DOI:10. 13382/j. jemi. B2508106

自适应特征增强的多尺度红外与可见光图像配准和融合算法*

孙溪成[1] 吕 伏[1,2] 尹艺潼[1]

(1. 辽宁工程技术大学软件学院 葫芦岛 125105;2. 辽宁工程技术大学基础教学部 葫芦岛 125105)

摘 要:目前的红外与可见光图像融合算法对图像的特征提取不充分,细节信息丢失。实际生活中的红外与可见光图像多为未配准图像,已有的配准算法仍存在伪影和偏差等问题。针对以上问题,提出了一种自适应特征增强的多尺度红外与可见光图像配准和融合算法。首先,在配准网络中使用多尺度卷积核和密集连接,以提取不同尺度的特征和防止信息丢失,并引入 ORB 特征点检测算法和设计的特征增强模块,以充分提取特征和适应复杂环境;其次,通过引入通道注意力和自学习参数设计了光照增强模块,以增强可见光图像的信息表达;然后,在融合网络里,利用不同池化和可变卷积设计了自适应多尺度池化卷积,以提取不同尺度的细节信息,并设计 EMA 特征融合模块将局部特征和全局特征进行融合;最后,设计了流场一致性损失函数,从而减小配准误差。为了更好地验证方法的实用性,建立了红外与可见光图像数据集。在公开数据集 TNO、Roadscene 和自建数据集上进行对比实验和消融实验,实验表明,在主观评价上,配准图像偏差小无伪影,融合图像清晰可见,客观评价上,在指标 MSE、MI、NCC、SD、EN 上相比于其他算法提高了 20%、7%、4%、15%、8%左右。另外,在 YOLOv8 上进行融合结果的目标检测性能实验,检测性能表现良好。

关键词: 图像配准;图像融合;多尺度;自适应;特征增强

中图分类号: TP391;TN911. 73 **文献标识码**: A **国家标准学科分类代码**: 520. 60

Multi-scale infrared and visible image registration and fusion algorithm with adaptive feature enhancement

Sun Xicheng[1] Lyu Fu[1,2] Yin Yitong[1]

(1. School of Software, Liaoning Technical University, Huludao 125105, China;
2. Department of Basic Education, Liaoning Technical University, Huludao 125105, China)

Abstract: The current infrared and visible light image fusion algorithms often fail to fully extract image features, resulting in the loss of detail information. In real-world scenarios, infrared and visible light images are typically unregistered, and existing registration algorithms still suffer from artifacts and biases. To address these issues, this paper proposes an adaptive feature enhancement multi-scale infrared and visible light image registration and fusion algorithm. First, multi-scale convolutional kernels and dense connections are used in the registration network to extract features at different scales and prevent information loss. Additionally, an ORB feature point detection algorithm and a designed feature enhancement module are introduced to fully extract features and adapt to complex environments. Secondly, a lighting enhancement module is designed by incorporating channel attention and self-learning parameters to improve the information expression of visible light images. Then, in the fusion network, adaptive multi-scale pooling convolutions are designed using different pooling strategies and variable convolutions to extract detail information at multiple scales. An EMA feature fusion module is designed to integrate local and global features. Finally, a flow consistency loss function is introduced to minimize registration errors. To better validate the practical applicability of the proposed method, an infrared and visible light image dataset is established. Comparative and ablation experiments are conducted on the public datasets TNO, Roadscene, and a self-constructed dataset. The experimental results show that, in terms of subjective evaluation, the registered images have minimal bias and no artifacts, while the fused images are clear and visible. On objective evaluation, it improves about 20%, 7%, 4%, 15%, and 8% on the metrics

收稿日期: 2025-01-09 Received Date: 2025-01-09

*基金项目:国家自然科学基金面上项目(52274206)、国家自然基金面上项目(51874166)、国家自然基金青年基金项目(51904144)资助

MSE, MI, NCC, SD, and EN compared to other algorithms. Additionally, target detection performance experiments on YOLOv8 show that the fusion results exhibit good detection performance.

Keywords: image registration; image fusion; multi-scale; adaptive; feature enhancements

0 引　言

近年来,人们对红外和可见光图像融合(infrared and visible image fusion, IVIF)越来越感兴趣,其目的是合并红外和可见光传感器的互补信息,从而生成更加全面的图像。红外传感器成像利用物体发射的热辐射,不受光照变化的影响,能够补偿可见传感器在描述场景和物体的关键属性方面的局限性,特别是在不利的照明条件下。目前的很多 IVIF 已经应用到目标检测[1]、语义分割、场景理解[2]和自动驾驶等实际应用。除此之外,多模态融合技术在电力、遥感、军事和人脸识别等领域发挥着重要作用[3]。现有的基于深度学习的 IVIF 技术可以分为 3 类,即基于自动编码器的方法,端到端基于卷积神经网络(convolution netural networks,CNN)的方法和生成对抗网络的方法。

预训练的自动编码器用于实现特征提取和特征重建,其中融合规则由手动设计完成。Li 等[4]首先为 IVIF 引入了一个自动编码器网络。通过在编码器部分集成密集块,可以全面提取特征。考虑到重要信息通常从网络退化,Liu 等[5]采用不同的接收扩张卷积从多尺度前瞻性中提取特征,然后通过边缘注意机制合并这些提取特征。Zhao 等[6]提出了一种基于自动编码器的融合网络,其中编码器将图像分别分解为具有低频和高频信息的背景和细节特征图,然后通过解码器部分生成融合结果。

首次将 CNN 应用于红外与可见光图像融合的方法由 Zhang 等[7]提出,旨在提升图像的细节信息。CDDFuse[8]在突出物体和保留纹理方面表现良好,但在解决融合图像中的结构畸变和边缘方面遇到了困难。为了设计一个可以应用于多个融合任务的融合网络,Xu 等[9]提出了 U2Fusion。黄玲琳等[10]为了解决图像融合后出现的伪影、小目标不清晰等问题,将图像下采样到多个尺度再利用注意力模型进行特征融合。陈潮起等[11]将红外、可见光图像分解为多层次图像,再设计最优的融合策略。陈永等[12]利用密集连接网络对两种图像的多尺度信息进一步特征提取,最后利用全连接层构成的解码器进行图像重建。李天放等[13]通过引入卷积注意力机制和语义分割网络增强特征提取并约束融合网络。为了将融合方法应用到无人机端,童小钟等[14]利用知识蒸馏技术降低模型参数量。但是这些方法对图像的细节信息和长距离依赖关系提取不充分。

除此之外,已经产生了广泛的基于生成对抗网络(generative adversarial network,GAN)的融合方法。Ma 等[15]在可见光图像和融合结果之间建立对抗性游戏以增强纹理细节。然而,他们只使用了可见光图像中的信息,从而失去了融合结果上目标的对比度或轮廓。为了改善这个问题,后来引入了双鉴别器 GAN[16],其中红外和可见光图像都参与网络,从而显着提高了融合性能。随着更多的尝试,Li 等[17]引入了一种集成了多分类约束的端到端 GAN 模型。Liu 等[18]设计了一个带有一个生成器和双鉴别器的融合网络。

Transformer[19]首次应用在在自然语言处理领域。后来,Dosovitskiy 等[20]提出了用于图像分类的视觉转换器(vision transformer,ViT)。VS 等[21]提出了能够同时使用局部信息和全局信息的图像融合 Transformer 模型,弥补了 CNN 模型提取全局上下文信息的能力不足。Ma 等[22]提出了一种融合方法,它可以保留强度最大的源模态像素。陈彦林等[23]利用 Transformer 组件和卷积组成多尺度自注意力编码器-解码器,充分提取局部特征和长距离依赖关系。但是由于 Transformer 和自注意力的计算复杂度高,导致融合模型的运行效率不高。

目前这些融合方法是为手动预配准的红外和可见图像设计的,但是实际生活中多是未配准的图像。对于图像配准。文献[24]依赖于光流估计,目的是估计源图像和目标图像之间的密集像素对应关系。Zhou 等[25]提出了一种称为 CrossRAFT 的图像配准框架,用于扩展多模态图像的现成单模态光流估计模型。Nemar[26]同时训练图像平移和配准,导致复杂和相互破坏性的优化。文献[27]提出的 UMF 从粗到细的图像配准方法采用一阶段训练方式。然而,它仅依靠从两个尺度中提取的特征来估计最终的变形场,缺乏全局特征的表示。

目前的多模态图像配准算法仍存在错位、伪影等问题,而且缺少多尺度特征的提取,导致部分信息丢失。现有的融合算法存在特征提取不充分,细节信息丢失,模型参数量大等问题。针对以上问题,本文提出自适应特征增强的多尺度图像配准和融合算法,在配准网络里,引入 ORB 算法并设计特征增强模块,通过多尺度特征提取和密集连接使配准算法适用能力更强。在融合网络里,设计了自适应多尺度池化卷积和光感增强模块,以提取不同尺度的特征和增强可见光图像的特征。将本文算法与其他 7 种先进算法在公开数据集 TNO、Roadscene 和自建数据集上进行大量对比实验,在主观评价上融合图像纹理轮廓清晰;在客观评价上,EN、MI、SD 明显高于其他算法。除此之外,在融合图像上进行检测实验,检测性能表现良好。

1　本文方法

1.1　本文框架

本文的工作流程如图1所示。根据数据来源的不同,配准的过程分为两种情况,第1种是已经配准的红外与可见光图像,需要利用已有的生成对抗网络CPST[27]将红外图像I_{ir}错位变形成伪红外图像I_{ir}^{fake},如式(1)所示,然后将I_{ir}^{fake}和I_{ir}输入到多尺度密集配准网络(multi-scale dense registration network,MDRN)进行配准,如式(2)所示。另一种情况是给定一对未对齐的可见光图像I_{vis}和红外图像,其目标是将红外图像配准到可见图像上,从而输出配准图像I_{ir}^{reg},如式(3)所示。融合过程中,将配准后的图像与可见光图像输入到自适应特征融合网络(adaptive feature fusion network,AFFN),通过光感增强模块(light enhancement module,LEM)提高可见光图像信息表达,利用STB(swin-transformer block)和自适应多尺度池化卷积(adaptive multi-scale pooled convolution,AMPC)分别提取局部和全局特征并用高效多尺度注意力特征融合模块(EMA feature fusion module,EMAM)将其融合,最后输出融合图像I_{fuse},如式(4)所示。图1中绿色和橙色箭头分别表示配准过程的两者情况,红色和紫色箭头代表融合过程。

$$I_{ir}^{fake}=CPST(I_{ir}) \tag{1}$$

$$I_{ir}^{reg}=MDRN(I_{ir},I_{ir}^{fake}) \tag{2}$$

$$I_{ir}^{reg}=MDRN(I_{ir},I_{vis}) \tag{3}$$

$$I_{fuse}=AFFN(I_{vis},I_{ir}^{reg}) \tag{4}$$

式中:*CPST*代表生成对抗网络;*MDRN*代表多尺度密集配准网络;*AFFN*代表自适应特征融合网络。

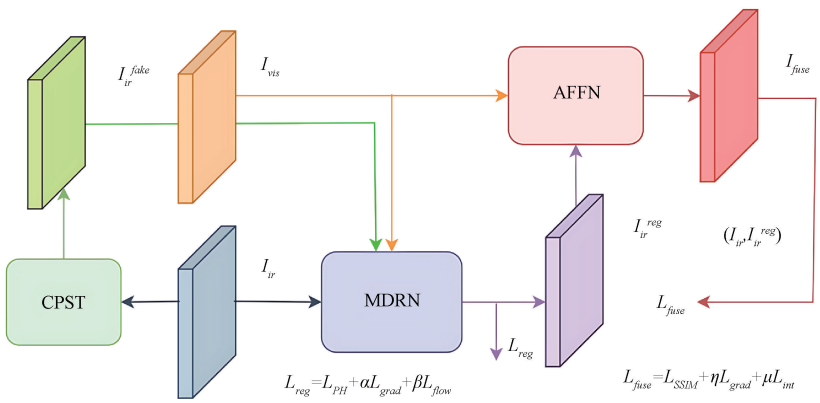

图1　红外与可见光图像配准和融合流程

Fig. 1　Flow chart of infrared and visible image registration and fusion

1.2　多尺度密集配准网络

由粗到细配准方法主要应用于处理图像的对齐问题。粗尺度初步配准通常从较低分辨率或粗略的特征开始,快速进行初步对齐。逐步细化配准将配准结果在更高分辨率的图像上进行微调,逐步引入更多细节信息。通过不同尺度的图像来确保更高的鲁棒性和精确度。密集连接可以保证配准过程里的信息重复使用,降低信息丢失的风险。多尺度密集配准网络如图2所示。

首先将红外与可见光图像分别输出到双分支特征提取分支,包括ORB算法[28]分支和卷积层分支。将两种特征输出后拼接并利用轻量化的深度可分离卷积进行特征融合,将通道维度恢复到之前维度。

$$M_{ir}^{1}=M(I_{ir}^{1}) \tag{5}$$

$$R_{ir}^{1}=ORB(I_{ir}^{1}) \tag{6}$$

$$F_{ir}^{1}=DWConv_1(Concate(M_{ir}^{1},R_{ir}^{1})) \tag{7}$$

$$F_{vis}^{1}=DWConv_1(Concate(M_{vis}^{1},R_{vis}^{1})) \tag{8}$$

式中:M为共享的多尺度特征提取器;R代表ORB算法;F_{ir}^{1}、F_{vis}^{1}分别为红外与可见光图像在两种特征提取器上的结合;Concate为拼接操作;$DWConv_1$为3×3的深度可分离卷积。然后将红外与可见光图像输入到变形场预测模块,预测出变形场为:

$$D_{ir}^{1}=DFE(F_{ir}^{1},F_{vis}^{1}) \tag{9}$$

式中:*DFE*为由粗到细的变形场预测模块[27];D为预测出的变形场。将红外图像按照预测的变形场进行变换,并与下一阶段的红外图像继续预测变形场为:

$$D_{ir}^{2}=DFE(ST(I_{vis}^{2},D_{ir}^{1}),I_{ir}^{2}) \tag{10}$$

式中:ST[29]为基于图像X和仿射参数p的空间变换的实现。将上一阶段的输出分别上采样到下面阶段的分辨率大小为:

$$Dense_{ir}^{2}=Upsample(D_{ir}^{2})\oplus D_{ir}^{1} \tag{11}$$

式中:*Upsample*表示上采样操作;*Dense*[30]表示密集连接操作。将最后的输出进行预测变形进行输:

$$I_{ir}^{reg}=ST(Dense_{ir}^{2}\oplus D_{ir}^{3}) \tag{12}$$

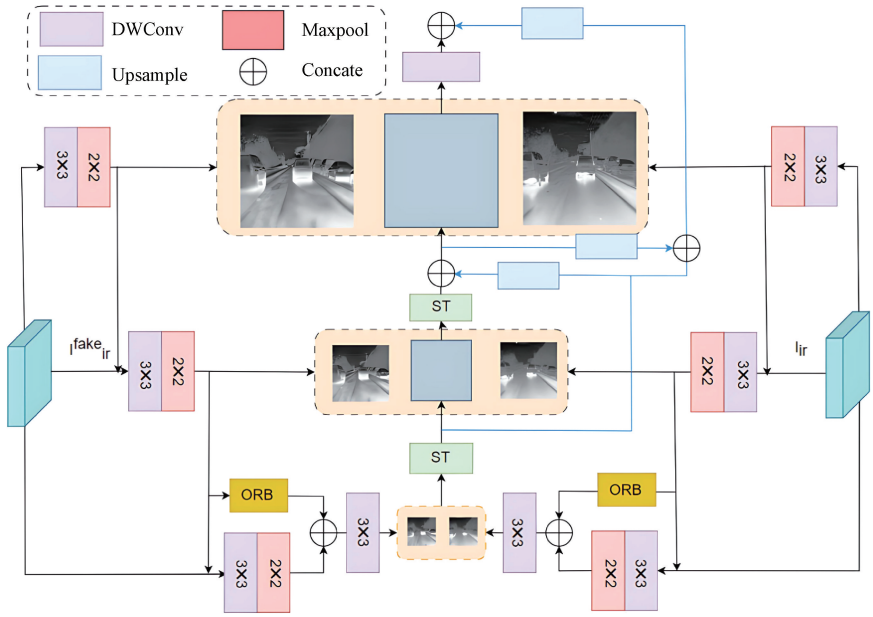

图 2　多尺度密集配准网络

Fig. 2　Multi-scale dense registration network

1) ORB 特征匹配算法

深度学习模型可能在面对大尺度的几何变换时表现不佳,特别是当数据中变换种类多样时,训练模型可能无法很好地应对各种变换的情况。

ORB 算法通过提取旋转不变的特征点,并且具有较好的尺度、旋转和光照变化的适应性。除此之外,ORB 算法还有着高效率、计算速度快、适合实时应用等优势,它在处理速度和资源有限的场景中表现优异。ORB 算法特征提取分支如图 3 所示。

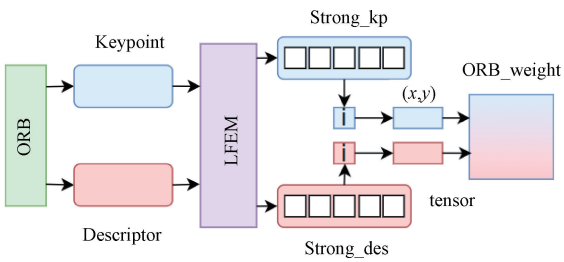

图 3　ORB 算法特征提取分支

Fig. 3　Feature extraction branch of ORB algorithm

2) 特征增强模块

ORB 描述子是通过基于像素值计算的局部特征生成的,具有较强的局部性和二值化特征。在复杂场景中,这些描述子可能由于光照等变化而表现不佳。针对这个问题,本文结合轻量化的深度可分离卷积设计了轻量化的特征增强模块(lightweight feature enhancement module, LFEM)。

LFEM 通过对 ORB 特征进行进一步的学习和增强,能够捕捉到更多复杂的特征信息,从而提升配准的准确性,尤其是在纹理较少或噪声较多的场景下。模型能够更好地应对不同视角、尺度、光照变化等外部因素的干扰,增强了图像配准的鲁棒性。

3) 自适应特征融合网络

本文设计的自适应特征融合网络由 3 部分组成,分别为特征增强部分,特征提取部分,特征融合部分,如图 4 所示。特征增强部分包括分组卷积和光感增强模块,分组卷积可以降低网络参数量,光感增强模块可以使可见光的重要部分更加明显。特征提取部分由两个分支组成,其中全局特征提取分支由 swin transformer block[31] 组成,以提取图像的全局特征和长距离依赖关系,局部特征提取分支由自适应卷积和两组不同的池化组成,以提取图像的细节特征。特征融合部分由轻量化的残差模块[32] 和 EMA[33] 组成,以融合图像的局部和全局特征。

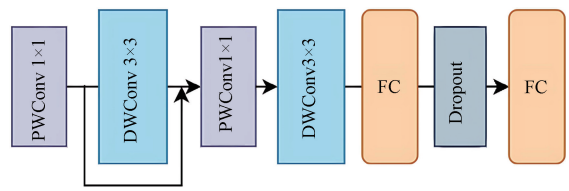

图 4　轻量化特征增强模块

Fig. 4　Lightweight feature enhancement module

由于拍摄场景多为光照差的条件,所以在可见光图像后引入光照增强模块,增加可见光的有效信息。然后将红外与可见光图像同时输入分组卷积:

$$I_{ir}^{g} = Groupconv(I_{ir}) \tag{13}$$

$$I_{vis}^{g} = Groupconv(LEM(I_{vis})) \tag{14}$$

式中：*Groupconv* 代表分组卷积；*LEM* 代表光感增强模块；I_{ir}^{g}、I_{vis}^{g} 代表红外与可见光图像在分组卷积后的输出。然后将图像输入到双分支特征提取部分，同时提取局部特征和全局特征：

$$I_{ir}^{A} = AFFN(I_{ir}^{g}) \tag{15}$$

$$I_{ir}^{S} = STB(I_{ir}^{g}) \tag{16}$$

式中：*AFFN* 代表自适应多尺度池化卷积；*STB* 代表 swin transformer block；I_{ir}^{A}、I_{ir}^{S} 代表提取的红外图像的局部和全局特征。将红外与可见光图像的局部特征和全局特征分别融合，并利用残差模块进行进一步融合特征：

$$I_{fuse}^{S} = Residual(I_{ir}^{S} \oplus I_{vis}^{S}) \tag{17}$$

$$I_{fuse}^{A} = Residual(I_{ir}^{A} \oplus I_{vis}^{A}) \tag{18}$$

式中：*Residual* 代表残差模块；I_{fuse}^{S}、I_{fuse}^{A} 代表融合后的全局和局部特征。为了将融合后的局部-全局特征进行更合适的权重融合，这里使用 EMA 进行融合，然后利用 sigmoid 生成权重分别与局部和全局特征相乘，为保留原有特征，最后与融合的局部-全局特征相融合并输出融合图像：

$$I_{fuse}^{EMA} = Sigmoid(EMA(I_{fuse}^{S} + I_{fuse}^{A})) \tag{19}$$

$$I_{fuse} = I_{fuse}^{EMA} \otimes I_{fuse}^{S} + I_{fuse}^{EMA} \otimes I_{fuse}^{L} + I_{fuse}^{L} + I_{fuse}^{S} \tag{20}$$

式中：*Sigmoid* 代表激活函数；*EMA* 代表高效多尺度注意力。

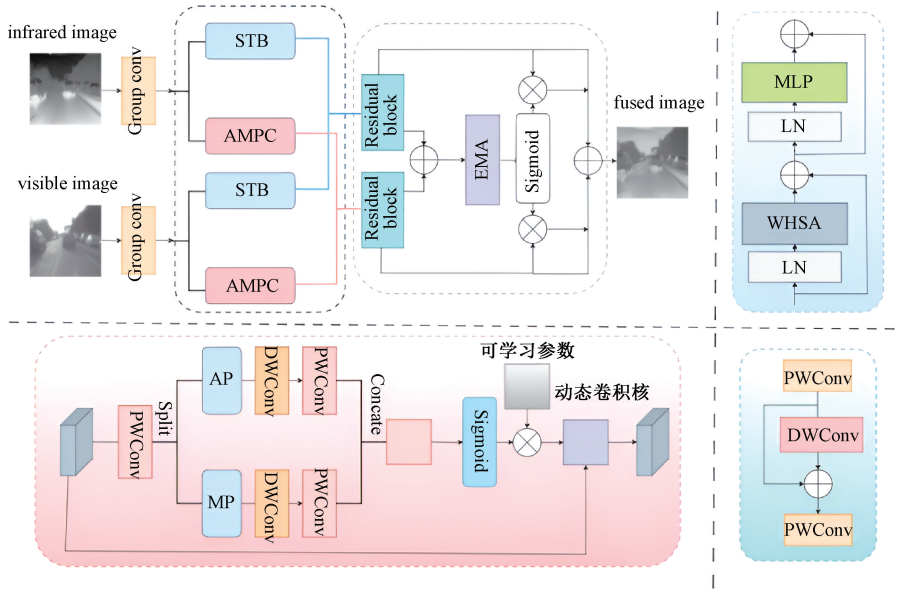

图 5　自适应特征融合网络

Fig. 5　Adaptive feature fusion network

(1) Swin Transformer Block

Swin Transformer 是一个改进的视觉 Transformer 模型，在图像处理任务中具有较高的性能和效率。机制弥补了普通窗口划分可能带来的局部性问题，使网络在捕捉全局信息时更具优势。全局特征提取分支可以表示为：

$$I_{ir}^{S} = MLP(LN(SWMA(LN(I_{ir}^{g})) + I_{ir}^{g})) \tag{21}$$

式中：*SWMA* 代表滑动窗口注意力；*MLP* 代表多层感知机；*LN* 代表层归一化。

(2) 自适应多尺度池化卷积

实际生活中的各个场景可能存在于不同的尺度上。自适应多尺度池化卷积可以动态调整卷积尺度，从而在不同尺度下都能准确地识别物体。相比于传统的固定卷积核结构，这种适应性使其在检测大范围和细小范围方面表现更好。

自适应多尺度池化卷积由深度可分离卷积、不同池化层和动态卷积构成，通过动态调整卷积核，使得网络能够更好地适应输入数据特征的变体，从而更加准确地捕获细节和全局特征。这提高了网络的灵活性和复杂度，在不同的场景下可以提高模型的表现能力，提高模型的泛化性能。相比使用固定的大卷积核或多个卷积核组合，自适应卷积能在相对较低的计算复杂度下，获得更好的特征提取能力。

首先输入到逐点卷积升高维度，按通道维度将输入平均分成两部分：

$$X_1, X_2 = Split(PWConv(I_{ir}^{g})) \tag{22}$$

式中：*PWConv* 代表逐点卷积；*Split* 代表划分操作；X_1、X_2 代表划分出的两个分支。两个分支分别利用平均池化和最大池化进行提取特征，然后利用深度可分离卷积提取局部特征，逐点卷积降低通道维度：

$$Y_1 = PWConv(DWConv(Avgpool(X_1))) \tag{23}$$

$$Y_2 = PWConv(DWConv(Maxpool(X_2))) \tag{24}$$

式中：*DWConv* 代表 3×3 的深度卷积；*Avgpool* 代表平均池化；*Maxpool* 代表最大池化；Y_1、Y_2 代表两个分支的输出。将两个分支的特征进行拼接，利用 sigmoid 生成权重，可学习参数生成动态卷积核，最后与输入相乘：

$$I_{ir}^{A} = (Sigmoid(Concate(Y_1, Y_2)) \otimes L) \otimes I_{ir}^{g} \quad (25)$$

式中：*Sigmoid* 代表激活函数；L 代表学习参数。

(3) EMA 特征融合模块

EMA 不仅通过类似 CA 的处理避免了维度减缩，还通过 3×3 卷积捕获特征，并通过跨空间学习方法聚合多个并行子网络的输出特征图。

EMA 特征融合模块由轻量化的 Residual block 和 EMA 注意力组成，其通过动态加权和聚焦重要特征，显著提高了模型的表现能力并使全局特征和局部特征更好地进行结合。

(4) 光感增强模块

可见光图像在光线不足时会显得较为模糊或噪声较大。针对以上问题本文基于通道注意力设计了光感增强模块。

LEM 由 3 个部分组成，即联合特征提取部分、通道注意力和学习参数校正。首先使用深度可分离卷积层对三通道进行联合特征提取，以捕获跨通道的光照变化。然后根据改进的 SE[34] 为输入图像的不同通道分配不同的权重，以实现自适应增强。最后通过对 RGB 3 个通道分别应用不同的 Gamma 值进行调整，达到细腻的亮度增强效果。

LEM 可以自适应调整图像的对比度和亮度，以便在低光环境下仍能突出重要的图像特征，如物体的轮廓、细节等。这样可以避免图像过暗或过亮，保证目标物体的清晰可见。

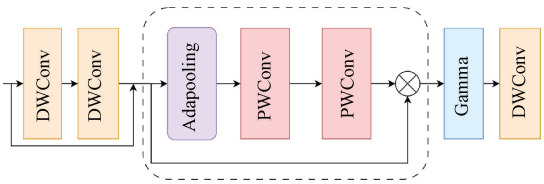

图 6　轻量化光感增强模块

Fig. 6　Lightweight light enhancement module

1.4　损失函数

1) 配准网络的损失函数

通过设计 ORB 特征点的流场一致性，能够确保配准后的图像在具有较大位移或变化的情况下依然能够保持较高的精度。ORB 特征点在不同模态(如红外与可见光图像)的配准中尤其有效。即使图像的像素值差异较大，ORB 特征点仍能提取出关键点，保证在不同模态图像中的配准一致性。

光度损失确保图像的整体一致性，梯度损失确保边缘和细节的保持，而流场一致性损失则帮助优化全局与局部特征的对齐，所以本文的配准网络使用光度损失、梯度损失、流场一致性损失相结合的损失函数。

光度损失用于衡量配准后图像的像素级差异，确保图像的全局对齐，光度损失函数定义为：

$$L_{PH}(I_{ir}^{reg}, I_{ir}) = \frac{1}{HW}\sum_{i,j}^{H,W} \| I_{ir}^{reg}(i,j) - I_{ir}(i,j) \|_1 \quad (26)$$

梯度损失通过衡量图像中像素强度变化(即边缘信息)的差异，有助于保留图像中的结构信息：

$$L_{grad}(I_{ir}^{reg}, I_{ir}) = \frac{1}{HW}\sum_{i,j}^{H,W} \| \nabla I_{ir}^{reg}(i,j) - \nabla I_{ir}(i,j) \|_2^2 \quad (27)$$

流场一致性损失表示为：

$$L_{flow} = \sum_{i=1}^{N} \| (flow_{reg\to ir}(p_i) - flow_{ir\to reg}(p_i) \|_2^2 \quad (28)$$

其中，$flow_{reg\to ir}(p_i)$ 表示特征点 p_i 从配准图像红外图像的光流估计，而 $flow_{ir\to reg}(p_i)$ 是反向流场，表示从红外图像到配准图像的光流估计。

总的配准网络损失函数可以表达为：

$$L_{reg} = L_{PH} + \alpha L_{grad} + \beta L_{flow} \quad (29)$$

式中：α、β 代表权重系数，这里设置为 10 和 2。

2) 融合网络的损失函数

本文利用结构相似性(SSIM)损失函数来保持融合图像的更清晰的强度分布，定义为：

$$L_{SSIM} = (1 - SSIM(I_{fuse}, I_{ir}^{reg})) + (1 - SSIM(I_{fuse}, I_{vis})) \quad (30)$$

式中：*SSIM*[35] 表示结构相似度度量，可以从光线、对比度和结构 3 个角度度量图像失真。

为了获得更清晰的纹理，融合图像的梯度被迫接近红外和可见图像梯度之间的最大值。

$$L_{grad} = \| \nabla I_{fuse} - \max(\nabla I_{ir}^{reg}, \nabla I_{vis}) \|_1 \quad (31)$$

融合图像还期望融合源图像中的强度信息，特别是红外图像中的重要目标。因此，本文使用强度最大化损失 L_{int} 来引导融合网络自适应地整合源图像的强度信息：

$$L_{\text{int}} = \frac{1}{HW} \| I_{fuse} - \max(I_{ir}^{reg}, I_{vis}) \|_1 \quad (32)$$

总融合网络损失函数可以表达为：

$$L_{Fuse} = L_{SSIM} + \omega L_{grad} + \mu L_{\text{int}} \quad (33)$$

式中：ω、μ 代表权重系数，这里取 10 和 3。

2　实验

2.1　实验数据和细节

1) 数据集和预处理

(1) 公开数据集

TNO 数据集[36] 和 Roadscene 数据集[9] 两个公开的

数据集包含了不同类型的场景。为了保证实验的公平性,将两个数据集打乱并随机分组,按照比例4∶1∶1划分为训练集、验证集和测试集。为了评估模型的泛化能力,划分之前将两个数据集混合打乱,即让训练集、验证集和测试集都有复杂的场景。

(2)自建数据集

为了验证本文方法在实际生活中的实用性,使用无人机搭载红外和可见光双摄像机在鄂尔多斯城市街道采集了红外与可见光图像数据集。无人机型号为DJI Matrice 300 RTK,相机模组为DJI Ienmuse H20T,该无人机将两种摄像机高度集成,同时支持红外和可见光拍摄,并支持同步拍摄。拍摄场景包含了白天、夜间下的汽车、行人、树木等,保证了数据集的复杂性及模型的泛化能力。处理后的数据集包含500对红外和可见光图像,统一处理为640×512的分辨率,按比例6∶1∶1划分为训练集、验证集和测试集,并确保数据随机分布均匀。

2)对比方法

配准网络的对比实验选择7种经典的配准方法,分别为Nemar、CrossRAFT、Superfusion[37]、UMF、IMF[38]、Yang等[39]和Li等[40]。融合网络的对比实验选择7种融合方法,分别为DIDFuse、RFN[41]、UMF、CDDFuse、CoCoNet[42]、LRRNet[43]、IMF。

3)评价指标

本文使用3个指标评估图像的配准结果,即均方误差(mean squared error,MSE)、互信息(mutual information,MI)[44]和归一化互相关(normalized cross correlation,NCC)[45]。使用熵(EN)、互信息、视觉信息保真度(visual information fidelity,VIF)[46]、SSIM[47]和标准差(standard deviation,SD)评估融合图像。

4)实验细节

配准网络和融合网络都在PyTorch 1.10.1中实现,使用NVIDIA Geforce RTX3090 GPU进行训练和测试。选择Adam优化器(β_1 = 0.9,β_2 = 0.999)用于优化模型。初始学习率设置为0.001,每隔100个epoch减小为之前的1/10,共经过300个epoch,Batch size设置为16。

2.2　对比实验

1)配准网络对比实验

为了直观地展示本文配准算法的优越性,在图7中的真实世界错位多模态数据集中提供了不同方法的配准情况。由于真实的错位数据集缺乏目标红外图像,通过配准图像和可见图像之间的误差图来评估配准精度。观察后,本文方法产生了理想的配准效果并抑制边缘重影。

表1为配准网络的对比实验定量分析,加粗字体为指标中最出色的算法,下划线的算法为第2名。本文方法与其他7种配准方法相比,本文方法在所有指标(即MSE、NCC和MI)中排名第1,在3个数据集上都优于其他方法。综合分析本文方法在图像配准实验中表现最佳,也说明了多尺度特征提取和密集连接的重要性。

表1　配准网络在3个数据集上的对比实验定量比较

Table 1　Quantitative comparison of registration networks on three datasets

方法	年份	TNO			Roadscene			自建数据集		
		MSE	NCC	MI	MSE	NCC	MI	MSE	NCC	MI
MIS	none	0.007	0.876	1.558	0.011	0.894	1.602	0.013	0.899	1.615
Nemar	2020	0.140	0.309	0.438	0.081	0.846	0.986	0.076	0.873	1.120
Yang[40]	2021	0.006	0.628	1.648	0.007	0.783	1.835	0.007	0.932	1.413
Li[41]	2022	0.007	0.772	1.563	0.008	0.847	1.736	0.009	0.892	1.336
CrossRAFT	2022	0.008	0.858	1.602	0.009	0.910	1.744	0.015	0.923	1.653
Superfusion	2022	0.007	0.886	1.507	0.004	0.953	1.820	0.009	0.874	**1.713**
UMF	2022	0.004	0.926	1.648	0.004	0.963	1.833	0.009	0.882	1.672
IMF	2023	**0.003**	**0.957**	**1.804**	**0.003**	**0.967**	**1.993**	**0.006**	**0.916**	1.637
本文	none	0.002	0.973	1.963	0.002	0.978	2.103	0.005	0.953	1.758

2)融合对比实验

图8所示为自建数据集的定性结果。本文方法在融合之前使用了多尺度密集配准网络进行精细配准,而其他方法使用了IMF的配准网络。从综合表现来看,RFN、DIDFuse、LRRNet和CoCoNet的融合图像光度较暗,场景不明显。其他三者方法的融合结果对比度较低,纹理细节不够清晰。本文方法融合的图像清晰可见,对比度合适。

表2~4所示为TNO、Roadscene和自建数据集上的定量比较结果。从3个数据集的综合表现来看,本文方法在EN、SD、MI上明显超过其他方法,说明融合结果最佳,在VIF、SSIM指标上分别排名第3和第2,表现良好。

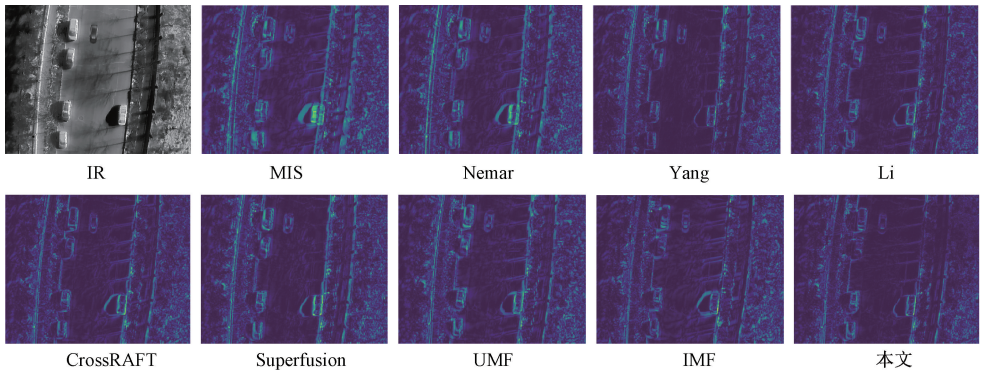

图 7　配准网络在自建数据集上的对比实验定性比较

Fig. 7　Qualitative comparison of the registration network on a self-built dataset

表 2　融合网络在数据集 TNO 上的对比实验定量比较

Table 2　Quantitative comparison of fusion networks on the dataset TNO

模型	年份	EN	SD	VIF	SSIM	MI
DIDFuse	2020	6. 97	45. 12	0. 61	0. 81	1. 69
RFN	2021	6. 83	34. 50	0. 51	0. 92	1. 21
UMF	2022	6. 98	40. 73	1. 01	0. 47	1. 53
CDDFuse	2023	7. 12	46. 00	0. 77	1. 03	2. 19
CoCoNet	2023	7. 77	46. 37	0. 89	1. 17	2. 31
LRRNet	2023	6. 85	43. 45	0. 71	0. 70	1. 87
IMF	2023	7. 27	46. 27	1. 00	0. 47	2. 16
本文	none	8. 02	47. 18	0. 98	1. 12	2. 54

表 3　融合网络在数据集 Roadscene 上的对比实验定量比较

Table 3　Quantitative comparison of fusion networks on the dataset Roadscene

模型	年份	EN	SD	VIF	SSIM	MI
DIDFuse	2020	7. 43	51. 58	0. 58	0. 86	2. 11
RFN	2021	7. 21	41. 25	0. 54	0. 90	1. 68
UMF	2022	7. 37	44. 73	0. 88	0. 50	1. 87
CDDFuse	2023	7. 44	54. 67	0. 69	0. 98	2. 30
CoCoNet	2023	7. 69	40. 67	0. 76	1. 16	2. 53
LRRNet	2023	6. 97	43. 73	0. 77	0. 69	2. 17
IMF	2023	7. 67	48. 73	0. 87	0. 49	2. 40
本文	none	8. 12	55. 87	0. 74	1. 04	2. 67

表 4　融合网络在自建数据集上的对比实验定量比较

Table 4　Quantitative comparison of fusion networks on self-built datasets

模型	年份	EN	SD	VIF	SSIM	MI
DIDFuse	2020	7. 44	52. 52	0. 57	0. 84	2. 05
RFN	2021	7. 32	40. 72	0. 52	0. 87	1. 55
UMF	2022	7. 33	43. 73	0,66	0. 46	1. 67
CDDFuse	2023	7. 37	50. 61	0. 67	0. 87	2. 18
CoCoNet	2023	7. 67	41. 72	0. 66	1. 15	2. 36
LRRNet	2023	6. 83	44. 62	0. 56	0. 62	2. 14
IMF	2023	7. 72	48. 37	0. 77	0. 53	2. 27
本文	none	7. 98	53. 17	0. 78	0. 94	2. 46

3)运行效率

除了融合效果的比较,模型大小和运行速度在实际应用中也很重要。因此,本文验证了所提出模型的内存消耗和计算效率。对于运行时间,从 TNO 中随机选择 10 个图像来计算平均时间。

表 5 所示为几种最先进的方法中模型大小、浮点数和运行时间的定量结果。由于引入了光感增强模块和 EMA 特征融合模块,导致模型参数量相比 DIDFuse 和 LRRNet 更大。但是,因为用的是轻量化的特征增强模块和残差模块,使得本文算法参数量相比其他模型表现

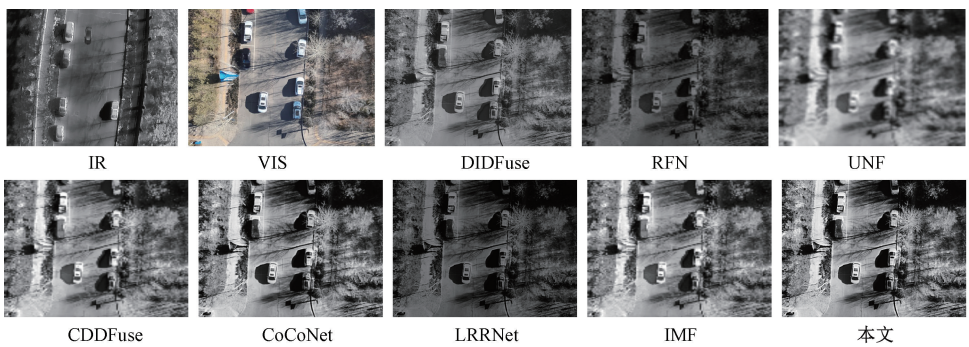

图 8　融合网络在自建数据集上的对比实验定性比较

Fig. 8　Qualitative comparison of fusion networks on self-biult dataset

表5　各种算法在数据集TNO上的运行效率对比实验

Table 5　Comparative experiments on the running efficiency of various algorithms on the TNO dataset

	DIDFuse	RFN	UMF	CDDFuse	CoCoNet	LRRNet	IMF	本文
TP/($\times10^6$)	0.261	10.936	7.732	4.273	9.130	0.492	6.736	2.127
浮点数/GFLOPs	18.71	676.09	438.27	267.73	115.37	134.79	203.71	220.68
用时/s	0.055	0.239	0.337	0.064	0.052	0.079	0.094	0.058

不错。除此之外,本文算法的运行时间仅次于DIDFuse和CoCoNet,相比于最快的CoCoNet仅慢了11.54%。

2.3　消融实验

1)配准网络的消融实验

为了验证配准网络内各个模块和损失函数的作用,在TNO、Roadscene和自建数据集上进行了配准网络消融实验。在第1组实验里,将ORB算法去除,其他保持不变,记作w/o ORB。在第2组实验里,将特征增强模块去除,其他不变,记作w/o LFEM。第3组实验里,将多尺度特征提取和密集连接去除,其他不变,记作w/o M-D。第4组实验,将设计的损失函数去除,其他不变,记作w/o Loss。第5组实验为本文算法,记作w/o Ours。

配准网络消融实验的定量结果如表6所示。可以看出缺少多尺度特征提取和密集连接、损失函数的结果很差,MSE提高了5倍,NCC和MI也降低不少。缺少ORB算法和缺少LFEM的网络效果略差。综合来看,本文设计的模块和损失函数不可或缺。

表6　配准网络在3个数据集上的消融实验定量比较

Table 6　Quantitative comparison of registration networks in ablation experiments on three datasets

metric	TNO			Roadscene			自建数据集		
	MSE	NCC	MI	MSE	NCC	MI	MSE	NCC	MI
w/o ORB	0.005	0.893	1.897	0.006	0.884	1.879	0.008	0.883	1.472
w/o LFEM	0.005	0.884	1.889	0.007	0.794	1.783	0.007	0.793	1.447
w/o M-D	0.009	0.783	1.847	0.008	0.778	1.683	0.009	0.773	1.364
w/o Loss	0.010	0.732	1.784	0.009	0.735	1.573	0.009	0.693	1.412
w/o Ours	0.002	0.973	1.963	0.002	0.978	2.103	0.005	0.953	1.658

2)融合网络的消融实验

为了验证融合网络内各个模块的作用,在TNO、Roadscene和自建数据集上进行了融合网络消融实验。在第1组实验里,将光感增强模块去除,其他保持不变,记作w/o LEM。在第2组实验里,将swin transformer block去除,其他不变,记作w/o STB。第3组实验里,将自适应多尺度池化卷积去除,其他不变,记作w/o AMPC。第4组实验,将设计的EMA特征融合模块去除,其他不变,记作w/o EMA。第5组实验为本文算法,记作w/o Ours。

融合网络消融实验的定量结果如表7所示。第3组实验的4个指标明显下降,说明设计的自适应多尺度池化卷积对于多模态图像融合的重要性。第1、2、4组实验在4个指标略微下降,说明模块的必要性。

表7　融合网络在3个数据集上的消融实验定量比较

Table 7　Quantitative comparison of ablation experiments of fusion networks on three datasets

metric	TNO				Roadscene				自建数据集			
	EN	SD	SSIM	MI	EN	SD	SSIM	MI	EN	SD	SSIM	MI
w/o LEM	7.78	44.72	0.93	2.18	7.84	51.18	0.89	2.36	7.19	47.82	0.77	2.25
w/o STB	7.67	43.28	0.79	2.01	7.73	50.19	0.79	2.33	7.14	48.62	0.78	2.27
w/o AMPC	6.73	39.28	0.55	1.58	6.62	45.29	0.54	1.68	5.73	37.92	0.48	1.78
w/o EMA	7.27	41.82	0.68	1.87	7.25	48.28	0.68	2.45	6.89	44.38	0.56	1.94
w/o Ours	8.02	47.18	1.17	2.54	8.12	55.87	1.04	2.67	7.98	53.17	0.94	2.46

为了验证自适应多尺度池化卷积的作用,在TNO、Roadscene和自建数据集上进行了消融实验。在第1组实验里,将最大池化分支去除,其他保持不变,记作w/o MAXP。在第2组实验里,将平均池化分支去除,其他不变,记作w/o AVGP。第3组实验里,将动态卷积去除,其他不变,记作w/o DC。第4组实验为本文算法,记作w/o Ours。

自适应多尺度池化卷积消融实验的定量结果如表8所示。第1、2、3组实验在4个指标略微下降,说明不同池化和动态卷积模块的必要性。

表 8　自适应多尺度池化卷积在 3 个数据集上的消融实验定量比较

Table 8　Quantitative comparison of ablation experiments of AMPC on three datasets

metric	TNO				Roadscene				自建数据集			
	EN	SD	SSIM	MI	EN	SD	SSIM	MI	EN	SD	SSIM	MI
w/o MAXP	7. 73	44. 46	0. 91	2. 06	7. 38	51. 38	0. 84	2. 19	7. 22	47. 73	0. 57	2. 23
w/o AVGP	7. 64	43. 22	0. 74	2. 01	7. 74	50. 11	0. 74	2. 12	7. 18	48. 65	0. 58	2. 27
w/o DC	7. 73	43. 28	0. 85	1. 87	6. 84	45. 83	0. 74	1. 75	7. 33	47. 92	0. 68	2. 18
w/o Ours	8. 02	47. 18	1. 17	2. 54	8. 12	55. 87	1. 04	2. 67	7. 98	53. 17	0. 94	2. 46

为了验证 EMA 融合模块内部模块的作用，在 TNO、Roadscene 和自建数据集上进行了消融实验。在第 1 组实验里，将 EMA 去除，其他保持不变，记作 w/o EMA。在第 2 组实验里，将两个残差模块去除，其他不变，记作 w/o RES。第 3 组实验为本文算法，记作 w/o Ours。

EMA 融合模块消融实验的定量结果如表 9 所示。第 1、2 组实验的 4 个指标明显下降，说明内部的 EMA 和残差模块对于多模态图像融合的重要性。

表 9　EMA 融合模块在 3 个数据集上的消融实验定量比较

Table 9　Quantitative comparison of ablation experiments of EMA fusion module on three datasets

metric	TNO				Roadscene				自建数据集			
	EN	SD	SSIM	MI	EN	SD	SSIM	MI	EN	SD	SSIM	MI
w/o EMA	7. 45	42. 83	0. 73	2. 02	7. 35	51. 01	0. 79	2. 36	7. 46	47. 36	0. 67	2. 05
w/o RES	7. 35	44. 34	0. 78	1. 93	7. 67	48. 33	0. 98	2. 45	6. 94	51. 38	0. 76	2. 34
w/o Ours	8. 02	47. 18	1. 17	2. 54	8. 12	55. 87	1. 04	2. 67	7. 98	53. 17	0. 94	2. 46

2. 4　目标检测实验

为了进一步体现本文算法的优越性，本文使用最先进的检测器 YOLOv8[48] 对 Roadscene 数据集进行目标检测实验。为了保证公平的比较，本文使用了 YOLOv8s 模型，并将各种方法的融合结果直接输入到检测器中进行再训练。定性比较实验结果如图 9 所示。检测框显示了每个类别和相应的置信度。DIDFuse、UMF、CoCoNet、IMF 检测到的类别明显缺少，只能检测到其中两种类别。RFN 和 LRRNet 检测出的效果很好，但是置信度没有本文算法高。综合来看，本文算法在目标检测性能上表现最佳。

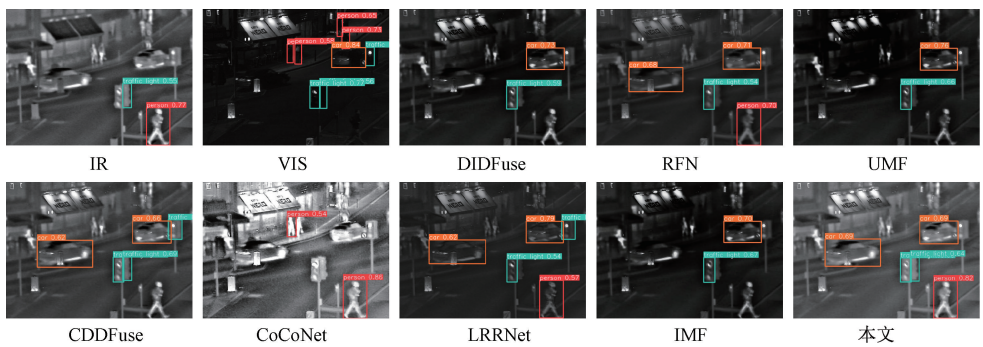

图 9　融合网络在数据集 Roadscene 上的目标检测性能定性比较

Fig. 9　Qualitative comparison of object detection performance of the fusion network on the dataset Roadscene

3　结　论

为了应对目前红外与可见光图像的难以配准以及融合效果不佳等问题，本文提出了一种自适应特征增强的多尺度红外与可见光图像配准和融合算法。具体来说，在配准网络中，使用 ORB 算法和特征增强模块提取特征，并利用多尺度卷积和密集连接进行配准。在融合网络中，利用光感增强模块增强可见光图像的特征，然后利用 swin transformer block 和自适应多尺度池化卷积提取全局和细节特征，最后利用注意力融合模块进行融合。未来仍有很多的工作需要开展，比如降低网络的参数量，如何应用到更多的多模态图像融合任务里。

参考文献

[1] LIU J, FAN X, HUANG Z, 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, 2022: 5802-5811.

[2] WANG D, LIU J, LIU R, et al. An interactively reinforced paradigm for joint infrared-visible image fusion and saliency object detection[J]. Information Fusion, 2023, 98: 101828.

[3] 李云红, 刘宇栋, 苏雪平, 等. 红外与可见光图像配准技术研究综述[J]. 红外技术, 2022, 44(7): 641-651.
LI Y H, LIU Y D, SU X P, et al. A survey of infrared and visible image registration techniques[J]. Infrared Technology, 2022, 44(7): 641-651.

[4] LI H, WU X J, KITTLER J. Infrared and visible image fusion using a deep learning framework[C]. 2018 24th International Conference on Pattern Recognition (ICPR). IEEE, 2018: 2705-2710.

[5] LIU J, FAN X, JIANG J, et al. Learning a deep multi-scale feature ensemble and an edge-attention guidance for image fusion[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 32(1): 105-119.

[6] ZHAO Z, XU S, ZHANG C, et al. DIDFuse: Deep image decomposition for infrared and visible image fusion[J]. ArXiv preprint arXiv:2003.09210, 2020.

[7] ZHANG X, YE P, LEUNG H, et al. Object fusion tracking based on visible and infrared images: A comprehensive review[J]. Information Fusion, 2020, 63: 166-187.

[8] ZHAO Z, BAI H, ZHANG J, et al. Cddfuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 5906-5916.

[9] XU H, MA J, JIANG J, et al. U2Fusion: A unified unsupervised image fusion network[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 44(1): 502-518.

[10] 黄玲琳, 李强, 路锦正, 等. 基于多尺度和注意力模型的红外与可见光图像融合[J]. 红外技术, 2023, 45(2): 143-149.
HUANG L L, LI Q, LU J ZH, et al. Infrared and visible image fusion based on multi-scale and attention model[J]. Infrared Technology, 2023, 45(2): 143-149.

[11] 陈潮起, 孟祥超, 邵枫, 等. 一种基于多尺度低秩分解的红外与可见光图像融合方法[J]. 光学学报, 2020, 40(11): 1110001.
CHEN CH Q, MENG X CH, SHAO F, et al. An infrared and visible image fusion method based on multi-scale low-rank decomposition[J]. Acta Optica Sinica, 2020, 40(11): 1110001.

[12] 陈永, 张娇娇, 王镇. 多尺度密集连接注意力的红外与可见光图像融合[J]. 光学精密工程, 2022, 30(18): 2253-2266.
CHEN Y, ZHANG J J, WANG ZH. Infrared and visible image fusion based on multi-scale densely connected attention[J]. Optics and Precision Engineering, 2022, 30(18): 2253-2266.

[13] 李天放,孙一宸,于明鑫,等.结合语义分割与跨模态差分特征补偿的红外与可见光图像融合方法[J].电子测量与仪器学报,2024,38(7):34-45.
LI T F, SUN Y CH, YU M X, et al. Infrared and visible image fusion method combining semantic segmentation and cross-modal differential feature compensation[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(7): 34-45.

[14] 童小钟,赵宗庆,苏绍璟,等.基于知识蒸馏自适应DenseNet的无人机对地目标可见光与红外图像融合[J].仪器仪表学报,2024,45(5):20-32.
TONG X ZH, ZHAO Z Q, SU SH J, et al. Visible and infrared image fusion for UAV-to-ground targets based on knowledge distillation-adaptive DenseNet[J]. Chinese Journal of Scientific Instrument, 2024, 45(5): 20-32.

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

[16] MA J, XU H, JIANG 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] 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: 72-86.

[18] LIU J, FAN X, HUANG Z, 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, 2022: 5802-5811.

[19] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Neural Information Processing Systems, Neural Information Processing Systems, 2017.

[20] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[J]. Computer Vision and Pattern Recognition, 2020.

[21] VS V, VALANARASU J M J, OZA P, et al. Image fusion transformer [C]. 2022 IEEE International Conference on Image Processing (ICIP). IEEE, 2022: 3566-3570.

[22] MA J, TANG L, FAN F, 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): 1200-1217.

[23] 陈彦林, 王志社, 邵文禹, 等. 红外与可见光图像多尺度 Transformer 融合方法[J]. 红外技术, 2023, 45(3): 266-275.
CHEN Y L, WANG ZH SH, SHAO W Y, et al. Infrared and visible image fusion method based on multi-scale transformer[J]. Infrared Technology, 2023, 45(3): 266-275.

[24] ILG E, MAYER N, SAIKIA T, et al. Flownet 2.0: Evolution of optical flow estimation with deep networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2462-2470.

[25] ZHOU S, TAN W, YAN B. Promoting single-modal optical flow network for diverse cross-modal flow estimation[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022: 3562-3570.

[26] ARAR M, GINGER Y, DANON D, et al. Unsupervised multi-modal image registration via geometry preserving image-to-image translation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 13410-13419.

[27] WANG D, LIU J, FAN X, et al. Unsupervised misaligned infrared and visible image fusion via cross-modality image generation and registration[J]. ArXiv preprint arXiv:2205.11876, 2022.

[28] RUBLEE E, RABAUD V, KONOLIGE K, et al. ORB: An efficient alternative to SIFT or SURF[C]. 2011 International Conference on Computer Vision. IEEE, 2011: 2564-2571.

[29] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial Transformer Networks[M]. Cambridge: MIT Press, 2015.

[30] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.

[31] LIANG J, CAO J, SUN G, et al. Swinir: Image restoration using swin transformer[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 1833-1844.

[32] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[33] OUYANG D, HE S, ZHANG J, et al. Efficient multi-scale attention module with cross-spatial learning[J]. ArXiv, 2023, abs12305.13563.

[34] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.

[35] WANG Z, SIMONCELLI E P, BOVIK A C. Multiscale structural similarity for image quality assessment[C]. The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers. IEEE, 2003: 1398-1402.

[36] XU H, WANG X, MA J. DRF: Disentangled representation for visible and infrared image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1-13.

[37] TANG L, DENG Y, MA Y, et al. Superfusion: A versatile image registration and fusion network with semantic awareness [J/OL]. IEEE/CAA Journal of Automatica Sinica, 2022: 2121-2137.

[38] WANG D, LIU J, MA L, 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.

[39] YANG N, YANG Y, LI P, et al. Research on infrared and visible image registration of substation equipment based on multi-scale retinex and ASIFT features[C]. Sixth International Workshop on Pattern Recognition. International Society for Optics and Photonics, 2021, 11913: 1191303.

[40] LI Y, WANG J, YAO K. Modified phase correlation algorithm for image registration based on pyramid[J]. Alexandria Engineering Journal, 2022, 61(1): 709-718.

[41] 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: 72-86.

[42] LIU J, LIN R, WU G, et al. Coconet: Coupled contrastive learning network with multi-level feature ensemble for multi-modality image fusion [J].

International Journal of Computer Vision, 2024, 132(5): 1748-1775.

[43] LI H, XU T, 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.

[44] MAHAPATRA D, GE Z, SEDAI S, et al. Joint registration and segmentation of Xray images using generative adversarial networks[C]. Machine Learning in Medical Imaging, Lecture Notes in Computer Science, 2018: 73-80.

[45] CAO X, YANG J, WANG L, et al. Deep learning based inter-modality image registration supervised by intra-modality similarity[J]. ArXiv Preprint arXiv: 1804.10735, 2018.

[46] HAN Y, CAI Y, CAO Y, et al. A new image fusion performance metric based on visual information fidelity[J]. Information Fusion, 2013, 14(2): 127-135.

[47] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[48] WANG G, CHEN Y, AN P, et al. UAV-YOLOv8: A small-object-detection model based on improved YOLOv8 for UAV aerial photography scenarios[J]. Sensors, 2023, 23(16): 7190.

作者简介

孙溪成,2022年于辽宁工程技术大学获得学士学位,现为辽宁工程技术大学硕士研究生,主要研究方向为多模态图像融合。
E-mail: 3079194134@qq.com

Sun Xicheng received his B. Sc. degree from Liaoning Technical University in 2022. Now he is a M. Sc. candidate at Liaoning Technical University. His main research interests include multimodal image fusion.

吕伏(通信作者),2021年于东北大学获得博士学位,现为辽宁工程技术大学副教授,主要研究方向为图像处理和人工智能。
E-mail: 13274280854@163.com

Lyu Fu (Corresponding author), received her Ph. D. degree from Northeastern University in 2021. Now she is an associate professor at Liaoning Technical University. Her main research interest includes image processing and artificial intelligence.

尹艺潼,2023年于辽宁工程技术大学获得学士学位,现为辽宁工程技术大学硕士研究生,主要研究方向为图像处理和网络安全。
E-mail: 1377584962@qq.com

Yin Yitong received her B. Sc. degree from Liaoning Technical University in 2022. Now she is a M. Sc. candidate at Liaoning Technical University. Her main research interests include Image processing and cybersecurity.