小 型 微 型 计 算 机 系 统
Journal of Chinese Computer Systems
DOI:10.20009/j.cnki.21-1106/TP.2025-0182
2026 年 5 月 第 5 期
Vol.47 No.5 2026

频率空间协作的红外可见光图像融合网络

曹春红[1,2],蒋云云[1],王彩瑞[1]

[1](东北大学 计算机科学与工程学院,沈阳 110169)
[2](东北大学 医学影像智能计算教育部重点实验室,沈阳 110169)
E-mail:caochunhong@cse.neu.edu.cn

摘 要: 红外可见光图像融合旨在生成既突出显著目标又包含丰富纹理的图像.现有的融合方法主要关注空间域特征,忽略了频率域信息.因此,本文提出一种频率空间协作的红外可见光图像融合网络.首先通过频率分解模块将源图像分解为高频部分(模态特有特征)和低频部分(模态共有特征).同时,粗略提取源图像的空间域特征,以保留良好的空间结构.最后,利用跨域自适应融合模块学习自适应权重,动态调整频率域和空间域特征,以缓解域间差异并生成高质量融合图像.在 TNO 和 VOT2020-RGBT 数据集上的定量和定性实验结果表明,本文方法在 6 项评价指标上表现优异,且相比 7 种先进的融合方法,能更有效地融合多模态互补信息,生成显著性目标突出、细节丰富的融合图像.

关 键 词: 红外图像;可见光图像;图像融合;频率分解;跨域自适应融合
中图分类号: TP391 **文献标识码:** A **文 章 编 号:** 1000-1220(2026)05-1182-08

Frequency-spatial Collaboration Network for Infrared and Visible Image Fusion

CAO Chunhong[1,2], JIANG Yunyun[1], WANG Cairui[1]

[1](School of Computer Science and Engineering, Northeastern University, Shenyang 110169, China)
[2](Key Laboratory of Intelligent Computing of Medical Images, Ministry of Education, Northeastern University, Shenyang 110169, China)

Abstract: Infrared and visible image fusion aims to generate a single image highlighting salient objects and rich textures. Existing fusion methods predominantly focus on spatial characteristics while ignoring valuable frequency information. Therefore, this paper proposes a frequency-spatial collaboration network for infrared and visible image fusion. Firstly, the frequency decomposition module decomposes the source image into high-frequency (modality-specific features) and low-frequency components (modality-shared features). Simultaneously, the spatial characteristics are roughly extracted to preserve the spatial structure of the fused image. Finally, the cross-domain adaptive fusion module learns adaptive weights to dynamically adjust features in both frequency and spatial domains, thereby mitigating inter-domain differences and generating high-quality fused images. Quantitative and qualitative experimental results on the TNO and VOT2020-RGBT datasets demonstrate that the proposed method performs excellently across six evaluation metrics. Compared with seven state-of-the-art methods, it effectively integrates multi-modal complementary information and generates fused images with prominent salient targets and fine-grained texture information.

Keywords: infrared image; visible image; image fusion; frequency decomposition; cross-domain adaptive fusion

0 引 言

红外可见光图像融合的本质是将红外图像和可见光图像中的互补信息进行有效融合,从而生成一幅具有丰富细节信息和显著性特征的高质量图像.红外图像在凸显场景中的显著目标方面具有优势,并且其成像不易受到照明、天气等环境变化的影响.但红外图像通常包含较差的场景结构信息和纹理细节.相反,可见光图像较好地呈现了丰富的场景信息,但成像易受到环境条件的限制.红外可见光图像融合技术可以将两者的优势集成到单幅图像中,生成的融合图像在诸多领域有广泛应用,如目标检测[1-4]、RGB-T 跟踪[5]、场景理解[6]等.

近年来,越来越多的学者投入到红外可见光图像融合的研究中[7].传统的图像融合方法往往缺乏针对性的特征提取能力,且需要依赖先验知识进行设计,这严重限制了此类方法对复杂场景的适应能力,如基于多尺度变化的方法,基于稀疏表示的方法.

目前,基于深度学习的方法在整合源图像互补信息方面取得了显著的进步.DenseFuse[8]是最经典的基于自动编码器(AE)的图像融合方法之一,它使用 dense blocks 做为自动编码器的基础组件提取特征,使用相加和 L1 正则的融合策略生成融合图像.CAEFuse[9]提出了一种卷积自编码融合网络,它通过将卷积神经网络和自编码器相结合的方式改善融合性能.然而,上述方法均依赖手工制作的融合规则,这会导致生成不理想的结果.ProFuse[10]以 U-Net 为骨干提取输入图像的多尺度特征,并对多种层次的特征进行融合处理.为了促使网

收稿日期:2025-04-30 收修改稿日期:2025-05-29 基金项目:国家自然科学基金项目(U21A20487)资助. 作者简介:**曹春红**,女,1976 年生,博士,副教授,CCF 会员,研究方向为计算机视觉、图像处理、计算机图形学、计算机辅助设计;**蒋云云**,女,2001 年生,硕士研究生,研究方向为计算机视觉、图像处理;**王彩瑞**,女,2004 年生,研究方向为计算机视觉、图像处理.

络有目的地提取和融合有意义的特征,SeAFusion[11] 以高层语义信息的返回流指导融合任务的执行,提高了融合图像在高层视觉任务上的性能,但此方法在极端环境下会生成较差的融合结果. LRRNet[12] 是一个轻量的融合网络,它通过分解源图像实现多模态图像融合,但其生成的图像整体亮度偏低,不符合人类的视觉感知. 基于生成对抗网络(Generative Adversional Networks,GAN)的图像融合方法能很好地规避红外可见光图像融合任务缺少真实值图像这一挑战. FusionGAN[13] 是首个处理图像融合任务的生成对抗网络,它将红外可见光图像融合看作是生成器与判别器之间的博弈,判别器控制生成器合成包含更多细粒度特征的融合图像. 随后,GANMcC[14] 提出使用双判别器改善融合质量,但是基于 GAN 的融合方法仍面临着训练不稳定的问题. 然而,上述方法的网络架构均依赖于普通卷积,这限制了它们对全局特征的提取能力. 因此,YDTR[15] 提出将提取局部特征的卷积神经网络(CNN)和建模远程依赖关系的 Transformer 结合使用,以优化融合图像的质量. 类似地,孙等人[16] 也提出了一种红外与可见光图像分组融合的视觉 Transformer,但这种结合 Transformer 的网络结构在训练阶段增加了计算开销,对设备性能的要求也较高.

然而,现有技术主要集中研究源图像的空间域特征,忽略了有价值的频率域信息. 针对这一技术难题,本文提出了一个频率空间协作的红外可见光图像融合网络(Frequency-Spatial Collaboration Network for Infrared and Visible Image Fusion, FSCN). 具体来说,本研究设计了一个频率分解模块(Frequency Decomposition Module,FDM),它通过将八度卷积[17] 作为基本组件实现了对源图像频率域信息的在线学习. 同时,为了保护融合结果的空间结构,图像的空间域信息依然受到关注. 此外,考虑到频率域和空间域之间的域间差异,本文引入了一个跨域自适应融合模块(Cross-Domain Adaptive Fusion Module,CDAFM)来集成两域的特征,生成最终的融合结果. 实验结果表明,本文方法生成的融合结果能在强调显著性对象的同时保留丰富的细粒度特征. 本文的主要贡献总结如下:

1)本文提出一个频率空间协作的红外可见光图像融合网络,该网络能够深入挖掘源图像的频率特征,并有效提取空间域的粗略融合特征. 通过频率和空间两域特征的协同作用,最终生成高质量的融合图像.

2)考虑到现有方法对频率信息的忽视,本文设计了频率分解模块,它通过在线学习的方式,灵活地将源图像分解为低频部分和高频部分.

3)针对频率域和空间域之间存在差异的问题,本文引入了一个跨域自适应融合模块,其通过学习自适应权重,实现了两域特征的最优融合.

4)本文方法在多个数据集上均表现出色,实现了良好的融合效果. 定性实验和定量实验表明,本文方法在红外可见光图像融合任务中展现出较强的竞争力.

1　八度卷积

Chen 等人[17] 假设“卷积层的输出特征映射可以看作是不同频率信息的混合”,并提出了可以将图像划分为低频分量和高频分量的八度卷积. 低频分量对应于具有温和强度变换的像素点,例如大色块,通常代表场景的基本结构和物体的主要轮廓;高频分量是指变化强烈的像素,例如图像中物体的纹理边缘. 此外,该工作表明八度卷积可以很容易地插入到任意网络中,并保证较少的内存占用和较低的计算成本.

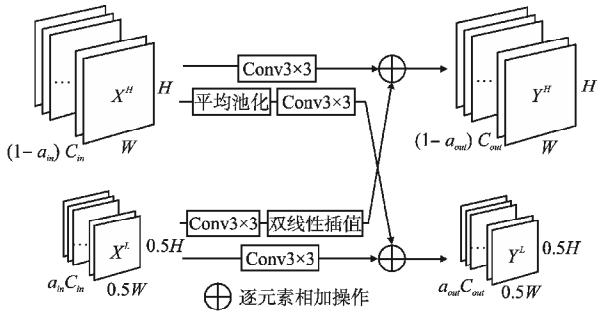


图 1　八度卷积

Fig. 1　Octave convolution

八度卷积的结构如图 1 所示,它通过 4 条处理路径实现了相同频率特征内部的信息更新和不同频率特征之间的信息交换,其工作过程可表示为公式(1):

$$\begin{aligned} Y^H &= Conv_{3\times3}(X^H) + Up(Conv_{3\times3}(X^L), X^H) \\ Y^L &= Conv_{3\times3}(X^L) + Conv_{3\times3}(pool(X^H, 2)) \end{aligned} \tag{1}$$

其中,$Conv_{3\times3}(\cdot)$代表 3×3 卷积,$pool(X,k)$代表核大小为 $k\times k$ 的平均池化操作,$Up(X,Z)$代表通过双线性插值的方式将 X 上采样到 Z 的大小.

2　本文方法

2.1　网络总体结构

给定一对可见光图像 I_{vi} 和红外图像 I_{ir},本文提出的频率空间协作的红外可见光图像融合网络(FSCN)的目标是生成一个含有丰富细节信息和显著特征的高质量融合图像 I_f. 如图 2 所示,FSCN 由频率分解模块、空间特征提取模块和跨域自适应融合模块 3 部分构成.

具体来说,频率分解模块将源图像分解为低频部分和高频部分$[\phi_l^x, \phi_h^x] = E(x, \Theta_E)$,其中 $x \in \{ir, vi\}$,Θ_E 表示频率分解模块的可学习参数. 然后,本文使用 4 个普通卷积将多模态图像的高频特征和低频特征转换到图像域($I_l^{ir}, I_h^{ir}, I_l^{vi}, I_h^{vi}$). 在图像域内,通过两个相加操作和两个卷积层,即可得到两个模态对应频率的融合特征(ϕ_l, ϕ_h). 同时,为了保证融合结果具有良好的空间结构,本文使用由 1 个卷积组成的空间特征提取模块获得多模态图像在空间域上的粗略融合结果 ϕ_s. 最后,将频率特征(ϕ_l, ϕ_h)和空间特征 ϕ_s 喂入跨域自适应融合模块,该模块通过学习到的自适应权重,协调两域特征的融合比重以生成高质量的融合结果 I_f.

2.2　频率域分解模块

在红外可见光图像融合任务中,两模态图像的低频部分代表模态的相同信息,高频特征代表各自模态的独特特征. 例如,对于表示同一场景的红外图像和可见光图像,两个模态的低频分量对应相同的背景或者关键物体的主要轮廓等关联性较强的内容;高频分量则对应各自模态中的特有信息,如可见

光图像中的纹理细节和红外图像中的热辐射信息. 受此启发，本文提出了一种能够将多模态图像分解为高频特征和低频特征的频率分解模块(Frequency Decomposition Module，FDM)，其结构如图 2(a)所示.

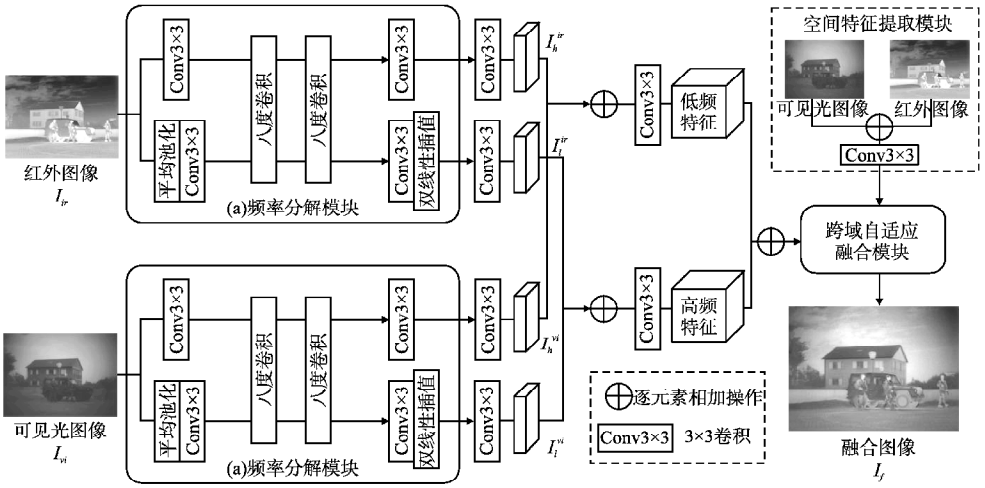


图 2 本文方法的总体框架图

Fig. 2 Overall network structure

FDM 首先通过 3×3 卷积和池化操作获得多模态图像对应的初始高频特征和低频特征，然后利用八度卷积实现对初始频率特征的同频信息更新和不同频信息的交换，以充分提取到多模态图像的频率信息，考虑到方法性能和网络结构的计算量和复杂度，本文的 FDM 模块共使用了 2 个八度卷积. 最后再利用卷积层将八度卷积处理的频率信息映射到特征域，并利用双线性插值保持提取到的高低频特征大小一致，以便网络对特征的进一步处理. 考虑到两种模态之间的差距，本文倾向于使用两个结构相同但参数不同的 FDM 分别提取多模态图像的频率信息. FDM 的工作过程可表示为公式(2)：

$$\begin{aligned}\phi_h^x &= Conv_{3\times3}(Oct(Oct(Conv_{3\times3}(I_x))))\\ \phi_l^x &= Up(Conv_{3\times3}(Oct(Oct(Conv_{3\times3}(pool(I_x,2))))),\phi_h^x)\end{aligned} \tag{2}$$

其中，$x\in\{ir,vi\}$，$Conv_{3\times3}(\cdot)$代表 3×3 卷积，$Oct(\cdot)$表示八度卷积，$pool(X,k)$代表核大小为 $k\times k$ 的平均池化操作，$Up(X,Z)$代表通过双线性插值的方式将 X 上采样到 Z 的大小.

2.3 跨域自适应融合模块

如图 2 所示，频率空间协作的红外可见光图像融合网络(FSCN)通过将可见光图像 I_{vi}和红外图像 I_{ir}相加并喂入由 1 个普通卷积组成的空间特征提取模块 $F(\cdot)$，即可获得多模态图像在空间域上的粗略融合结果 $\phi_s=F(I_{ir}+I_{vi})$，其包含良好的空间结构特征. ϕ_s 可被视为对频率分解模块提取的频率特征(ϕ_l,ϕ_h)的补充信息，两域特征协同合作从而提升网络的融合性能.

考虑到频率域和空间域之间的域间差异，如图 3 所示，本

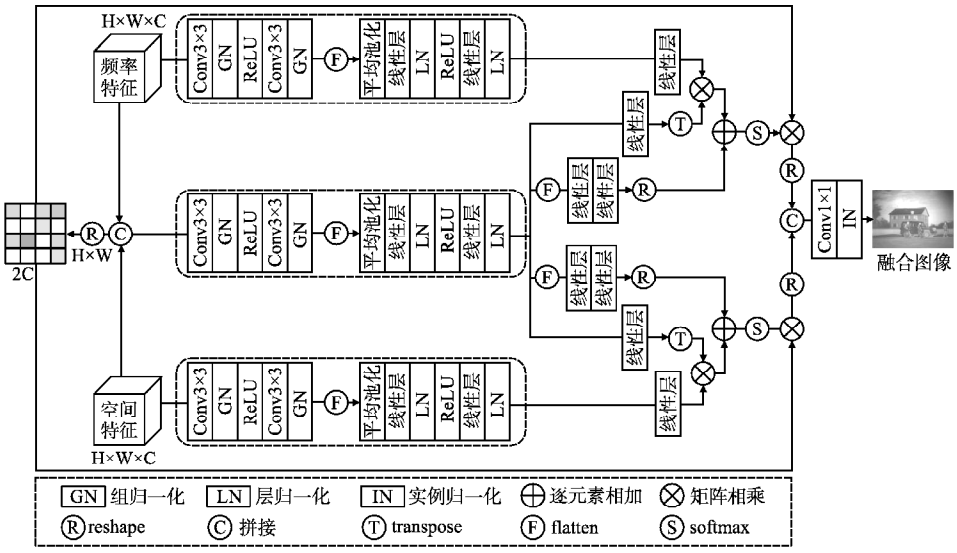


图 3 跨域自适应融合模块

Fig. 3 Cross-domain adaptive fusion module

文引入了一个跨域自适应融合模块(Cross-Domain Adaptive Fusion Module,CDAFM)来集成两域的特征,它在计算相似度得分时引入了一个可学习参数 W,从而实现了对两域特征的动态调整. 具体来说,本文将高低频特征相加得到频率粗略融合 $\phi_f=\phi_l+\phi_h$,然后将 $\phi_f\in R^{H\times W\times C}$ 变形后喂入到线性层生成 $Q_f\in R^{C\times d}$. 采用相同的策略,从空间粗略融合特征 $\phi_s\in R^{H\times W\times C}$ 中获得 $Q_s\in R^{C\times d}$,其中 C 代表特征通道数. 此外,将 ϕ_s 和 ϕ_f 拼接得到包含两个域全部信息的初始特征 $\phi_{init}\in R^{H\times W\times 2C}$. 类似地,将 ϕ_{init} 变形后喂入到线性层生成 $[K_s,K_f,W_s,W_f]$,其中 K_s 和 K_f 的维度为 $R^{2C\times d}$,W_s 和 W_f 的维度为 $R^{C\times 2C}$,融合值 $V\in R^{2C\times L}$ 由 ϕ_{init} 变形得到,其中 $L=H\times W$. 随后,将键值对 (Q_s,K_s),(Q_f,K_f) 分别相乘,所得结果与对应的可学习参数 (W_s,W_f) 相加,再经过 softmax 函数处理即可获得最终的相似性得分 S_s 和 S_f. 最后,将相似性得分 (S_s,S_f) 与融合值 V 相乘得到融合图像 I_f. 令 $Conv_{1\times1}(\cdot)$ 代表 1×1 卷积,$Softmax(\cdot)$ 和 $R(\cdot)$ 分别代表 softmax 函数和 reshape 变形,$[\cdot]$ 表示拼接操作. 则 CDAFM 的工作过程可以表示为公式(3):

$$S_f=Softmax\left(\frac{Q_fK_f^T+W_f}{\sqrt{d}}\right)$$

$$S_s=Softmax\left(\frac{Q_sK_s^T+W_s}{\sqrt{d}}\right)$$

$$I_f=Conv_{1\times1}([R(S_fV),R(S_sV)]) \tag{3}$$

2.4　损失函数

为了使融合图像在保留红外图像和可见光图像细粒度纹理信息的同时突出显著性特征. 如公式(4)所示,本文定义了训练网络的损失函数 I_{over},它由强度损失 $I_{intensity}$ 和梯度损失 I_{grad} 两部分组成:

$$L_{over}=L_{intensity}+L_{grad} \tag{4}$$

从像素强度的角度分析,强度损失 $I_{intensity}$ 控制融合图像的细节信息和整体结构信息达到最优权衡,它被定义为公式(5)的形式:

$$L_{intensity}=\frac{1}{HW}\|I_f-max(I_{ir},I_{vi})\|_2 \tag{5}$$

其中,I_{ir},I_{vi} 和 I_f 分别代表红外图像,可见光图像和融合图像. H 和 W 表示图像的高度和宽度. $max(\cdot)$ 代表逐个元素最大选择操作,$\|\cdot\|_2$ 代表 L2 范数.

此外,本文使用梯度损失 I_{grad} 确保融合图像保留多模态图像重要的细粒度纹理特征. 令 ∇ 表示 Sobel 梯度算子,$|\cdot|$ 表示绝对值操作,I_{grad} 可表示为公式(6):

$$L_{grad}=\frac{1}{HW}\|\,|\nabla I_f|-max(|\nabla I_{ir}|,|\nabla I_{vi}|)\,\|_2 \tag{6}$$

3　实　验

3.1　实验设置

3.1.1　数据集

本文在 MSRS[18] 训练集上训练网络,该训练集包含 1083 对图像. 为了证明模型的泛化性,本文遵循 LRRNet[12] 和 CrossFuse[19] 的测试策略,使用包含 21 对红外可见光图像对的 TNO[20] 数据集和包含 40 对红外可见光图像对的 VOT2020-RGBT[21] 作为测试集. 具体而言,TNO[20] 数据集中的图像场景更为复杂,其可见光图像中往往缺失显著性目标. 而 VOT2020-RGBT[21] 中的图像则大多展现了含有较小显著性目标的街道场景. 并且,上述两个测试集中的图像尺寸均不固定,呈现出多样性.

3.1.2　实现细节

本文的网络模型在 Pytorch 中实现,在 NVIDIA RTX 3090 GPU 上进行训练. 在训练阶段,所有 MSRS 训练集图像都将转换为灰度图并调整为 128×128 的大小喂入网络. 本文设置学习率为 1×10^{-5},批大小和历元分别为 8 和 30.

3.1.3　对比方法和评价指标

本文选择了 7 种当前较先进的主流融合方法与本文方法进行对比,以证明方法的优越性. 包括:DenseFuse[8]、MFEIF[22]、DeF[23]、YDTR[15]、DDFM[24]、LRRNet[12] 和 CS2Fusion[25].

本文使用了 6 种指标从多个方面来客观评估融合效果,包括:熵(EN)[26]、标准差(SD)[27]、空间频率(SF)[28]、视觉信息保真度(VIF)[29]、平均梯度(AG)[30] 和加权特征互信息(FMI_w)[31].

熵(EN)[26] 是度量图像信息量的客观评价指标,EN 值越高,融合图像信息越丰富. 令 $p=\{p_1,p_2,\cdots,p_n\}$ 表示图像的灰度分布,p_f 表示图像中灰度值为 f 的像素所占的比例,n 为灰度等级,其可表示为公式(7):

$$EN=-\sum_{f=1}^{n}p_f\log p_f \tag{7}$$

标准差(SD)[27] 用来衡量融合图像的对比度,SD 值越高说明图像对比度和细节越丰富. 令 μ 表示融合图像的灰度均值,M,N 表示图像的宽和高,$F(i,j)$ 表示图像在第 i 行第 j 列对应的像素值,则可定义标准差为公式(8):

$$SD=\sqrt{\sum_{i=1}^{M}\sum_{j=1}^{N}(F(i,j)-\mu)^2} \tag{8}$$

空间频率(SF)[28] 用于评估图像细节的丰富度,SF 值越高说明纹理信息越丰富. 令 F_L,F_U,F_{LU} 和 F_{RU} 分别为左、上、左上、右上方向上的一阶梯度,则融合图像的空间频率可表示为公式(9):

$$SF=\sqrt{(F_L)^2+(F_U)^2+(F_{LU})^2+(F_{RU})^2} \tag{9}$$

平均梯度(AG)[30] 通过计算平均梯度来评估图像的清晰度,AG 值越大意味着融合图像包含越丰富的梯度信息,即图像细节更清晰. 令 M,N 表示图像的宽和高,$F(i,j)$ 表示图像在第 i 行第 j 列对应的像素值,其可定义为公式(10):

$$AG=\frac{1}{(M-1)(N-1)}\times\sum_{i=1}^{M-1}\sum_{j=1}^{N-1}\times\sqrt{\frac{(F(i+1,j)-F(i,j))^2+(F(i,j+1)-F(i,j))^2}{2}} \tag{10}$$

视觉信息保真度(VIF)[29] 用于评估融合图像的视觉信息保真度,VIF 值越高表示融合图像在视觉效果上越真实;加权的特征互信息(FMI_w)[31] 用于评估融合结果和源图像之间的相似性,FMI_w 越高,意味着融合图像中包含越多的源图像信息. VIF 和 FMI_w 的计算公式较为复杂,详细计算过程可参考相应的参考文献.

3.2 TNO 数据集对比实验

3.2.1 定量比较

在 TNO[20] 数据集上的定量比较结果如表 1 所示，本文方法在 EN、SD、SF、AG 和 FMI_w 指标上均排名第一，在 VIF 指标上的测试结果仅次于 CS2Fusion[25] 排名第二.

以上数据有力地证实了以下结论：首先，较大的 EN 和 SD 证明本文方法能够生成信息更丰富、对比度更强的融合结果. 此外，在指标 AG 和 SF 上获得第一，表明本文方法生成的融合图像在细节清晰度和保护细粒度特征方面具有较好的表现. VIF 指标用来量化融合图像与源图像视觉信息一致性，虽然本文方法在 VIF 上仅次于 CS2Fusion[25] 排名第二，但在 FMI_w 指标上本文方法达到最优，这表明本文方法可以保留更多的源图像特征. 此外，结合图 4 的定性结果可以看出，本文方法生成的融合结果更符合人类视觉感知，在红外可见光图像融合任务中更具有优势.

表 1 不同融合方法在 TNO[20] 数据集上的定量比较

Table 1 Quantitative comparison of different fusion methods on TNO[20] dataset

方法	年份	EN ↑	SD ↑	SF ↑	VIF ↑	AG ↑	FMI_w ↑
DenseFuse[8]	2019	6.1738	22.5446	6.0508	0.5343	2.3443	0.4166
MFEIF[22]	2021	6.5386	30.5456	6.6998	0.6213	2.6423	0.4030
DeF[23]	2022	6.3309	26.3028	5.6513	0.5091	2.2885	0.2800
YDTR[15]	2022	6.2268	24.0540	6.9244	0.5461	2.4679	0.3560
DDFM[24]	2023	6.7155	32.5963	7.4420	0.2912	3.0133	0.2450
LRRNet[12]	2023	6.8384	39.4994	9.3306	0.5508	3.5841	0.3290
CS2Fusion[25]	2024	6.6670	36.6245	8.9417	**0.7968**	3.3125	0.3702
本文方法		**6.9268**	**40.9632**	**10.3726**	0.7890	**3.9815**	**0.4352**

注：加粗字体表示各列最优结果，加下划线表示各列次优结果.

3.2.2 定性比较

为了进一步说明方法的有效性，本文在 TNO[20] 数据集中的 21 对图像上进行了定性分析. 如图 4 所示，本文方法能够很好地保留源图像中的互补信息，生成更符合人类视觉感知的高质量图像. 为了便于观察，本文对各类方法的融合结果进行了局部放大处理.

对于云层特征，DenseFuse[8]、MFEIF[22]、DeF[23] 和 YDTR[15] 虽然保留了云层结构，但是它们生成的融合结果整体亮度偏暗，影响视觉观察. DDFM[24] 生成的融合结果中包含了大量的伪影，LRRNet[12] 的处理使云层部分被模糊化，而 CS2Fusion[25] 的融合结果中甚至直接丢失了云层特征. 相比之下，本文方法在保留源图像信息方面具有优势. 针对窗户边缘的处理，DenseFuse[8]、MFEIF[22]、DeF[23]、YDTR[15]、DDFM[24] 和 CS2Fusion[25] 均产生了模糊不清的轮廓. 而 LRRNet[12] 生成的结果整体亮度偏暗不利于人们观察，例如窗户旁边的树. 相比之下，本文方法在保护源图像的细粒度特征方面效果良好.

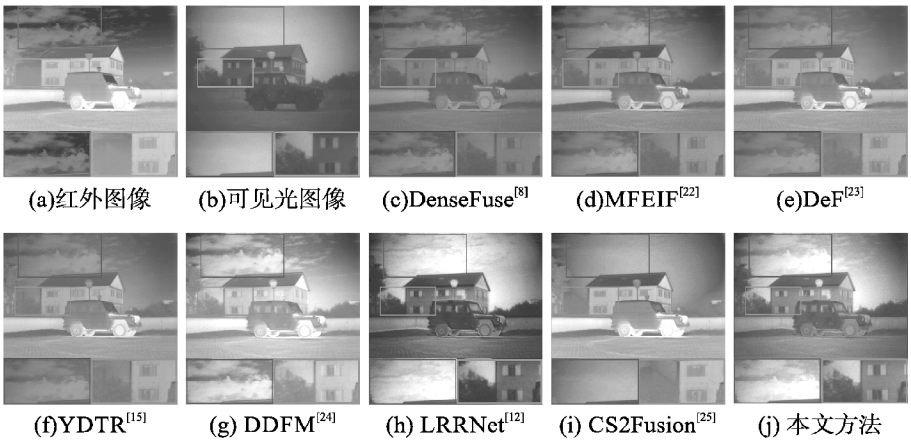


图 4 不同融合方法在 TNO[20] 数据集上的定性比较

Fig. 4 Qualitative comparison of different fusion methods on TNO[20] dataset

3.3 VOT2020-RGBT 数据集对比实验

3.3.1 定量比较

为了证明模型的泛化能力，本文在 VOT2020-RGBT[21] 数据集上同样进行了定量比较和定性比较. 如表 2 所示，本文方法在 VOT2020-RGBT[21] 数据集上实现了全指标领先，本文选用的 6 个指标分别从信息量、细节丰富度、对比度、保真度等多个角度对生成结果进行了较为全面的评估. 因此，表 2 中的定量比较数据可以有力地证明本文方法的综合性能显著优于其他方法. 特别是在 VIF 指标上，本文方法超越了在 TNO[20] 数据集上表现最优的 CS2Fusion[25] 方法，这表明本文方法相比于 CS2Fusion[25] 更能关注到较小的显著性目标，更适用于自动驾驶等现实应用.

表 2　不同融合方法在 VOT2020-RGBT[21] 数据集上的定量比较

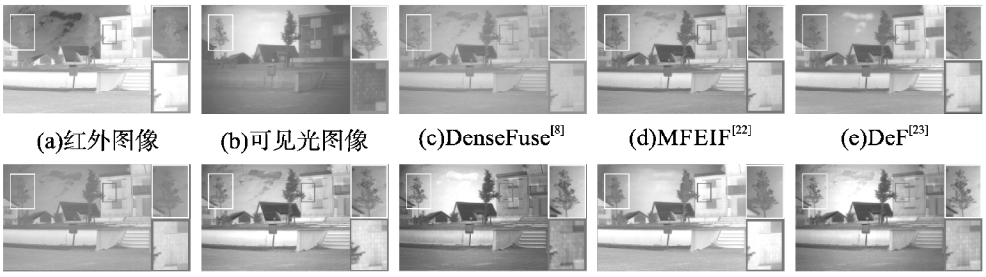
Table 2　Quantitative comparison of different fusion methods on VOT2020-RGBT[21] dataset

方法	年份	EN ↑	SD ↑	SF ↑	VIF ↑	AG ↑	FMI_w ↑
DenseFuse[8]	2019	6.3568	25.8544	6.6471	0.5829	2.4575	0.4188
MFEIF[22]	2021	6.6923	34.6952	7.5791	0.6792	2.8714	0.4059
DeF[23]	2022	6.5306	30.7549	6.5467	0.5604	2.5077	0.2804
YDTR[15]	2022	6.4047	29.0864	8.3152	0.6049	2.8172	0.3529
DDFM[24]	2023	6.8188	34.6992	7.3361	0.2675	2.8799	0.2142
LRRNet[12]	2023	<u>6.9645</u>	<u>41.5869</u>	<u>9.4734</u>	0.5751	<u>3.5341</u>	0.3324
CS2Fusion[25]	2024	6.8098	39.4802	9.1304	<u>0.7424</u>	3.377	<u>0.3746</u>
本文方法		**6.9799**	**41.8115**	**10.2662**	**0.7872**	**3.7865**	**0.4359**

注:加粗字体表示各列最优结果,加下划线表示各列次优结果.

3.3.2　定性比较

如图 5 所示,为了证明本文方法在 VOT2020-RGBT[21] 数据集上的视觉效果仍具有优势,同样,本文进行了定性比较实验,并对各类方法的融合结果进行了局部放大处理.



(a)红外图像　(b)可见光图像　(c)DenseFuse[8]　(d)MFEIF[22]　(e)DeF[23]

(f)YDTR[15]　(g) DDFM[24]　(h) LRRNet[12]　(i) CS2Fusion[25]　(j) 本文方法

图 5　不同融合方法在 VOT2020-RGBT[21] 数据集上的定性比较

Fig. 5　Qualitative comparison of different fusion methods on VOT2020-RGBT[21] dataset

对于房屋表面的瓦片特征,DenseFuse[8]、MFEIF[22]、DeF[23] 和 YDTR[15] 的处理使融合结果中的瓦片被模糊化,而 DDFM[24] 和 CS2Fusion[25] 的生成结果亮度偏高,导致瓦片特征出现了一定程度地丢失. 在红外图像中,场景中的树木和背景是难以区分的. 相比于其他方法,本文方法在图像对比度、权衡红外图像和可见光图像的融合比重方面具有明显的优势. 例如,CS2Fusion[25] 的树木茂密区域丢失了可见光图像中原本的特征. DeF[23] 和 YDTR[15] 的树木与背景仍然较难区分.

3.4　消融实验

3.4.1　频率分解模块分析

为证明频率分解模块的有效性,如表 3 中的‘w/o FDM’所示,它表示去除了频率分解模块中的八度卷积,此时该模块退化成只有几个普通卷积组成的简单特征提取块,且不再能对源图像的频率特征进行学习. 对比‘w/o FDM’和本文方法在 6 种评价指标上的结果可以发现,在去除频率分解操作后,网络生成的融合结果在 EN、SD、VIF、AG 和 FMI_w 指标上均降低,其中 EN、SD 和 VIF 分别下降了 0.0991、2.6139 和 0.0423. 添加频率分解操作后,用于评估图像细节丰富度的 SF 指标降低了 0.1392,因为频率分解模块中会使用平均池化操作对源图像进行下采样从而提取低频特征,而下采样一定程度上会影响到图像的细节保留效果,从而导致 SF 指标略微变差. 但结合图 6 第 3 列和第 6 列分析,第 1、2、4 组图像可以看到,添加频率分解模块有助于突出显著性特征,第 3 组图像可以证明,频率域分解模块对源图像的纹理细节信息具有较好地保留效果.

3.4.2　空间特征提取模块分析

为了证明空间粗略融合能改善网络的融合性能,如表 3 的‘w/o Spatial’所示,它表示移除掉在跨域自适应融合模块之前的空间特征提取模块,只使用频率域的特征生成融合结果. 对比‘w/o Spatial’和本文方法在 6 种评价指标上的结果可以发现,在去除空间粗略融合后,网络生成的融合结果在所有指标上均会有所下降. 结合图 6 第 4 列和第 6 列分析可以发现,添加空间粗略融合有助于保护图像的空间结构信息和纹理细节特征.

3.4.3　跨域自适应融合模块的可学习参数分析

如表 3 的‘w/o CDAFM-w’所示,它表示去除掉跨域自适应融合模块的可学习参数 W. 定量对比‘w/o CDAFM-w’和本文方法可以发现,去除可学习参数 W 后,网络生成的融合结果在 6 种评价指标上的结果均会降低,这充分地证明了可学习参数 W 的有效性. 对比图 6 的第 5 列和第 6 列可以发现,去

表 3　本文方法在 TNO[20] 数据集上的消融实验

Table 3　Ablation experiments of the proposed method on the TNO[20] dataset

	EN ↑	SD ↑	SF ↑	VIF ↑	AG ↑	FMI_w ↑
w/o FDM	6.838	38.349	**10.512**	0.747	3.960	0.426
w/o Spatial	6.823	38.632	10.177	0.752	3.922	0.419
w/o CDAFM-w	6.857	39.918	10.120	0.759	3.894	0.400
本文方法	**6.927**	**40.963**	10.373	**0.789**	**3.982**	**0.435**

注:加粗字体表示各列最优结果.

除跨域自适应融合模块的可学习参数会严重影响融合图像的显著性信息的保留能力,从而导致网络的融合性能变差.

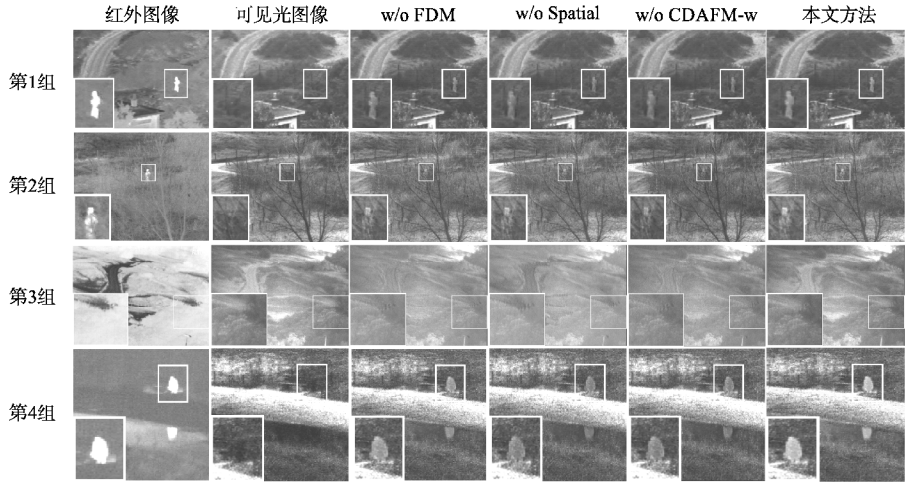


图 6 频率空间协作的红外可见光图像融合网络的消融实验结果

Fig. 6 Ablation experiment results of the frequency-spatial collaboration network for infrared and visible image fusion

4 结 论

鉴于当前多模态图像融合方法普遍未能充分重视频率特征的现状,本文提出了一种频率空间协作的红外可见光图像融合网络(FSCN). 本文设计了一个频率分解模块,旨在深入挖掘源图像在频率域中的丰富信息. 该模块能够灵活地将输入图像分解为高频部分(模态特有特征)和低频部分(模态共有特征). 随后,将多模态图像对应频率的特征分别融合,生成频率域融合结果. 同时,为了确保融合结果具备出色的空间结构,本文巧妙地提取了源图像的空间粗略特征,它可被视为对频率特征的有效补充. 与频率融合特征相互协作,共同提升网络的融合性能. 此外,本文引入了一个跨域自适应融合模块,以整合两个域中的信息,有效应对频率域和空间域之间存在的域间差异挑战. 最后,通过大量实验评估 FSCN 的性能,实验结果表明,本文方法在保持融合图像的显著性特征和纹理细节方面具有优越性.

本文提出的研究思路具备拓展至其他图像融合任务的潜力. 当前的研究主要集中在红外可见光图像融合领域,并通过大量严谨的实验验证了其有效性和可行性. 未来,研究将探索一种通用的图像融合框架,以灵活适应不同领域的图像融合需求,例如医学图像融合. 同时,本文将从多个方面对融合任务的损失函数进行深入研究,以更好地调控网络训练过程,从而生成更为优质的融合结果,提升网络的整体性能与鲁棒性.

References:

[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: 5792-5801.

[2] YU M Y, LIU X Y. Ship detection method based on multimodal visible and infrared image fusion[J/OL]. Computer Engineering, 2025, 1-10, https://doi. org/10. 19678/j. issn. 1000-3428. 0070436, 2025-01-15.

[3] WANG Y T, LIU Z M, WAN Y P, et al. Target detection under low light conditions based on visible and infrared images[J]. Computer Engineering, 2024, 50(8): 270-281.

[4] CHENG Q H, JIAN H F, ZHENG S K, et al. Illumination-aware infrared/visible fusion for object detection[J]. Computer Science, 2025, 52(2): 173-182.

[5] Tang Z, Xu T, Li H, et al. Exploring fusion strategies for accurate RGBT visual object tracking[J]. Information Fusion, 2023, 99: 101-881, doi: 10. 48550/arXiv. 2201. 08673.

[6] 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: 101-828, doi: 10. 1016/j. inffus. 2023. 101828.

[7] LIU D, ZHANG G Y, SHI Y Q, et al. Zero-shot infrared and visible image fusion based on fusion curve[J]. Pattern Recognition and Artificial Intelligence, 2025, 38(3): 268-279.

[8] Li H, Wu X J. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2019, 28(5): 2614-2623.

[9] YANG Y, LIU J X, HUANG S Y, et al. Convolutional auto-encoding fusion network for infrared and visible image fusion[J]. Journal of Chinese Computer Systems, 2019, 40(12): 2673-2680.

[10] QIU D F, HU X Y, LIANG P W, et al. A deep progressive infrared and visible image fusion network[J]. Journal of Image and Graphics, 2023, 28(1): 156-165.

[11] Tang L, Yuan J, Ma J. 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, doi: 10. 1016/j. inffus. 2021. 12. 004.

[12] Li H, Xu T, Wu X, 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.

[13] Ma J, Yu W, Liang P, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. Information Fu-

sion,2019,48:11-26,doi:10.1016/j.inffus.2018.09.004.

[14] Zhou H,Wu W,Zhang Y,et al. Semantic-supervised infrared and visible image fusion via a dual-discriminator generative adversarial network[J]. IEEE Transactions on Multimedia,2021,25:635-648,doi:10.1109/TMM.2021.3129609.

[15] Tang W,He F,Liu Y. YDTR:infrared and visible image fusion via Y-shape dynamic transformer[J]. IEEE Transactions on Multimedia,2022,25:5413-5428,doi:10.1109/TMM.2022.3192661.

[16] SUN X H,GUAN Z,WANG X. Vision transformer for fusing infrared and visible images in groups[J]. Journal of Image and Graphics,2023,28(1):166-178.

[17] Chen Y,Fan H,Xu B,et al. Drop an octave:reducing spatial redundancy in convolutional neural networks with octave convolution[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2019:3434-3443.

[18] Tang L,Yuan J,Zhang H,et al. PIAFusion:a progressive infrared and visible image fusion network based on illumination aware[J]. Information Fusion,2022,83-84:79-92,doi:10.1016/j.inffus.2022.03.007.

[19] Li H,Wu X. CrossFuse:a novel cross attention mechanism based infrared and visible image fusion approach[J]. Information Fusion,2024,103:102147,doi:10.1016/j.inffus.2023.102147.

[20] Alexander T. TNO image fusion dataset[J]. Figshare Dataset,2014,doi:10.6084/m9.figshare.1008029.v2.

[21] Kristan M,Leonardis A,Matas J,et al. The eighth visual object tracking VOT2020 challenge results[C]//Computer Vision-ECCV Workshops,European Conferenceon Computer Vision,2020:1-56.

[22] 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,2022,32(1):105-119.

[23] Liang P,Jiang J,Liu X,et al. Fusion from decomposition:a self-supervised decomposition approach for image fusion[C]//European Conference on Computer Vision,2022,doi:10.1007/978-3-031-19797-0_41.

[24] Zhao Z,Bai H,Zhu Y,et al. DDFM:denoising diffusion model for multi-modality image fusion[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision,2023:8082-8093.

[25] Wang X,Guan Z,Qian W,et al. CS2Fusion:contrastive learning for self-supervised infrared and visible image fusion by estimating feature compensation map[J]. Information Fusion,2024,102:1-15,doi:10.1016/j.inffus.2023.102039.

[26] Roberts J W,Van Aardt J A,Ahmed F B. Assessment of image fusion procedures using entropy,image quality,and multispectral classification[J]. Journal of Applied Remote Sensing,2008,2(1):1-28.

[27] MA Jiayi,MA Yong,LI Chang. Infrared and visible image fusion methods and applications:a survey[J]. Information Fusion,2019,45:153-178,doi:10.1016/j.inffus.2018.02.004.

[28] 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.

[29] Yu H,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.

[30] Cui G,Feng H,Xu Z,et al. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition[J]. Optics Communications,2015,341:199-209,doi:10.1016/j.optcom.2014.12.032.

[31] Haghighat M,Razian M A. Fast-FMI:non-reference image fusion metric[C]//IEEE 8th International Conference on Application of Information and Communication Technologies,2014:1-3.

附中文参考文献:

[2] 于梦源,刘向阳. 基于多模态可见光和红外图像融合的船舶检测方法[J/OL]. 计算机工程,2025,1-10,https://doi.org/10.19678/j.issn.1000-3428.0070436,2025-01-15.

[3] 王昱婷,刘志明,万亚平,等. 基于可见光与红外图像的弱光条件下目标检测[J]. 计算机工程,2024,50(8):270-281.

[4] 程清华,鉴海防,郑帅康,等. 基于光照感知的红外/可见光融合目标检测[J]. 计算机科学,2025,52(2):173-182.

[7] 刘　铎,张国印,史一岐,等. 基于融合曲线的零样本红外与可见光图像融合方法[J]. 模式识别与人工智能,2025,38(3):268-279.

[9] 杨　勇,刘家祥,黄淑英,等. 卷积自编码融合网络的红外可见光图像融合[J]. 小型微型计算机系统,2019,40(12):2673-2680.

[10] 邱德粉,胡星宇,梁鹏伟,等. 红外与可见光图像渐进融合深度网络[J]. 中国图象图形学报,2023,28(1):156-165.

[16] 孙旭辉,官　铮,王　学. 红外与可见光图像分组融合的视觉Transformer[J]. 中国图象图形学报,2023,28(1):166-178.