文章编号:1001-5078(2023)07-1052-08

·红外技术及应用·

引入特征交互的红外与可见光图像自适应融合

陈从平[1],闫焕章[1],郁春明[1],江高勇[1],凌　阳[2],戴国洪[1]
(1. 常州大学机械与轨道交通学院,江苏 常州 213164;2. 常州大学材料科学与工程学院,江苏 常州 213164)

摘　要:在图像融合领域,现有的基于卷积神经网络(CNN)或 Transformer 架构的方法存在两个局限性:首先,浅层纹理特征与深层语义特征之间无法有效聚合;其次,红外与可见光特征的权重比例无法自适应变化。本文提出一种引入特征交互的红外与可见光图像自适应融合方法。首先,构建一种基于 Transformer 的特征交互模块,聚合跨尺度特征信息,增强特征表达能力。其次,设计一种融合模块,自适应地调整特征权重比例。所提出的融合方法通过两阶段训练策略完成。第一个阶段,应用创新的特征交互概念训练编码器,增强特征表达,重建特征图像。第二个阶段,基于设计的权重自适应调整模块训练红外与可见光特征融合任务。公开数据集的实验结果表明,与现有方法相比,本方法在主观和客观的评价方面均优于其他典型方法。

关键词:图像融合;Transformer;特征交互;自适应融合;跨尺度
中图分类号:TP391.41;TN219　　**文献标识码:**A　　**DOI:**10.3969/j.issn.1001-5078.2023.07.011

Adaptive fusion of infrared and visible light images with introduction of feature interaction

CHEN Cong-ping[1], YAN Huan-zhang[1], YU Chun-ming[1], JIANG Gao-yong[1], LING Yang[2], DAI Guo-hong[1]
(1. School of Machinery and Rail Transit, Changzhou University, Changzhou 213164, China
2. School of Materials Science and Engineering, Changzhou University, Changzhou 213164, China)

Abstract: In the field of image fusion, existing methods based on Convolution Neural Network (CNN) or Transformer architectures have two limitations: first, effective aggregation between shallow texture features and deep semantic features is not possible; second, the weight ratio of infrared and visible light features cannot be adaptively changed. In this paper, an adaptive fusion method of infrared and visible images that introduce feature interaction is proposed. Firstly, a feature interaction module based on Transformer is constructed to aggregate cross-scale feature information and enhance feature representation. Secondly, a fusion module is designed to adaptively adjust the feature weight ratio. The proposed fusion method is completed by a two-stage training strategy. In the first stage, the encoder is trained using innovative feature interaction concepts to enhance feature representation and reconstruct feature images. In the second stage, the weight adaptive adjustment module based on the design trains the infrared and visible feature fusion task. The experimental results on publicly available datasets show that this method is superior to other typical methods in terms of subjective and objective evaluation compared with the existing methods.

Keywords: image fusion; Transformer; feature interaction; adaptive fusion; cross scale

基金项目:江苏省产业前瞻与关键核心技术－碳达峰碳中和科技创新专项资金项目(No. BE2022044)资助。
作者简介:陈从平(1976－),男,博士,教授,博士生导师,主要研究领域为机器视觉与机器学习。E-mail:mechencp@cczu.edu.cn
通讯作者:闫焕章(1996－),男,硕士研究生,主要研究领域为深度学习与图像融合。E-mail:771468952@qq.com
收稿日期:2022-09-14;**修订日期:**2022-11-01

1 引 言

红外成像传感器与可见光成像传感器在自动驾驶[1]、无人机[2]和视频监测[3]等领域得到了广泛的应用,其中,红外成像传感器所捕获的图像目标具有极高的像素亮度,不易受光照、天气等环境因素的影响,但其缺乏纹理细节信息;而可见光成像传感器捕获的图像具有丰富的纹理细节,但易受夜间、雨天、雾天等能见度低的环境影响,若能将红外图像与可见光图像融合,可对这两种图像的特征信息互补,显著提高目标识别的准确性。

传统的图像融合方法如金字塔变换[4]、小波变换[5]、轮廓波变换[6]、稀疏编码[7]及显著性融合方法[8]等,虽然取得了良好的融合性能,但存在以下问题:(1)泛化能力差,融合任务变化后,会导致模型迁移的效果大幅下降;(2)高度依赖人的主观经验,无法构建通用的特征提取方法和融合策略,完全基于设计良好的手工特征;(3)计算复杂度高,受人工定义特征的影响较大。

相对于传统的图像融合方法,深度学习在大数据驱动下的优势尤为显著,Li 等[9]在编码器中使用密集连接,加强中间层特征之间信息交流以提高表征能力,但忽略了图像的多尺度特征,致使特征提取过程中存在信息冗余。Yu 等[10]使用不同的融合策略实现图像融合,网络结构简单。Zhao 等[11]对双通道特征进一步学习,提取各不相同的特征,有效地在融合图像中保留了源图像显著特征,但融合策略简单,无法根据特征进行自适应融合。Ma 等[12]首次使用生成对抗网络实现图像融合任务,但视觉效果较差,缺少纹理效果。Wang 等[13]使用 Transformer 结构提取图像特征,但没有进行浅层纹理特征与深层语义特征之间的信息交互,造成融合结果中丢失纹理细节。

为了解决上述问题,本文提出了一种引入特征交互的红外与可见光自适应融合网络,旨在将编码器提取的多尺度特征通过特征交互模块,进行跨特征层的长距离信息交流,加强特征表达能力。这种跨空间、跨尺度的特征交互模块基于 Feature Pyramid Transformer[14],文中的特征交互模块主要分为两个部分,第一个部分是 Self-Transformer 模块,用于捕获特征图内的远距离依赖,建立全局依赖关系。第二部分是 Interaction-Transformer 模块,用于当前特征图与其他尺度特征图之间的特征交互,实现浅层纹理特征与深层语义特征信息的关联。此外,本文基于 SKNet[15]中动态重组特征方法,提出了动态融合模块,借助通道之间的特征关系去平衡红外特征与可见光特征,实现了红外特征与可见光特征融合权重的自适应调整。

2 融合方法

2.1 网络总体结构

引入特征交互的红外与可见光图像自适应融合方法的原理框架如图 1 所示,该网络是一个端到端的融合网络,在特征提取、特征交互、特征融合及特征重建过程中,共包括四个部分:编码器 Encoder、特征交互模块 Feature Interaction Module、融合模块 Fuse Module 和解码器 Decoder。其中编码器 Encoder 提取源图像 I^m 的多尺度特征,通过特征交互模块 FIM 学习浅层纹理特征与深层语义特征之间的关系,增强各尺度特征的表达能力,自适应融合红外与可见光特征,最后将融合特征重建,输出与输入相同大小的图像。

给定输入图像 $I^m \in R^{H\times W\times C_{\text{in}}}$($H$,$W$ 和 C_{in} 分别表示输入图像的高、宽和通道数,$m = ir$ 代表红外图像,$m = vis$ 代表可见光图像),在训练阶段图像被调整为固定大小。

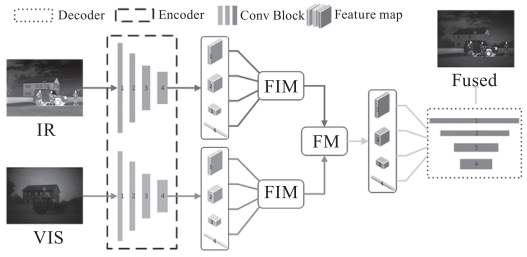

图 1 网络总体架构

Fig. 1 Overall network architecture

2.2 网络细节

2.2.1 编码器(Encoder)

如图 1 所示,多尺度特征 ϕ^m 由两个卷积核均为 3×3 的卷积块取得,其中第一个卷积操作的步长为 1,用于增加网络通道数,第二个步长为 2,用于筛除图像冗余的特征信息,经 4 次循环后,可提取源图像多尺度特征 $\phi^m = [\phi_1^m, \phi_2^m, \phi_3^m, \phi_4^m]$。

2.2.2 特征交互模块(FIM)

特征交互模块以编码器 Encoder 输出的多尺度

特征 $\phi_i^m \in R^{H_i \times W_i \times C_i}$ 作为输入，分别经过 Self-Transformer（ST）模块和 Interaction-Transformer（IT）模块，前者学习特征图 ϕ_i^m 自身的长距离依赖关系，捕获全局特征信息（如图 2（a）中 ST 模块），后者进行特征图 ϕ_i^m 与其他尺度特征图 ϕ_j^m 之间的信息交互（如图 2 中的 IT 模块），建立多尺度特征之间的关系（如图 2（b）中每个尺度与其他三个尺度交互得到与当前尺度相同大小的交互特征）。

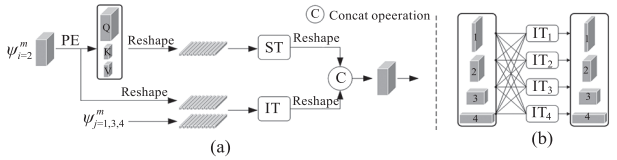

图 2 特征交互模块架构

Fig. 2 Feature interaction module architecture

①Self-Transformer（ST）：使用深度可分离卷积[16]对特征图 ϕ^m 进行位置编码，用于适应不同尺度的特征图作为输入，同时还能保证像素特征转换成序列后仍然保留空间位置信息，然后选用一层 Transformer[17] 网络提取位置编码后的特征图 ψ^m 的全局特征。

Block_{ST} 模块的网络结构如图 3（a），其中虚线框中的结构为 Transformer 结构，数据在输入 MSA（Multi-head self-attention）之前和经过 MSA 之后均进行层归一化 Layer Norm 操作调整数据分布，然后通过多层感知机 MLP 进行通道之间的信息交互，同时使用残差结构防止信息丢失。结构 MSA 在一个滑窗中计算局部关注度，与原 Transformer 结构中 MSA 的输入不同，本文输入序列向量来特征图自于 ψ_i^m 下采样和未下采样的两个尺度的特征，通过式（1）计算矩阵 Q、K 和 V，其中序列向量 $Z_i^m \in R^{H_i \times W_i \times C_i}$ 由特征图 $\psi_i^m \in R^{H_i, W_i, C_i}$ 转换而来，序列向量 $L_i^m \in R^{W_i \times H_i \times c_i / n^2}$（$n$ 为下采样倍数）由特征图 $\psi_i^m \in R^{H_i, W_i, C_i}$ 下采样后转换而来，以此计算特征图 ψ^m 的全局特征序列，减少计算复杂度。最后将全局特征序列转换成与特征图 ϕ_i^m 相同尺度的特征图 $\overline{\psi^m} \in R^{H_i \times W_i \times C_i}$。该计算过程可表示为：

$$Q = ZW_Q, K = LW_K, V = LW_V \tag{1}$$

$$\overline{\psi_i^m} = Reshape(Block_{ST}(Q_i^m, K_i^m, V_i^m))(i = 1, \cdots, 4) \tag{2}$$

②Interaction-Transformer（IT）：将特征图 ψ_i^m 与其他尺度特征图 ψ_j^m 之间的空间、通道特征信息进行交互，充分利用特征的多尺度结构。

通道信息交互模块 $\text{Block}_{\text{IT-Channel}}$ 的网络架构如图 3（b）所示，使用全局平均池化操作提取浅层特征图 ψ_i^m 的空间信息，并通过 1×1 卷积对通道进行升维得到浅层特征的空间特征信息，以该特征对深层特征 ψ_{i+1}^m 在通道上进行加权调整，完成浅层特征图与深层特征图之间信息交互，同时将未通过信息交互的浅层特征进行下采样，作为分支直接与信息交互后的特征相加，用于减少信息交互过程中因使用全局平均池化噪声造成的信息丢失，得到最终的通道信息交互特征 γ_{ic}^m。

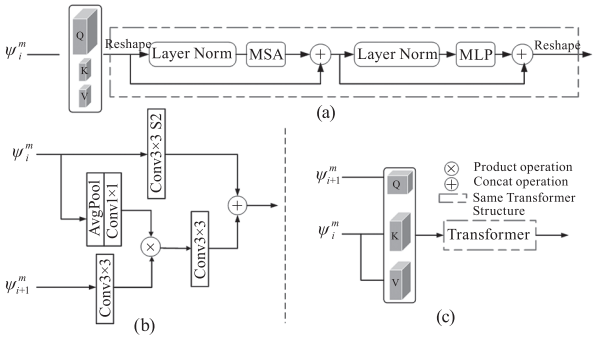

图 3 特征交互子模块网络结构

Fig. 3 Feature interaction submodule network structure

空间信息交互模块 $\text{Block}_{\text{IT-Spatial}}$ 的网络架构如图 3（c）所示，使用与 Self-Transformer（ST）中完全相同的 Transformer 结构，仅将输入从原来的同一特征图与其下采样特图更换成不同特征图。通过对深层特征图 ψ_i^m 与浅层特征图 ψ_j^m 在空间上进行信息交互，得到空间信息交互特征 γ_{is}^m。该结构中 MSA 以浅层特征 ψ_i^m 与深层特征 ψ_j^m 的序列化向量作为输入，计算局部区域内的关注度，以此调整依赖关系为：

$$\gamma_{ic}^m = Block_{IT-Channel}(\psi_i^m, \psi_j^m)(i, j = 1, \cdots, 4; j \neq i) \tag{3}$$

$$\gamma_{is}^m = Block_{IT-Spatial}(\psi_i^m, \psi_j^m)(i, j = 1, \cdots, 4; j \neq i) \tag{4}$$

最后，将 ST 模块、IT 模块及 Encoder 的输出特征图进行通道合并，再经过卷积核为 1×1 的卷积操作进行通道特征信息交互：

$$\alpha_i^m = Conv(Concat(\psi_i^m, \overline{\psi}_i^m, \gamma_{ic}^m, \gamma_{is}^m))(i, \cdots, 4) \tag{5}$$

2.2.3 融合模块（FM）

针对红外与可见光特征在最终融合图像的权重占比，本文基于 SKNet 动态调整权重的思想，设计了

一种自适应调整红外与可见光特征权重的融合结构(如图4 所示)。由于简单的红外特征与可见光特征逐像素相加也能得到不错的融合质量,故使用红外特征与可见光特征之和作为融合的引导特征,提取其通道特征,通过归一化后更新红外特征与可见光特征权重比例。

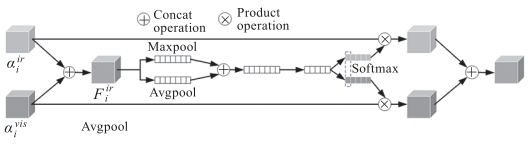

图4　融合网络结构

Fig. 4 Converged network structure

以红外特征 α_i^{ir} 与可见光特征 α_i^{vis} 之和作为引导特征 F_i :

$$\boldsymbol{F}_i = \alpha_i^{ir} + \alpha_i^{vis}(i = 1,\cdots,4) \tag{6}$$

提取 F 在空间上的全局平均特征 S_{ap} 以及全局最大池化特征 S_{mp},其中第 c 层特征按下式计算:

$$S_{ap}^{c} = \frac{1}{H \times W}\sum_{j=1}^{H}\sum_{k=1}^{W}\boldsymbol{F}_i^{c}(j,k) \tag{7}$$

$$S_{mp}^{c} = \max(\boldsymbol{F}_i^{c}) \tag{8}$$

最后,将红外特征分别与可见光特征的最大、平均特征相加,并计算红外与可见光基于通道的自适应权重:

$$q^{c} = \frac{\exp(S_{ap}^{ir-c} + S_{mp}^{ir-c})}{\exp(S_{ap}^{ir-c} + S_{mp}^{ir-c}) + \exp(S_{ap}^{vis-c} + S_{mp}^{vis-c})} \tag{9}$$

$$p^{c} = \frac{\exp(S_{ap}^{vis-c} + S_{mp}^{vis-c})}{\exp(S_{ap}^{ir-c} + S_{mp}^{ir-c}) + \exp(S_{ap}^{vis-c} + S_{mp}^{vis-c})} \tag{10}$$

其中, P^c 代表红外在第 c 层的自适应权重;q^c 代表红外在第 c 层的自适应权重,且 $p^c + q^c = 1$,基于通道注意力重建的红外特征 $\hat{\alpha}^{ir}$ 与可见光特征 $\hat{\alpha}^{vis}$ 计算如式(10):

$$\hat{\alpha}_i^{ir} = q_i \cdot \alpha_i^{ir},\quad \hat{\alpha}_i^{vis} = p_i \cdot \alpha_i^{vis},\quad (i = 1,\cdots,4) \tag{11}$$

2.2.4　解码器(Decoder)

如图2 所示,由多个转置卷积模块构成,利用跨尺度特征进行特征重建。考虑到不同尺寸在卷积和池化后尺寸不一致问题,在上采样后进行尺寸对齐保持特征尺寸一致,最后一个卷积层使用 Sigmoid 作为激活函数,保证计算损失时输入输出的数值范围一致,其他卷积层的激活函数均采用 LeakyReLU。

2.3　损失函数

为了保证特征交互模块在纹理特征和语义特征聚合功能的专向性,避免红外特征与可见光特征融合任务对特征聚合产生干扰,本模型分为两个阶段训练红外与可见光融合模型,第一个阶段完成编码器 Encoder、特征交互模块 Interaction Module 以及解码器 Decoder 的训练,第二个阶段完成融合模块的训练。

2.3.1 特征提取、交互及重建阶段

为了保证最终重建图像保留红外与可见光各自的图像特征,本文使用 $\boldsymbol{L}_{\text{ssim}}$ 结构相似度损失和 $\boldsymbol{L}_{\text{pixel}}$ 作为损失函数监督网络的训练。结构相似度函数能够以图像的结构、亮度和对比度三大属性来衡量两幅图像之间的相似性,在重建任务中可用于提取输入图像结构、亮度和对比度特征,如式(12)所示:

$$\boldsymbol{L}_{ssim} = 1 - SSIM(Output, Input) \tag{12}$$

同时,使用 L_{pixel} 对重建图像 $O_{rec}^{m}(m = ir,vis)$ 与输入图像 I^m 进行逐像素约束,其中 $\|\cdot\|_F$ 代表 Frobenius 范数:

$$\boldsymbol{L}_{pixel} = \| O_{rec}^{m} - I^{m} \|_F \tag{13}$$

第一个阶段训练时的总损失函数为:

$$\boldsymbol{L}_{stage1} = \omega_1 \boldsymbol{L}_{pixel}^{m} + \omega_2 \boldsymbol{L}_{ssim}^{m} \quad (m = ir,vis) \tag{14}$$

其中, ω_1 和 ω_2 是 $\boldsymbol{L}_{pixel}$ 损失与 $\boldsymbol{L}_{ssim}$ 损失之间的权衡参数。

2.3.2　特征融合阶段

固定第一个阶段训练好的编码器 Encoder、特征交互模块以及解码器 Decoder 的权重参数,通过约束红外特征、可见光特征与融合特征之间的感知损失[18]完成红外特征 α_i^{ir} 与可见光特征 α_i^{vis} 融合模块的训练,可使融合图像保留精细的结构信息同时保留着前景亮度和背景细节。感知损失 $\boldsymbol{L}_{\text{Perceptual}}$ 计算如式(14), α_i^f 表示第 i 层红外特征 α_i^{ir} 与可见光特征 α_i^{vis} 融合后的特征。

$$\boldsymbol{L}_{\text{Perceptual}} = \sum_{i=1}^{4} \| \alpha_i^f - (\alpha_i^m + \alpha_i^m) \| \tag{15}$$

为直接控制输出图像的亮度强弱与结构信息,本文使用结构相似性函数计算融合图像局部特征:

$$\boldsymbol{L}_{\text{local}}^{\boldsymbol{m}} = 1 - SSIM(Output, Input^m) \quad (m = ir, vis) \tag{16}$$

同时采用 L1 损失函数补充融合图像的可见光

细节特征,保证融合图像在整体上更符合人类视觉感知:

$$\boldsymbol{L}_{\text{global}} = \| O_{\text{rec}}^{m} - \boldsymbol{I}^{vis} \|_1 \tag{17}$$

阶段二的总损失函数如式(17)所示,其中 ω_3, ω_4,ω_5 为超参数,用于平衡三个损失函数差异。

$$\boldsymbol{L}_{\text{stage2}} = \omega_3 \boldsymbol{L}_{\text{local}}^{ir} + \omega_4 \boldsymbol{L}_{\text{local}}^{vis} + \omega_5 \boldsymbol{L}_{\text{global}} \tag{18}$$

3 实 验

3.1 实验参数设定

在训练阶段,选取 TNO[19]、RoadScene[20] 和 OTCBVS[21]等数据集中的部分红外与可见光图像为训练集,用于训练本文设计的双阶段模型。所有图像均以滑动步长为 20,裁剪为 224 × 224 大小的红外与可见光图像对,共 30595 对,灰度范围转换为[-1,1]。此外,选用优化器 Adam,学习率为 0.02,第一阶段每批次训练量为 16,第二阶段每批次训练量为 32。超参数[ω_1,ω_2,ω_3,ω_4,ω_5]分别为[2,5,5,2,3]。训练及测试配置为 Intel(R)Xeon(R)Gold 6248R CPU、128 GB RAM、NVIDIA GeForce GTX 3090 GPU。

在测试阶段,从 TNO 数据集选取部分图像作为验证集,将源图像的灰度范围转换为[-1,1],直接使用源图像大小。同时选取 6 种典型的融合方法,包括 DenseFuse、RFN-Nest、FusionGAN、IFCNN、U2Fusion[22]以及 SDNet 与本文方法进行比较,使用互信息(Mutual Information,MI)[23]、视觉信息融合保真度(the Visual Information Fidelity for fusion,VIF)[24]、相关差异和(the Sum of the Correlation Differences,SCD)[25]、信息熵(entropy,EN)[26]、标准差(Standard Deviation,SD)[27]和均方误差(mean squared error)共六项客观评价指标进行数据上的客观对比。

3.2 消融实验

3.2.1 跨尺度、跨空间信息交互有效性验证

为验证跨尺度、跨空间信息交互的有效性,设计四组实验对比验证,使用跨尺度、跨空间信息交互的一组记作 Y;不使用跨尺度、跨空间信息交互的一组记为 N,另外两组为 RFN-Nest 和 IFCNN 网络,它们在特征提取过程中均未进行跨尺度、跨空间信息交互。

如图 5 所示,图中方框突出显示显著目标,左下方框为放大的细节特征。N 组丢失了部分可见光细节,红外特征表现不显著,目标与背景对比度低,而 Y 组由于跨空间、跨尺寸信息交互,增强本地和远程依赖关系,生成了更清晰的内容。本文选取 TNO 数据集中的 3 组图像以及 6 个指标,分别从主观和客观进行评价,在图像 Kaptein 中,Y 组更精细地重建了林间树枝的细节,在图像 soldiers_with_jeep 中,Y 更清晰的重建了云层细节,同时,以上图像中的人、汽车等显著目标突出。从表 1 的客观评估数据上看,Y 组数据多项指标优于未进行信息交互的 RFN-Nest 网络、IFCNN 网络以及 N 组,证明特征交互模块的有效性。

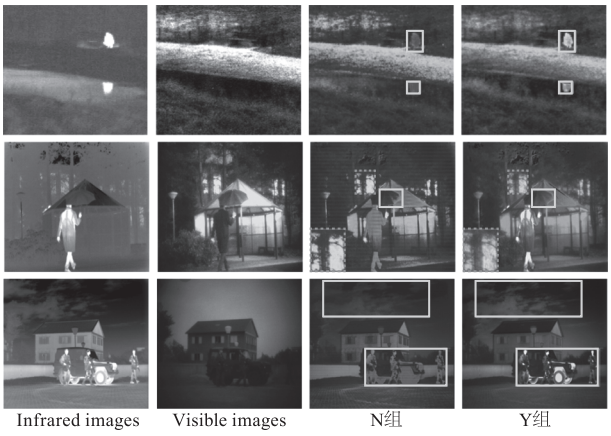

图 5 特征交互模块的定性比较结果

Fig. 5 Qualitative comparison results of feature interaction module

表 1 特征交互模块的定量比较结果

Tab. 1 Quantitative comparison results of feature interaction module

Method	MI	VIF	SD	MSE	EN	SCD
IFCNN	2.05268	0.78691	8.86915	0.04102	6.71019	1.71853
RFN-Nest	2.11841	0.81829	**9.12493**	0.04681	6.68766	1.72389
N 组	2.01622	0.67508	8.63791	0.04692	6.62328	1.59015
Y 组	**2.56670**	**0.84622**	9.10376	**0.03962**	**6.72526**	**1.72804**

3.2.2 自适应融合有效性验证

为验证自适应融合模块的有效性,使用 3 种经典的融合策略(add、average、$L_{1\text{-Norm}}$)作为对比实验,选取 TNO 数据集中三组图像进行主观分析,图 6 给出融合模块的定性比较结果,在求和融合策略 Sum 得到的图像中,虽然突出了可见光细节,但前景与背景的对比度区分不明显,整体图像偏暗。相反,平均融合策略 Average 得到的图像偏向于可见光图像,出现伪光晕。$L_{1\text{-Norm}}$无法更好平衡红外与可见光权重占比,出现细节不足、对比度不够的情况。与其他

融合策略相比,基于自适应融合方法获取的融合图像的背景细节特征精确,例如植被和云层,另外突出了目标的亮度,例如建筑、汽车和人。

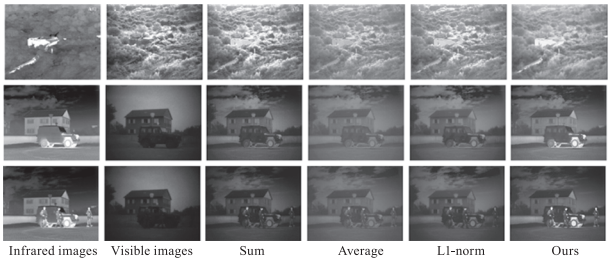

图 6 融合模块的定性比较结果

Fig. 6 Qualitative comparison results of fusion modules

3.2.3 参数的影响

在第二阶段的损失函数,本文使用超参数[ω_3, ω_4, ω_5]来平衡三个损失函数差异,在本次消融研究中,分别设置[1 :1 :1]、[5 :2 :5]、[5 :2 :3]共三组实验来验证不同比例的参数对融合性能的影响,选取 TNO 数据集中的 soldiers_with_jeep 图像从视觉效果进行分析,其中方框区域为分析目标。

如图 7 所示,视觉差别十分明显,当红外图像的权重系数越大,在突出细节特征的同时,融合图像中红外亮度特征的表现逐渐显著,并且并非整体统一变亮,例如图中表现为汽车与人的亮度明显增加。

图 7 参数的定性比较结果

Fig. 7 Qualitative comparison results of the parameters

3.3 TNO 数据集对比

为进一步验证本方法的优越性,选择 6 种具有代表性的融合方法,包括 DenseFuse、RFN-Nest、FusionGAN、IFCNN、U2Fusion 以及 SDNet,从 TNO 数据集中随机抽取 21 幅图片计算评价指标,同时选取其中 7 组红外与可见光图像作为效果展示,包括 Kaptein_1654、Bunker、Soldiers_with_jeep、Nato_camp、Sandpath、Marne_03 和 Bench。使用 6 个客观评价指标进行评估,包括 MI、VIF、SD、SCD、MSE 和 EN,加粗表示最佳值,加下划线表示次优值。

TNO 数据集的定性比较结果如图 8 所示,观察结果发现,FusionGAN 融合结果中仅保留了红外显著特征,图像背景偏暗,纹理细节严重缺失。IFCNN、DenseFuse 融合图像保留了可见光图像的纹理细节,但由于融合是简单的红外与可见光特征在通道上连接,无法更好的平衡红外与可见光特征的权重,导致图像对比度低,局部区域内的细节不明显。SDNet 融合结果中纹理信息保留完好,但红外特征占比较高,图像整体偏亮。U2Fusion 和 RFN-Nest 由于对特征重用以及在融合策略上的调整,其融合结果整体较为良好,在图像清晰的同时,局部区域内的对比度较高,但标记中的纹理细节存在丢失。与以上方法相比,本方法有意地将特征的浅层纹理与深层语义进行聚合,可在保留更多的纹理细节的同时,突出背景与前景的对比度,尤其在局部区域内突出了细节与对比度,如 Kaptein_1654 中树枝细节和 Bunker 中树丛的纹理层次分明。主观评价结果表明,本文方法能够取得较好的融合结果。

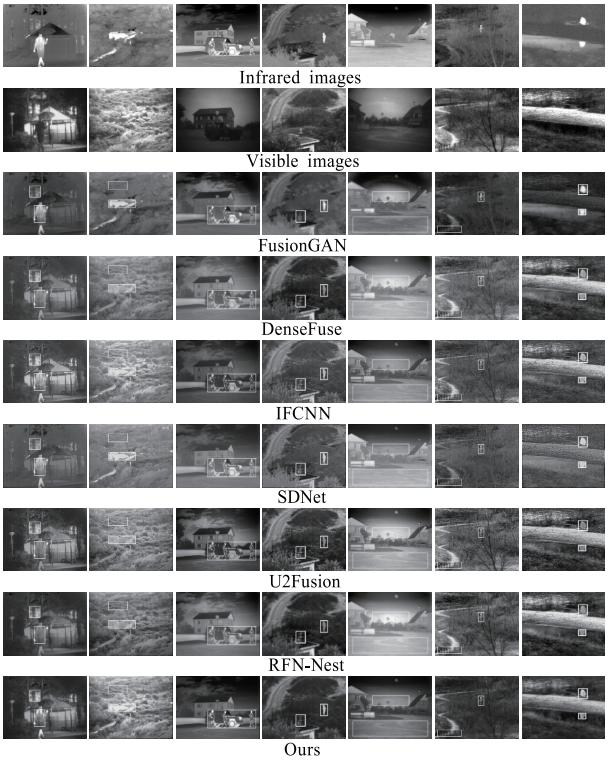

图 8 TNO 数据集的定性比较结果

Fig. 8 Qualitative comparison results of TNO datasets

表 2 的定量比较结果中,由于本方法通过加强特征信息跨空间和跨尺度交互,可保留源图像中更多的特征,故与源图像的均方误差 MSE 最小。表 2 互信息 MI 和信息熵 EN 指标表明,与仅关注局部或某个尺度的其他方法相比,本方法的融合图像包含的信息更多。VIF 指标表明,自适应调整融合图像中的红外与可见光特征,相比于固定权重的融合方法,能够在保证信息量的情况下,更好地平衡红外与

可见光特征,使融合图像更加符合人类视觉感知。表 2 中,本方法在客观评价指标 MI、VIF、SCD 和 MSE 均为最佳值,表明本方法在保留更多细节特征与的同时,突出显示红外亮度特征,客观评价指标从数据上表明,本方法是红外与可见光图像融合的有效融合架构。

表 2 TNO 数据集的定量比较结果

Tab. 2 Results of quantitative comparison of TNO datasets

Metrics	DenseFuse	RFN-Nest	FusionGAN	IFCNN	SDNet	U2Fusion	Ours
MI	2. 30190	2. 11841	2. 33522	2. 05268	2. 26058	2. 01020	**2. 56670**
VIF	0. 81745	0. 81829	0. 65413	0. 78691	0. 75917	0. 81967	**0. 84622**
SD	9. 03165	9. 12493	8. 43657	8. 86915	8. 84938	**9. 23972**	9. 10376
SCD	1. 72537	1. 72389	1. 37926	1. 71835	1. 55899	1. 72391	**1. 72804**
MSE	0. 04312	0. 04681	0. 05654	0. 04102	0. 04652	0. 04129	**0. 03962**
EN	6. 70418	6. 68766	6. 55802	6. 71019	6. 69482	**6. 99666**	6. 72526

4 结 论

本文提出了一种新颖有效的引入特征交互的红外与可见光自适应融合网络,结合卷积与自注意力机制聚合多尺度特征,自适应调整融合权重。网络由编码器 Encoder、特征交互模块 InteractionModule、FuseModule 和解码器 Decoder 四个部分组成。首先,输入图像经过编码器提取多尺度特征,使用构建的特征交互模块提取特征图自身的远距离依赖关系,以及不同尺度特征图之间浅层特征与深层语义特征之间的关联关系,增强特征表征能力,减轻特征信息丢失。其次,设计以红外特征与可见光特征之和作为引导特征,自适应调整红外特征与可见光特征的融合权重。消融实验验证了特征交互模块与融合模块的有效性,此外,本方法在公共数据集上与 6 种典型方法的主客观实验表明,本方法具有一定的优越性与鲁棒性。

参考文献:

[1] Zang Y, Zhou D, Wang C, et al. UFA-FUSE: a novel deep supervised and hybrid model for multi-focus image fusion [J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 1 – 17.

[2] Huang Yingjie, Mei Lingliang, Wang Yong, et al. Research on UAV detection based on infrared and visible image fusion [J]. Computer Knowledge and Technology, 222, 18 (7): 1 – 8. (in Chinese)
黄颖杰,梅领亮,王勇,等. 基于红外与可见光图像融合的无人机探测研究[J]. 电脑知识与技术,2022,18(7):1 – 8.

[3] Shen Ying, Huang Chunhong, Huang Feng, et al. Research progress of infrared and visible image fusion [J]. Infrared and Laser Engineering, 2021, 50(9): 152 – 169. (in Chinese)
沈英,黄春红,黄峰,等. 红外与可见光图像融合技术的研究进展[J]. 红外与激光工程,2021,50(9):152 – 169.

[4] Du J, Li W, Xiao B, et al. Union laplacian pyramid with multiple features for medical image fusion [J]. Neurocomputing, 2016, 194(jun. 19): 326 – 339.

[5] Pang H, Ming Z, Guo L. Multifocus color image fusion using quaternion wavelet transform [C]//International Congress on Image & Signal Processing, IEEE, 2013.

[6] Yang B, Li S, Sun F. Image fusion using nonsubsampled contourlet transform [C]//International Conference on Image & Graphics, IEEE Computer Society, 2007.

[7] Kwon H, Kaist Y, Lin S. Data-driven depth map refinement via multi-scale sparse representation [C]//Computer Vision & Pattern Recognition. IEEE, 2015.

[8] Han J, Pauwels E J, Zeeuw P D. Fast saliency-aware multi-modality image fusion [J]. Neurocomputing, 2013, 111(jul. 2): 70 – 80.

[9] Li H, Wu X J. Dense fuse: a fusion approach to infrared and visible images [J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614 – 2623.

[10] Yu Z A, Yu L B, Peng S C, et al. IFCNN: a general image fusion framework based on convolutional neural network [J]. Information Fusion, 2020, 54: 99 – 118.

[11] Zhao Z, Xu S, Zhang C, et al. DID fuse: deep image decomposition for infrared and visible image fusion [C]// Twenty-Ninth International Joint Conference on Artificial Intelligence and Seventeenth Pacific Rim International Conference on Artificial Intelligence, 2020.

[12] Ma J, Wei Y, Liang P, et al. Fusion GAN: a generative ad-

versarial network for infrared and visible image fusion [J]. Information Fusion, 2019, 48: 11 – 26.

[13] Wang Z, Chen Y, Shao W, et al. Swin fuse: a residual swin transformer fusion network for infrared and visible images [J]. ArXiv Preprint ArXiv: 2204. 11436, 2022.

[14] Zhang D, Zhang H, Tang J, et al. Feature pyramid transformer [C]//European Conference on Computer Vision. Springer, Cham, 2020: 323 – 339.

[15] Li X, Wang W, Hu X, et al. Selective Kernel Networks [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2020.

[16] Chollet F. Xception: deep learning with depthwise separable convolutions [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.

[17] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16 × 16 words: transformers for image recognition at scale [C]//International Conference on Learning Representations, 2021.

[18] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution [C]//European Conference on Computer Vision, Springer, Cham, 2016.

[19] A. Toet (2014). TNO image fusion dataset. figshare. data. [Online]. Available: https://figshare. co-m/articles/TN Image Fusion Dataset/1008029.

[20] H. Xu (2020). Roadscene database [EB/OL]. https://github. com/hanna-xu/RoadScene.

[21] S. Ariffin (2016). OTCBVS database [EB/OL]. http://vciplokstate. org/pbvs/bench/.

[22] 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 – 518.

[23] Eskicioglu A M, Fisher P S. Image quality measures and their performance [J]. IEEE Trans Commun, 1995, 43 (12): 2959 – 2965.

[24] 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.

[25] Aslantas, V, Bendes, et al. A new image quality metric for image fusion: the sum of the correlations of differences [J]. Aeu-international Journal of Electronics and Communications, 2015, 69 (12): 1890 – 1896.

[26] Aardt V, Jan. Assessment of image fusion procedures using entropy, image quality, and multispectral classification [J]. Journal of Applied Remote Sensing, 2008, 2 (1): 1 – 28.

[27] Rao Yunjiang. In-fibre bragg grating sensors [J]. Measurement Science and Technology, 1997, (8): 355 – 375.