which in turn affects the application performance of fused images in high-level visual tasks. Although existing methods attempt to cascade the fusion task with high-level visual tasks (segmentation, detection, etc.), this sequential connection has limited enhancement on semantic information. To balance visual effects and downstream tasks, this paper proposes a semantic-driven infrared and visible image fusion network, SDFusion. First, a shared feature encoder is adopted to perform multi-level cross-modal feature extraction for infrared and visible images,

收稿日期:2025-07-03;**修订日期**:2025-08-23.
基金项目:国家自然科学基金(No.62001450)
Supported by National Natural Science Foundation of China (No.62001450)
*通信联系人,E-mail:3175209736@qq.com

fully mining the useful information in both modal images. Then, through the parallel collaborative optimization of the image fusion decoder and the semantic segmentation decoder, while hierarchically injecting encoded features into decoded features to enhance feature representation, the joint modeling of fusion features and semantic features is achieved. Experimental results on public datasets show that compared with traditional methods, this method significantly improves seven objective evaluation indicators. Specifically, EN is improved by 3.7%, SD by 7.3%, MI by 45.3%, VIFF by 18.5%, and $Q^{AB/F}$ by 7.2%. The fusion results of this method demonstrate superior performance in downstream segmentation tasks compared to traditional approaches. These experiments fully demonstrate the effectiveness of the SDFusion method. The fusion results not only achieve obvious improvement in visual effects but also greatly promote the development of high-level visual tasks, providing new ideas and methods for the development of infrared and visible image fusion technology.

Key words: infrared and visible image fusion; semantic-driven; collaborative optimization

1 引　　言

红外与可见光图像融合技术在复杂视觉感知任务中具有重要应用价值。红外图像通过热辐射特性可穿透烟雾、雾霾等干扰,在低光照或无光照条件下清晰呈现目标物体的热分布信息,但其空间分辨率低且缺乏纹理细节;可见光图像依托光学反射信息能够捕捉丰富的场景结构与表面细节,但对光照变化敏感,在逆光、夜间或遮挡场景中易丢失关键目标信息。这两种图像的互补特性促使研究人员融合红外和可见光图像,并生成兼具目标显著性与纹理细节的融合图像。具有良好的视觉效果和丰富的信息的图像融合能够广泛应用于各种下游识别任务,例如目标检测[27]、语义分割[28]等等。

传统融合框架[1-3]一般通过人工预先制定多模态数据特征提取和整合策略以实现图像融合,导致在复杂动态场景下难以有效协调跨模态特征间的深层关联与互补性;随着深度学习的快速发展很多研究人员将其引入到图像融合领域,这些方法虽然往往具有很好的视觉效果,但它们只追求更高的融合评价指标,忽略了图像融合最重要的意义在于能否更好的服务于高级视觉任务;在此之后一些研究人员尝试在图像融合任务中引入语义信息,比如[7]利用分割掩码突出目标区域,但掩码局限性较大,对语义信息的增强效果有限;一些算法[5][7][8][9]意识到分割网络能够提供足够的语义信息,采用了在图像融合网络后级联语义分割网络的方式以促进高级视觉任务,具体地说,将融合网络的输出结果输入分割网络,利用分割网络的语义损失促使融合网络生成适合分割任务的融合图像。但此类方法语义反馈信号对融合过程的引导的效率较差,并没有充分利用分割网络提供的语义信息,且语义特征与视觉特征交互的动态性不足。

针对以上问题,本文提出一种新的目标语义分层驱动红外和可见光图像融合网络,本融合网络最终同时输出融合结果和分割结果,采用双任务联合约束网络进行特征提取的策略,让分割任务也参与到图像融合过程中,不仅仅作为一个后验约束,极大增强了融合图像的语义信息。且图像融合和语义分割是两个不同级别的任务,本方法同时巧妙弥合了两者之间的特征差距。综上所述,本文主要贡献如下:

(1)提出一种目标语义分层驱动的红外和可见光图像融合框架SDFusion,该框架能够在最大程度上实现语义信息对图像融合任务的促进。

(2)利用语义分割和图像融合双任务联合约束图像特征提取,有效克服了传统级联架构中存在的语义反馈滞后等问题,使算法在语义特征和视觉特征保留上均有较好的效果。

(3)在多个公开数据上验证了方法的有效性,融合结果在视觉质量和指标评估上表现优异。

2 相关工作

2.1 传统的融合方法

传统的红外与可见光图像融合方法主要基

于人工设计的特征提取与融合规则,以增强图像的信息互补性。多尺度变换方法[29]是早期主流技术,Chen等人[1]遵循管道并利用拉普拉斯金字塔变换来进行多尺度特征分解。小波变换[2]也是常见的多尺度变换的方法,Zhan等人采用小波变换[2]分解策略,通过分离图像的低频(全局结构)与高频(细节纹理)成分,并采用能量最大化、区域方差加权等规则融合不同尺度的特征信息。然而,这类方法对分解层数与融合规则敏感,易引入伪影或细节丢失。稀疏表示方法[3]通过构建完备字典对图像块进行稀疏编码,利用系数融合保留红外目标的显著性强度与可见光的细节特征,但其性能受限于字典的完备性,且稀疏求解过程计算复杂度较高;子空间学习方法则通过将图像投影到低维潜在空间实现特征解耦,例如将红外图像的稀疏热目标与可见光的低秩背景分离后重组,在正交子空间中融合方差最大的主成分;但这类方法依赖线性或低秩假设,难以捕捉复杂场景下的非线性模态关联。尽管传统方法在特定场景下效果显著,但人工定义的特征提取规则和融合策略较为复杂且会导致模型泛化能力弱,促使研究者转向数据驱动的深度学习范式。

2.2 基于深度学习的融合方法

近年来,基于深度学习的图像融合方法通过端到端的特征学习与自适应融合策略,显著提升了红外与可见光图像融合的质量与效率。早期研究以卷积神经网络(CNN)为核心,例如编码器-解码器架构[11-14]被广泛用于多层级特征提取与重建,通过卷积核的局部感知特性捕捉跨模态的纹理细节与热目标关联。然而,传统CNN方法对全局上下文建模能力不足,易导致显著性目标模糊或背景细节丢失。生成对抗网络(GAN)通过对抗训练策略优化融合结果的视觉自然性与任务适配性。典型方法如FusionGAN[15]与DDcGAN[16],其生成器通过重构损失与对抗损失生成细节丰富的图像,而判别器则约束融合结果与可见光图像的纹理一致性及红外目标的可辨识性。然而,GAN的稳定性与模态均衡性仍面临挑战。近年来,Transformer凭借全局长程建模能力,在图像融合领域起到了重要作用[17-20],SwinFusion[18]基于Transformer设计了一个基于自注意的域内融合单元和一个基于交叉注意的域间融合单元,分别对同一域内和跨域内的长程相关性进行建模和融合;PPT fusion[17]则设计了一个金字塔图像块Transformer作为低层视觉任务的通用特征提取模块,补充了Transformer在CV领域的局部特征提取能力。上述方法在提高红外和可见光图像融合视觉效果上取得了不错的成绩,但任务导向型不足,没有考虑融合之后与下游任务的适配性。

2.3 高级视觉任务驱动的融合方法

为解决红外和可见光图像融合缺乏语义信息的问题,很多工作已经尝试用高级视觉任务(语义分割,目标检测等)驱动融合网络强制保留语义信息。Tang等人[5]首次在图像融合任务后添加了分割任务并进行联合训练,通过引入语义损失指导图像融合过程,以实现增强融合图像语义信息的目的。类似地,Sun等人[6]设计了检测驱动的融合网络,将融合图像输入到检测网络中,以激励融合网络从目标检测网络中学习目标的特定信息。Liu等人[7]在此前研究基础上实现了图像融合与分割任务的交互促进的SegMiF网络,SegMiF同样采用在融合网络后面附加一个分割模型策略,通过巧妙地桥接两个分量之间的中间特征,从分割任务中学习的知识可以有效地辅助融合任务;同时,有益的融合网络支持分割网络更好地发挥作用。这类方法通过任务协同优化,不仅提升了融合图像在目标结构清晰度、边缘一致性等方面的表现,更显著改善了其在分割、检测等任务中的可用性;但此类级联的框架导致下游任务仅作为后验约束,未能实现特征层级的深度融合且模型泛化能力差,没有最大化利用分割或者检测任务对图像融合的促进作用,因此急需一种新的红外和可见光图像融合的方法,能够使融合图像获得更多的语义信息。

3 方 法

在本节将详细介绍本论文提出的目标语义分层驱动红外和可见光图像融合SDFusion的方法。首先,阐述提出此方案的动机,随后详细介绍了网络架构,最后介绍相关损失函数。

3.1 动机

在图像融合中引入语义信息能够促使图像

更好的应用于下游的高级视觉任务。针对现有结合语义信息的红外和可见光图像融合方法存在语义特征和视觉特征交互动态性不足，对于分割网络提供的语义信息利用不充分等问题，提出语义信息驱动的红外和可见光图像融合网络。与以往将融合网络与分割网络串联的框架不同，本网络架构采用了双任务指导特征提取的策略，在图像融合任务与语义分割任务的联合约束下，网络能够实现在提取丰富的视觉特征同时保留足够的语义信息，在避免视觉特征和语义特征域差问题的同时最大化了语义分割任务对融合图像在高级视觉任务中的促进作用，极大增强了融合图像的语义信息。整体网络如图1所示。随后将在3.2中详细介绍网络结构。

3.2 网络架构

整体网络框架由一个共享特征编码器，一个图像融合解码器，一个语义分割解码器组成，将红外和可见光图像经过通道维度拼接输入共享特征编码器，用双解码器约束共享特征编码器进行特征提取，同时将每层编码器特征跳跃连接到解码器特征，最后分别输出融合结果和分割结果，促使共享特征编码器能够提取足够的语义信息和视觉特征，生成既满足人类视觉观察，又有利于机器视觉感知的融合结果。为了保证能够取得良好的分割结果，本网络架构基于unet[21]实现。

共享特征编码器(SFE)：共享特征编码器SFE用于提取图像视觉纹理特征及语义信息，包含5个模块，每个模块由2个3×3卷积核构成，输出通道数分别为32，64，128，256，512；同时为了减少计算量，每个模块之间通过最大池化下采样*max pool*连接以减小特征图尺寸，*max pool*大小设置为2×2；每个模块的输出不仅作为下一个模块的输入，还通过跳跃连接到解码器的相应模块帮助其更好的利用不同层次的特征信息。最后一个模块输出的特征图被分别输送到图像融合解码器和语义分割解码器用于融合结果和分割结果重建。

图像融合/语义分割解码器(IFD/SSD)：图像融合/语义分割解码器(IFD/SSD)分别用于重建融合结果和分割结果，监督共享特征编码器进行特征提取。两个解码器采用相同的结构，由四个模块组成，同样每个模块包含2个3×3卷积核，输出通道数依次为256，128，64，32，每个模块之间采用上采样卷积up-conv恢复特征图尺寸，大小设置为2×2；，同时解码器将整合通过跳跃连接输入的相同尺寸的特征以提高分割准确性和图像细节特征的保留能力；最终的输出采用1×1卷积重建结果。

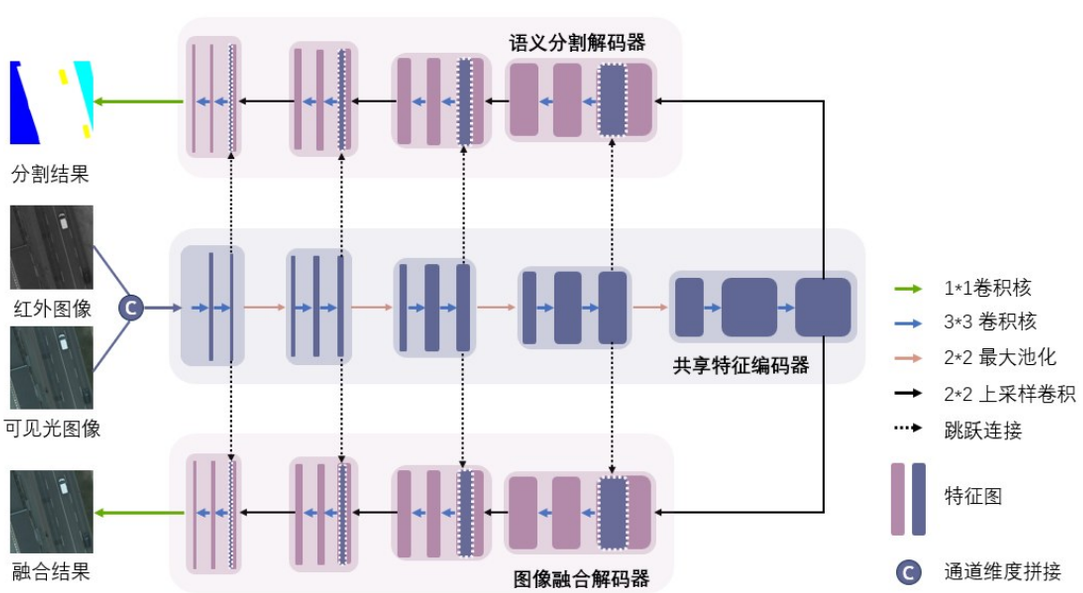

图1 本模型框架

Fig. 1 Model framework

3.3 损失函数

本框架采用多任务联合优化策略,需要同时提取图像纹理特征以及语义信息,通过融合任务与分割任务的协同训练提升模型性能,因此需要两部分损失函数。总体损失函数由融合损失L_{fuse}和分割损失L_{seg}加权组成,定义如下:

$$L_{total}=L_{fuse}+\alpha L_{seg}\ , \tag{1}$$

其中α为权衡两项任务重要性的超参数

3.3.1 融合损失

融合损失旨在提升红外与可见光图像特征融合的质量,因此采用结构相似性损失与强度损失联合约束融合结果,公式化为:

$$L_{fuse}=L_{ssim}+\alpha L_{seg}\ . \tag{2}$$

*SSIM*损失:考虑到结构相似性(*SSIM*)指标是最广泛使用的度量,能够测量融合结果与源图像在亮度、对比度和结构方面的相似性,采用*SSIM*损失L_{ssim}约束融合图像I_{fuse}和红外图像I_{ir}、可见光图像I_{vi}的结构相似性。同时考虑最大选择策略可以有效地聚合源图像中的纹理细节,具体而言*SSIM*损失定义为:

$$L_{ssim}=1-SSIM\left(I_{fuse},max\left(I_{ir},I_{vi}\right)\right), \tag{3}$$

其中$SSIM(\cdot)$表示进行结构相似性计算,$max(\cdot)$指逐元素最大选择。

强度损失:除了纹理细节,亮度信息保留也是图像融合任务需要考虑的重点,为了最大程度捕获图像强度信息,设置以下强度损失函数L_{int}:

$$L_{int}=\frac{1}{HW}\left\| I_{fuse}-max(I_{ir}-I_{vi})\right\|_1, \tag{4}$$

H和W分别代表图像的高度和宽度,$\|\cdot\|_1$表示l_1范数。

3.3.2 分割损失

为了能够解决分割任务类别不平衡与边界模糊问题,设计交叉熵损失L_{se}和*dice*损失L_{dice}联合约束分割结果,具体表示如下:

$$L_{seg}=L_{se}+L_{dice}\ . \tag{5}$$

交叉熵损失*Cross Entropy Loss*:交叉熵损失是常用的语义分割损失,其通过衡量模型预测概率分布与真实标签分布之间的差异,驱动网络优化语义分割的准确性,为此,设计以下交叉熵损失函数L_{se}:

$$L_{se}=cross\_entropy\_loss(\tilde{M},M)\ , \tag{6}$$

其中$cross_entropy_loss(\tilde{M},M)$为为交叉熵损失计算,$M$为分割真值,$\tilde{M}$为网络输出分割。

*dice*损失:为了优化分割边界,缓解类别不平衡问题,在交叉熵损失的基础上设置了*dice*损失,将L_{dice}写为:

$$L_{dice}=dice\_loss(\tilde{M},M)\ , \tag{7}$$

$dice_loss(\cdot)$为计算*dice*损失。

4 实 验

在本节中,首先给出了一些实配置和实验细节。随后,从定性和定量的角度比较了各种算法在几个数据集上的融合性能。之后,通过分割模型验证了该算法在高级视觉任务中的优势。最后通过消融实验验证了设计模块的有效性。

4.1 实验细节

本次实验基于pytorch框架实现,在单个NVIDIA 2080Ti GPU上训练,网络参数使用Adam优化器更新,批次大小和初始学习率分别设置为4和1×10^3,对于式(1)超参数α,将其设置为0.2。(见4.4消融实验)。

本次实验采用了Potsdam数据集进行训练,在TNO数据集、Roadscene数据集和Potsdam数据集上验证了融合性能。对比实验包括包括SwimFusion[18],YDTR[25],LRRNet[26],U2Fusion[24],PIAFusion[23],CrossFuse[30],TC-MoA[31]。此外,我们选用了SegFormer分割网络以测量Potsdam融合图像包含的语义信息。

本次实验使用五个度量标准来定量地衡量融合结果,分别为熵(EN),标准差(SD),互信息(MI),视觉信息保真度(VIF)和$Q^{AB/F}$。度量越高,表示融合图像越好,详细信息参见[22]。此外,像素交集(IoU)被用来量化分割性能。Potsdam数据集涉及六类对象,包括背景,建筑,低矮植被,树,车,杂物。

4.2 融合实验

4.2.1 定性比较分析

为了直观地观察各种算法融合性能的差异,本次实验从TNO数据集和Roadscene数据集中选取2对有代表性的图像,可视化结果见图2、图3和图4、图5所示。在图2中,SwinFusion,YDTR,LRRNet,U2Fusion,CrossFuse没有很好的呈现出房屋,PIAFusion和TC-MoA虽然可

以显示房屋，但相比本方法具有行人不突出，纹理较差等问题。在图3中，本方法与其他方法相比城墙与背景的对比度更强且城墙轮廓更清晰。在图4中，其他算法墙上的文字均没有本算法更清楚，说明本算法在真对可见光纹理细节保留上做的更好，且其他算法行人均不明显，说明存在红外图像热辐射信息缺失的问题。在图5中其他算法存在背景模糊不清晰等问题，本算法在整体视觉效果上表现得更好，更符合视觉感知。

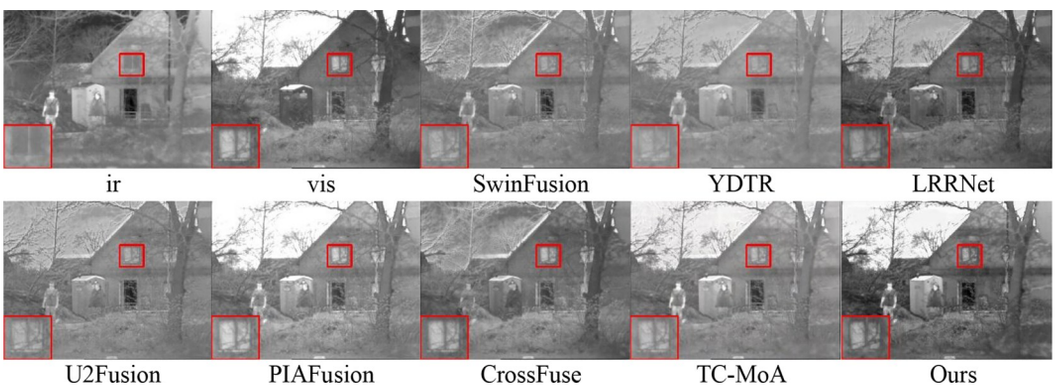

图2 基于TNO数据集的不同融合算法的视觉效果

Fig. 2 Visual effects of different fusion algorithms based on the TNO dataset

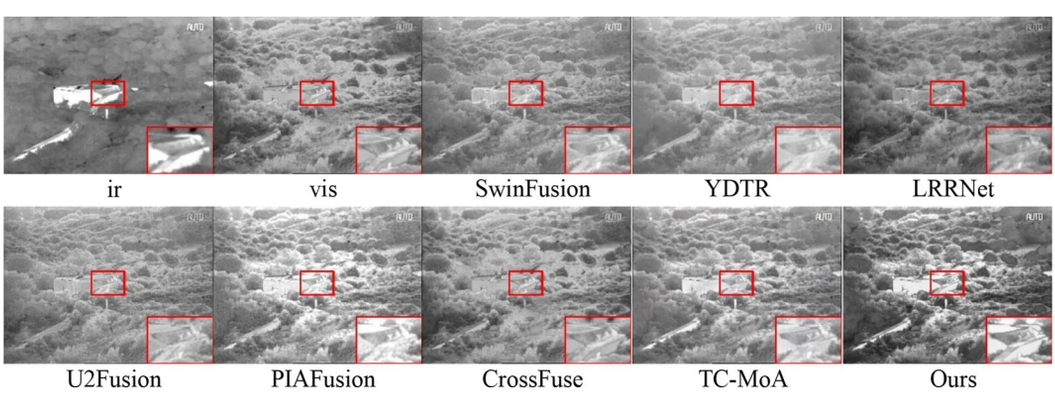

图3 基于TNO数据集的不同融合算法的视觉效果

Fig. 3 Visual effects of different fusion algorithms based on the TNO dataset

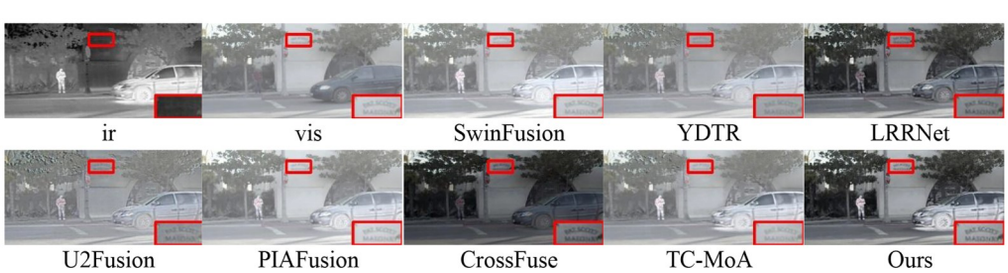

图4 基于Roadscene数据集的不同融合算法的视觉效果

Fig. 4 Visual effects of different fusion algorithms based on the RoadScene dataset

此外，为了验证算法的普适性从Potsdam数据集中另外选取了两张图片进行了测试，见图6和图7，在两组图中我们的方法展现了更好的视觉对比度，且强化了纹理信息的表达，证明了算法在能够使用于多种不同场景的数据集。

通过上述比较，可以发现本方法在保留红外

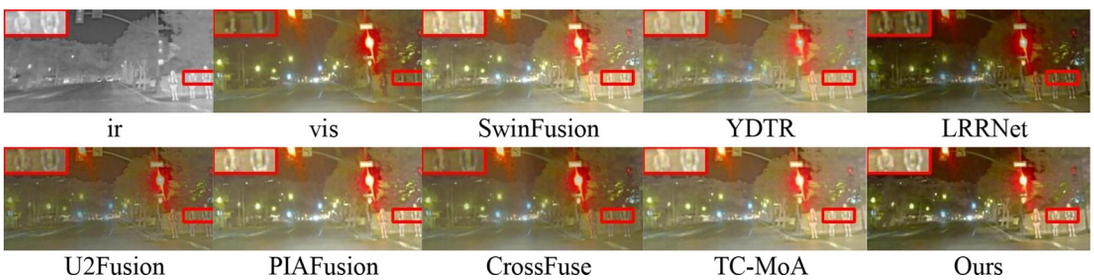

图5 基于Roadscene数据集的不同融合算法的视觉效果

Fig. 5 Visual effects of different fusion algorithms based on the RoadScene dataset

图像热辐射信息和可见光图像纹理信息上做的比其他算法更好,且目标信息与背景区域的对比度更强,证明了本论文采用的图像融合与语义分割联合驱动的网络在视觉效果上的优异性能。

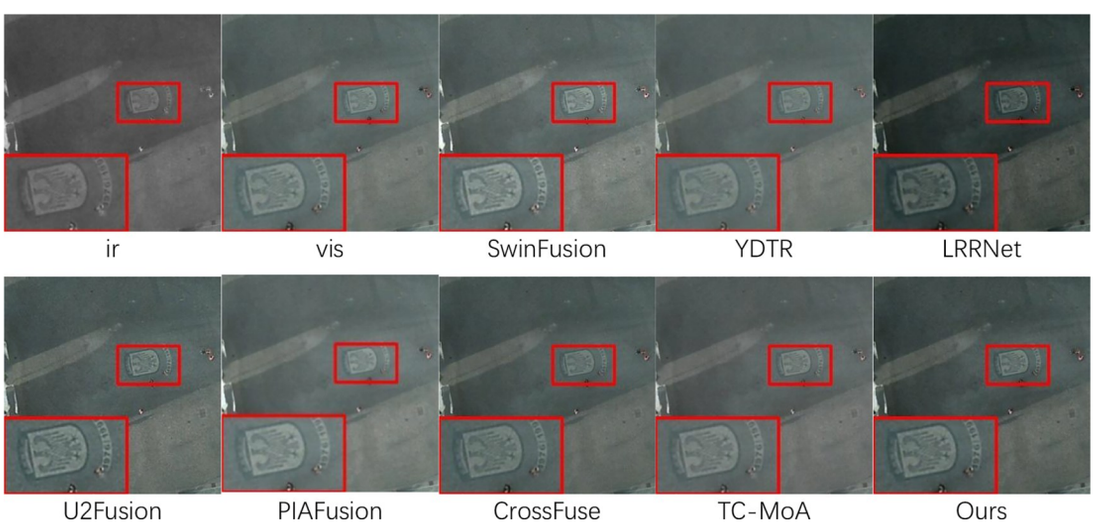

图6 基于Potsdam数据集的不同融合算法的视觉效果

Fig. 6 Visual effects of different fusion algorithms based on the Potsdam dataset

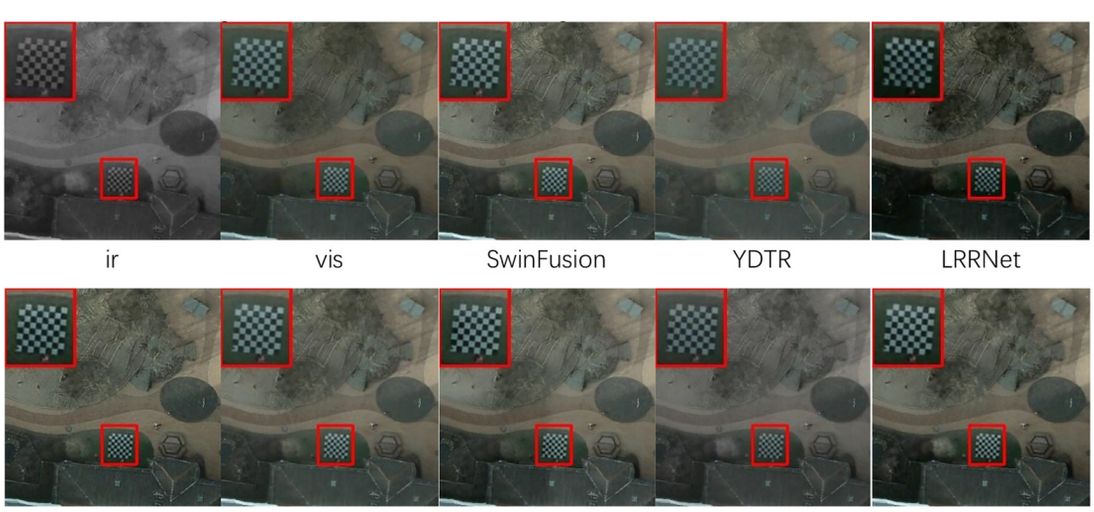

图7 基于Potsdam数据集的不同融合算法的视觉效果

Fig. 7 Visual effects of different fusion algorithms based on the Potsdam dataset

4.2.2 定量比较分析

在表1、表2和表3上分别评估了TNO数据集、Roadscene数据集和Potsdam数据集的融合结果在五个指标上的度量。本方法几乎在所有数据集上的度量指标都是最优和次优，从定量角度更加证明了该方法取得了良好的融合效果且适用于多种情况下的融合任务。另外，将本模型推理时间和其他方法进行了比较见表4，本模型的推理时间为0.130 6 s，仅次于YDTR和LRRNet，说明本模型无论在融合效果和推理时间上都具有一定优势。

表1 基于TNO数据集的不同融合算法的评价指标。粗体和下划线分别表示最佳值和次佳值。

Tab. 1 Evaluation metrics of different fusion algorithms based on the TNO dataset. Bold and underline indicate the best and second-best values, respectively.

	EN	SD	MI	VIFF	$Q^{AB/F}$
SwinFusion	6.886 5	39.612 8	2.599 2	0.716 0	0.546 7
YDTR	6.320 5	26.704 0	1.715 7	0.544 7	0.398 3
LRRNet	7.009 4	42.697 8	1.957 6	0.554 6	0.404 0
U2Fusion	6.835 8	34.762 4	1.477 2	0.540 1	0.461 7
PIAFusion	6.921 0	41.351 4	2.386 7	0.773 6	0.554 9
CrossFuse	6.973 0	41.697 2	2.223 1	0.718 6	0.470 2
TC-MoA	6.876 7	38.113 5	1.992 8	0.652 2	0.532 9
Ours	**7.120 9**	**45.833 3**	**3.771 4**	**0.916 2**	**0.594 9**

表2 基于Roadscene数据集的不同融合算法的评价指标。粗体和下划线分别表示最佳值和次佳值。

Tab. 2 Evaluation metrics of different fusion algorithms based on the Roadscene dataset. Bold and underline indicate the best and second-best values, respectively.

	EN	SD	MI	VIFF	$Q^{AB/F}$
SwinFusion	7.054 7	44.838 0	2.377 3	0.675 2	0.498 6
YDTR	6.776 1	33.238 2	2.137 9	0.616 5	0.468 0
LRRNet	7.002 1	40.067 6	1.977 7	0.494 1	0.348 3
U2Fusion	6.712 9	29.031 4	1.671 4	0.537 5	0.479 7
PIAFusion	6.982 6	44.073 3	2.480 3	0.697 0	0.466 7
CrossFuse	7.123 1	45.329 1	2.315 6	0.604 0	0.367 4
TC-MoA	7.124 1	42.274 7	2.284 2	0.672 5	**0.537 7**
Ours	**7.466 3**	**58.912 7**	**3.352 9**	**0.850 1**	0.521 0

表3 基于Potsdam数据集的不同融合算法的评价指标。粗体和下划线分别表示最佳值和次佳值。

Tab. 3 Evaluation metrics of different fusion algorithms based on the Potsdam dataset. Bold and underline indicate the best and second-best values, respectively.

	EN	SD	MI	VIFF	$Q^{AB/F}$
SwinFusion	6.892 4	**38.380 1**	3.320 6	1.563 3	0.733 3
YDTR	6.466 3	26.889 3	3.178 5	1.347 7	0.801 2
LRRNet	6.757 7	34.347 7	2.782 9	1.132 1	0.599 0
U2Fusion	6.793 2	32.101 4	2.743 2	1.448 7	0.696 2
PIAFusion	6.845 5	35.673 9	3.471 1	1.551 0	0.793 2
CrossFuse	6.638 4	28.873 0	2.830 4	1.341 1	0.751 1
TC-MoA	6.707 4	31.048 5	3.328 1	0.652 2	0.532 9
Ours	**6.920 3**	36.132 9	**4.143 0**	**1.579 6**	**0.812 7**

表 4 不同融合算法在推理时间上的比较

Table 4 Comparison of different fusion algorithms in terms of inference time

	SwinFusion	YDTR	LRRNet	U2Fusion	PIAFusion	CrossFuse	TC-MoA	ours
推理时间/s	0.571 6	0.105 4	0.058 1	0.654 6	0.267 4	0.304 4	0.162 8	0.130 6

4.3 分割实验

4.3.1 定性比较分析

为了证明本算法对下游任务的促进作用,基于在Potsdam数据集上的融合结果用于分割任务,segformer分割网络被重新训练以证明所提出算法的优越性能。图8为Potsdam数据集上分割结果的可视化,可以发现我们的算法在低矮植被和建筑物的分割边界相较于其他算法,其他算法更接近Ground Truth,在视觉上验证了本算法融合结果对下游分割性能的提高。

4.3.2 定量比较分析

在表5中,我们计算了segformer分割模型在Potsdam数据集上的每个类别的IoU和mIoU。总体而言,我们的算法在背景,建筑,低矮植被,杂物的IoU和mIoU上取得了最佳值,在树类别的IoU上取得了次佳值,更加验证了目标语义分层驱动的红外和可见光图像融合增强了对下游任务的促进作用。

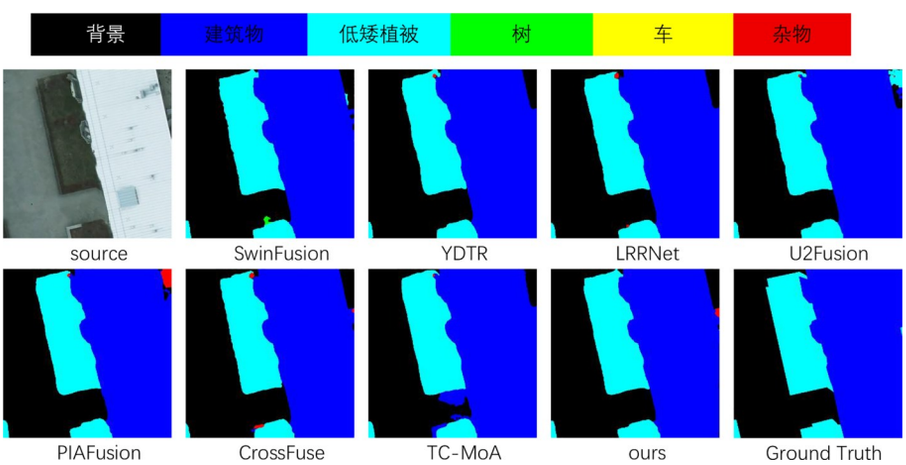

图8 不同融合方法在Potsdam数据集上的融合结果在Segmenter上的分割结果

Fig. 8 Segmentation results of different fusion methods on the Potsdam dataset using Segmenter

表 5 不同融合算法在Potsdam数据集上每类的分割结果. 黑体和下划线分别表示最佳值和次佳值.

Fig. 5 Segmentation results per class for different fusion algorithms on the Potsdam dataset. Bold and underlined values indicate the best and second-best performance, respectively.

算法	IoU						mIoU
	背景	建筑	低矮植被	树	车	杂物	
ir	78.93	89.95	72.29	71.56	76.66	29.80	69.89
vis	84.86	91.38	75.68	70.61	74.77	37.91	72.54
SwinFusion	85.76	92.73	77.00	72.15	**79.30**	39.02	74.33
YDTR	84.47	92.80	76.69	71.79	77.26	31.34	72.39
LRRNet	84.50	91.60	76.18	71.93	77.49	37.97	73.28
U2Fusion	85.54	93.28	77.58	71.89	77.87	37.80	73.99
PIAFusion	84.98	93.00	77.93	**72.94**	77.44	37.98	74.05
CrossFuse	84.70	92.47	76.30	71.26	78.70	37.49	73.49
TC-MoA	83.01	90.93	75.26	70.22	77.96	36.55	72.32
ours	**86.40**	**93.51**	**78.48**	72.52	78.31	**39.71**	**74.82**

4.4 消融实验

在这个部分将通过消融实验验证模块的有效性，并通过融合指标定量验证了融合效果，结果见表6和表7。

在表6中，分别去掉了网络的语义分割解码器SSD，结构相似性损失L_{ssim}以及强度损失L_{int}，从表中结果可以得知虽然融合效果尚可，但与本方法相比仍有一定差距。另外，本算法架构是一个双任务网络，需要选定式(1)中一个合适的参数α以确保分割损失和重建损失在优化期间保持数值平衡，从而防止任务优势偏差。为了保证融合结果良好，分别试取α为1、0.7、0.5、0.2并定量分析融合结果见表7。由表可知，当α=1、0.7、0.5时，融合结果与α=0.2时有较大差距，说明当α参数过大时，网络过分侧重语义分割任务，忽略了融合效果。

表6 消融研究在TNO数据集上的定量评价结果. 黑体表示最佳结果

Tab. 6 Quantitative evaluation results of ablation studies on the TNO dataset. Bold indicates the best results.

	EN	SD	MI	VIFF	$Q^{AB/F}$
w/o SSD	7.051 1	43.460 8	3.588 0	0.888 7	0.591 4
w/o L_{ssim}	7.086 8	44.313 6	3.395 1	0.874 8	0.585 8
w/o L_{int}	7.096 1	45.222 0	3.599 5	0.903 4	0.591 6
ours	**7.120 9**	**45.833 3**	**3.771 4**	**0.916 2**	**0.594 9**

表7 消融研究在TNO数据集上的定量评价结果. 黑体表示最佳结果

Tab. 7 Quantitative evaluation results of ablation studies on the TNO dataset. Bold indicates the best results.

	EN	SD	MI	VIFF	$Q^{AB/F}$
α=1	7.071 3	43.896 2	3.678 6	0.892 1	**0.596 9**
α=0.7	7.088 0	44.858 7	3.676 0	0.901 6	0.593 3
α=0.5	7.094 6	45.092 5	3.658 6	0.904 5	0.593 4
α=0.2	**7.120 9**	**45.833 3**	**3.771 4**	**0.916 2**	0.594 9

5 结 论

本文提出了一种目标语义分层驱动的红外和可见光图像融合方法SDFusion，为了解决以往图像融合弱化语义信息的问题，设计了一个双任务驱动的融合网络，通过同时生成融合结果和分割结果约束特征提取的方法使本网络在保留视觉特征的同时最大化利用了语义信息，并将相关特征分层注入特征重建过程，生成融合结果既能为人类视觉观察提供信息，又能更好服务高级视觉任务。实验结果显示，融合结果在五个客观评价指标EN，SD，MI，VIFF，$Q^{AB/F}$上较传统方法分别提升了3.7%，7.3%，45.3%，18.5%，7.2%；将融合结果用于分割实验，用像素交集(IoU)量化分割性能，本方法比较传统方法在大多数类别的IoU和mIoU上都具有优势；以上结果证明与现有的图像融合算法对比，本文提出的融合算法在视觉质量和高层语义方面具有相对优势。

参 考 文 献：

[1] ZHOU Z Q, *et al*. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters [J]. *Information fusion*, 2016, 30: 15-26

[2] ZHAN L C, ZHUANG Y, HUANG D L. Infrared and visible images fusion method based on discrete wavelet transform [J]. *Journal of Computers*, 2017, 28(2): 57-71

[3] LIU Y, CHEN X, RABAB K W, *et al*. Image fusion with convolutional sparse representation [J]. *IEEE Signal Process Letters*, 2016, 23(12): 1882 - 1886

[4] MA J Y, TANG L F, XU M, *et al*. STDFusionNet: An Infrared and Visible Image Fusion Network Based on Salient Target Detection [J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-13

[5] TANG L F, YUAN J T, MA J Y. 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: 28-42

[6] SUN Y M, *et al*. Detfusion: A detection-driven infrared and visible image fusion network [C]. Proceedings of the 30th ACM international conference on multimedia, 2022: 4003-4011.

[7] LIU J Y, LIU Z, WU G Y, *et al*. Multi-interactive feature learning and a full-time multi-modality benchmark for image fusion and segmentation [C]. Proceedings of the IEEE/CVF international conference on computer vision, 2023: 8115-8124

[8] CHEN J, DING J F, MA J Y. Hitfusion: Infrared and visible image fusion for high-level vision tasks using transformer [J]. *IEEE Transactions on Multimedia*, 2024, 26: 10145-10159.

[9] XIONG J X, LIU G, TANG H J, *et al*. SeGFusion: A semantic saliency guided infrared and visible image fusion method [J]. *Infrared Physics & Technology*, 2024, 140: 105344

[10] LI H, LIU L, HUANG W, *et al*. An improved fusion algorithm for infrared and visible images based on multi-scale transform [J] *Infrared Physics & Technology*, 2016, 74: 28-37.

[11] LI H, WU X J, and JOSEF K. Rfn-nest: An end-to-end residual fusion network for infrared and visible images [J]. *Information fusion*, 2021, 73: 72-86.

[12] LIANG P W, JIANG J J, LIU X M, *et al*. Fusion from decomposition: A self-supervised decomposition approach for image fusion [C]. European conference on computer vision, 2022:719-735.

[13] ZHAO Z X, XU S, ZHANG C X, *et al*. DIDFuse: Deep image decomposition for infrared and visible image fusion [C]. IJCAI, 2020:970 - 976

[14] LU Q, ZHANG H B, YIN L F. Infrared and visible image fusion via dual encoder based on dense connection [J]. *Pattern Recognition*, 2025, 163: 111476.

[15] MA J Y, YU W, LIANG P WP., *et al*. FusionGAN: A generative adversarial network for infrared and visible image fusion [J] *Information fusion*, 2019, 48: 11 - 26.

[16] MA J Y, XU H, JIANG J J, *et al*. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion [J]. *IEEE Transactions on Image Processing*, 2020, 29: 4980 - 4995.

[17] FU X, XU T Y, WU X J, *et al*. PPT fusion: Pyramid patch transformer for a case study in image fusion, *arxiv preprint*: 2017. 13967 (2021).

[18] MA J Y, TANG L F, FAN F, *et al*. SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer [C]. IEEE/CAA Journal of Automatica. Sinica,2022, 9(7): 1200 - 1217.

[19] TNAG W, HE F Z, LIU Z. ITFuse: An interactive transformer for infrared and visible image fusion [J]. *Pattern Recognition*, 2024, 156: 110822.

[20] YANG X, HUO H T, LI C, *et al*. Semantic perceptive infrared and visible image fusion transformer [J] *Pattern Recognition*, 2024, 149: 110223.

[21] RONNEBERGER O, FSICHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation [C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234 - 241.

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

[23] TANG L F, *et al*. PIAFusion: A progressive infrared and visible image fusion network based on illumination aware [J]. *Information Fusion*, 2022, 83: 79-92.

[24] XU H, *et al*. U2Fusion: A unified unsupervised image fusion network [J]. *IEEE transactions on pattern analysis and machine intelligence* 2020, 44(1): 502-518.

[25] WEI T, HE F Z, LIU Y. YDTR: Infrared and visible image fusion via Y-shape dynamic transformer [J]. *IEEE Transactions on Multimedia*, 2022, 25: 5413-5428.

[26] 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.

[27] YU H, GAO J C, ZHOU S P, *et al*. Cross-Modality Target Detection Using Infrared and Visible Image Fusion for Robust Objection recognition [J]. *Computers and Electrical Engineering*, 2025, 123: 110133.

[28] ZHANG Z L, TAO Z, and XU Z. EGFormer: Towards Efficient and Generalizable Multimodal Semantic Segmentation, *arxiv preprint arxiv*:2505.14014 (2025).

[29] ZHOU Z Q, WANG B, LI S, *et al*. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with Gaussian and bilateral filters [J]. *Information Fusion*, 2016, 30: 15 - 26.

[30] LI H, WU X J. CrossFuse: A novel cross attention mechanism based infrared and visible image fusion approach [J]. *Information Fusion*, 2024, 103: 102147.

[31] ZHU P F, *et al*. Task-customized mixture of adapters for general image fusion [C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2024: 7099-7108.

[32] 张永兴,连博文,顾乃庭等. 基于多尺度空间注意力互补的红外与可见光图像融合[J]. 光学精密工程,2025,33(07):1152-1168.
ZAHNG Y X, LIAN B W, GU N T, *et al*. Infrared and visible image fusion based on multi-scale spatial attention complementary [J]. *Optics and Precision Engineering*, 2025, 33(07): 1152-1168. (in Chinese)

[33] 贾轩,张叶,常旭岭,等. 基于深度度量学习和语义分割的场景识别[J]. 液晶与显示,2025,40(05):740-750.
JIA X, ZHANG Y, CHANG X L, *et al*. Scene Recognition Based on Deep Metric Learning and Semantic Segmentation [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(5): 740-750. (in Chinese)

[34] 张恒,王磊,张鹏超,等. 基于动态特征剔除与轻量化检测的视觉SLAM算法[J]. 液晶与显示,2025,40(05):727-739.
ZHANG H, WANG L, ZHANG P C, *et al*. Visual SLAM Algorithm Based on Dynamic Feature Elimination and Lightweight Detection [J] *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(5): 727-739. (in Chinese)

[35] 蔡忠祺,林珊玲,林竖普,等. 基于改进YOLOv8n-Pose的疲劳驾驶检测[J]. 液晶与显示,2025,40(04):617-629.
CAI Z Q, LIN S L, LIN J P, *et al*. Fatigue Driving Detection Based on Improved YOLOv8n-Pose [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(4): 617-629. (in Chinese)

作者简介:

陶俣卓,女,硕士研究生,2023年于渤海大学获得学士学位,主要从事图像处理方面的研究。E-mail:tyz64895671@163.com

赵　凡,女,博士,副教授,2017年于中国科学院长春光学精密机械与物理研究所获得博士学位,主要从事计算机视觉方面的研究。E-mail:3175209736@qq.com