DOI: 10.13382/j.jemi.B2508413

基于双支路特征提取和语义引导的偏振图像融合网络*

陈广秋 代宇航 段 锦 黄丹丹
(长春理工大学电子信息工程学院 长春 130022)

摘 要:为解决当前偏振图像融合技术着重于融合结果的视觉质量与统计指标,忽略了融合图像在后续高级视觉任务中的应用这一问题,提出一种由语义分割引导的偏振图像双支路特征提取网络结构。融合网络包括编码器、融合层和解码器。在编码器中,构建 2 个由梯度残差密集块 GRDB 和 SwinTransformer 组成的双支路特征提取器,用于提取源图像的局部偏振特征与全局强度信息;在融合层内,采用可逆神经网络 INN 建立两类特征相关性,用以无损增强偏振特征并进行融合;在解码器中,使用 Restormer 作为基本单元,恢复和保留融合特征中的高频特征和场景细节,以提升图像清晰度并获得融合图像。为了使融合结果包含丰富的语义信息,本文在训练阶段将融合网络与分割网络级联,利用语义分割损失引导高级语义信息回流并指导融合网络训练,提高融合图像在高级视觉任务中的应用性能。实验结果表明,提出的融合网络,其融合结果在主观评价和语义分割任务中均取得最优值,并在客观评价指标中信息熵 EN 和结构相似性指数 SSIM 分别比其他融合方法提升了 27%和 16.8%。

关键词: 图像融合;视觉任务;语义引导;偏振图像
中图分类号: TN919.81;TP391.4 **文献标识码**: A **国家标准学科分类代码**: 520.20

Polarization image fusion based on dual-branch feature extraction and semantic guidance

Chen Guangqiu Dai Yuhang Duan Jin Huang Dandan
(School of Electronic Information Engineering, Changchun University of Science and Technology, Changchun 130022, China)

Abstract: To address the current limitation in polarization image fusion technology—where the focus is predominantly on the visual quality and statistical metrics of the fused output while neglecting its applicability to subsequent high-level vision tasks—this paper proposes a dual-branch feature extraction architecture for polarization image fusion, guided by semantic segmentation. The fusion network comprises an encoder, a fusion layer, and a decoder. In the encoder, a dual-branch feature extractor—composed of GRDB and Swin Transformers—is constructed to extract local polarization features and global intensity information from the source images. Within the fusion layer, an INN is employed to model the inter-feature correlations, enabling lossless enhancement and effective fusion of the polarization characteristics. In the decoder, Restormer serves as the core building block to reconstruct and preserve high-frequency details and structural scene information from the fused features, thereby enhancing image clarity and generating the final fused result. To enrich the fused output with task-relevant semantics, the fusion network is cascaded with a segmentation network during training. The semantic segmentation loss is leveraged to guide the backpropagation of high-level semantic information, thereby optimizing the fusion network and improving the utility of the fused images for advanced vision tasks. Experimental results demonstrate that the proposed network achieves superior performance in both subjective visual assessment and downstream semantic segmentation tasks. Moreover, it outperforms existing fusion methods in objective metrics, with notable improvements of 27% in EN and 16.8% in SSIM.

Keywords: image fusion; visual tasks; semantic guidance; polarization images

收稿日期: 2025-05-27 Received Date: 2025-05-27
* 基金项目:国家自然科学基金重大仪器专项项目(62127813)资助

0 引 言

质量和目标识别能力。强度图像反映场景的强度细节和亮度信息,可以准确反映目标物体的外观,但对表面特性和反射信息的描述十分有限。而线偏振度图像则通过捕捉物体表面的反射偏振光信息,揭示表面粗糙度、材质特性等细节,对于透明和反射目标物体具有较强感知力。因此,将强度图像与线偏振度图像融合,能够同时获得亮度信息和物体反射特性,尤其是在路况复杂或恶劣天气环境下,可以增强目标检测和物体辨识精度。

目前,图像融合方法主要分为传统方法与深度学习两大类。主流的传统方法包括基于多尺度分析和基于稀疏表示学习的融合方法。前者有经典的拉普拉斯金字塔(laplacian pyramid, LP),小波变换(wavelet transform, WT)等;后者主要有联合稀疏表示(joint sparse representation, JSR),卷积稀疏表示(convolutional sparse representation, CSR)等。

随着深度学习的不断发展并根据神经网络结构特点,基于深度学习的融合方法主要包括卷积神经网络(convolutional neural network, CNN)、自编码器网络(autoencode, AE)和生成对抗网络(generative adversioal network, GAN)。

2020 年 Zhang 等[1]提出了一种无监督的偏振图像融合网络框架 PFNet,自动提取不同偏振角度图像特征并进行融合;2022 年 Li 等[2]在 PFNet 网络基础上,通过增加损失函数,构建了 TIPFNet 网络,在图像细节和信息保留方面得到提升;2023 年 Wu 等[3]构建了 DBPFNet 偏振图像融合网络,在注意机制基础上利用空间金字塔池提取多尺度关键全局特征,以获得具有显著红外偏振信息的融合图像;2024 年 Liu 等[4]通过在强度图像和线偏振度图像的特征之间建立全局特征交互和融合特征,构建 DT-F Transformer 网络实现互补信息挖掘。

但是现有的图像融合方法仍然倾向于追求更高的图像客观评价指标,忽略了融合图像本身呈现的视觉效果以及后续在语义分割或目标检测等高级视觉任务中的应用表现,导致融合图像依然存在重要特征缺失、纹理模糊及分割结果较差等问题。目前针对视觉任务需求的偏振图像融合算法较少,相应的文献稀缺,但其应用逐渐成为研究热点,如语义分割、目标检测及跟踪等。因此,本文提出一种基于双支路特征提取和语义引导的偏振图像融合网络。在编码器中通过双支路特征提取器,由梯度密集残差块(gradient residual dense block, GRDB)捕捉边缘、纹理等局部偏振细节,利用 SwinTransformer 的移动窗口机制和层级结构来捕获全局强度信息;在融合层内,采用可逆神经网络(invertible neural networks, INN)对呈现互补作用的两类特征进行增强与融合;在解码器内,使用 Restormer 多尺度重构融合特征,逐层恢复融合图像。另外本文还为一偏振数据集建立分割标签,网络训练时将实时语义分割网络 BiSeNet[5]与融合网络级联,利用语义损失优化融合网络,使得融合图像既能包含稳定的强度和偏振信息,同时在后续语义分割视觉任务中具有良好的应用表现。

1 网络整体框架

整个网络框架包含融合网络和语义分割网络,如图 1 所示。融合网络中包括编码器、融合层和解码器,语义分割网络则使用经典的 BiSeNet 作为主体。

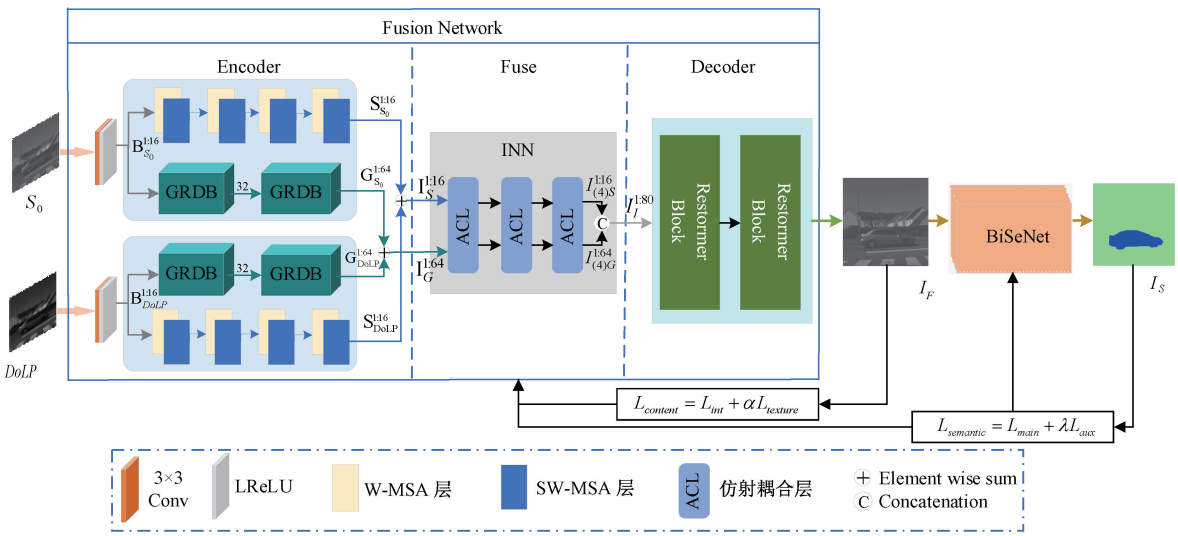


图 1 网络框架

Fig. 1 Network architecture diagram

在训练阶段,将融合网络和语义分割网络级联,编码器对配准后的强度图像 S_0 与线偏振度图像 *DoLP* 进行特征提取,然后在融合层中将特征图进行整合,获得融合特征图 I_1 ,最后在解码器中,对融合特征图进行多尺度重构,得到融合图像 I_F ,随后,BiSeNet 将根据本文制作的偏振语义分割标签对融合图像进行分割并获得分割结果 I_S ,利用语义损失函数引导高级语义信息回流到融合网络,指导融合网络的训练,增强融合图像的高级视觉任务应用属性。

在测试阶段,将 S_0 和 *DoLP* 图像输入到训练优化后的融合网络中获取融合图像。

1.1　融合网络

1) 编码器

编码器主要由双支路特征提取器构成,其中包括梯度密集残差块(gradient residual danse block,GRDB)支路和 SwinTransformer 支路,结构如图 1 所示。强度图像 S_0 与线偏振度图像 *DoLP* 先分别经过一组 3×3 卷积和 LReLU 激活函数,获得 16 通道的基础特征图 $B_X^{1:16}$,其中 $X \in \{S_0, DoLP\}$;然后进入双支路特征提取器,其中 GRDB 支路采用变体残差块结构,GRDB 内部结构如图 2 所示。

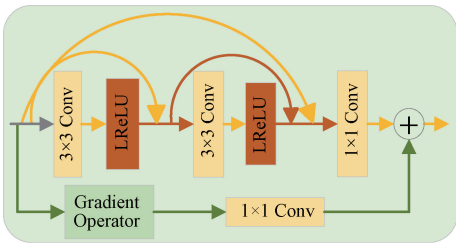


图 2　GRDB 内部结构

Fig. 2　Internal structure of GRDB

主流为密集连接块,由 3×3、1×1 卷积及 LReLU 激活函数构成卷积组合。残差流为梯度提取块,包括 1 个 Sobel 梯度算子和 1×1 卷积层,最后通过逐元素加法将主流和残差流输出特征整合,提取局部偏振特征和颗粒度细节,基础特征图 $B_X^{1:16}$ 经过 GRDB 支路后,输出为 64 通道梯度特征图 $G_X^{1:64}$ 。

SwinTransformer 支路由 patch partition、linear embedding、patch merging 等层和 swinTransformer Blocks 组成,分成 4 个阶段通过层级结构,利用移动窗口机制提取图像特征。

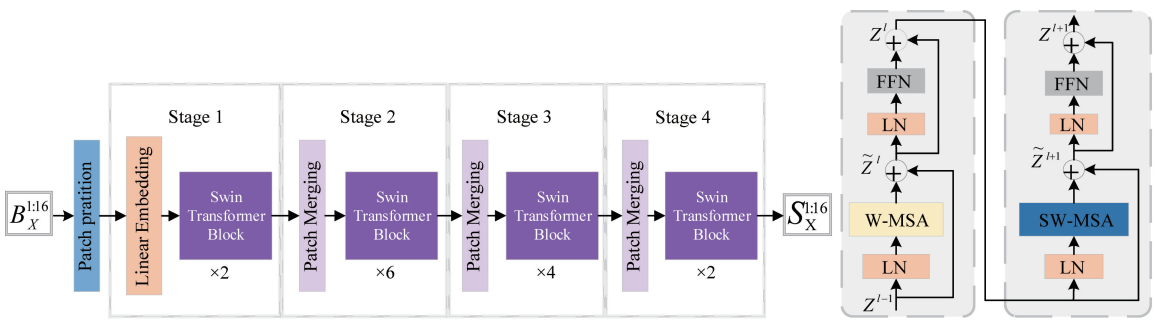
基础特征图 $B_X^{1:16}$ 首先经过 Patch Partition 分割成一系列重叠的小块;随后在第一阶段由线性嵌入(linear embedding)层进行编码,将图像块级像素转换成嵌入向量,进入到 2 个连续 SwinTransformer Blocks 中,如图 3(b)所示,分别包含 W-MSA(窗口自注意力机制)和 SW-MSA(移动窗口自注意力机制),以此构建窗口间的特征相关性。两个连续 SwinTransformer Blocks 输入输出计算过程如式(1)所示。

$$\begin{aligned}
\tilde{Z}^{l} &= \text{W-MSA}(\text{LN}(Z^{l-1})) + Z^{l-1} \\
Z^{l} &= \text{FFN}(\text{LN}(\tilde{Z}^{l})) + \tilde{Z}^{l} \\
\tilde{Z}^{l+1} &= \text{SW-MSA}(\text{LN}(Z^{l})) + Z^{l} \\
Z^{l+1} &= \text{FFN}(\text{LN}(\tilde{Z}^{l+1})) + \tilde{Z}^{l+1}
\end{aligned} \tag{1}$$

式中:W-MSA(·)与 SW-MSA(·)代表窗口自注意力机制和移动窗口自注意力机制处理;LN(·)表示层归一化;⊕表示元素相加;FFN(·)表示前馈网络计算。

阶段 1 的输出送到阶段 2 的块状拼接(Patch Merging)层,将相邻 Patch 合并,缩减分辨率和增加通道数,再送入 6 个连续的 SwinTransformer Blocks 中,阶段 3、4 的 SwinTransformer Blocks 数分别是 4、2 个,阶段流程如图 3(a)所示。

基础特征图 $B_X^{1:16}$ 经过 4 阶段的 W-MSA 和 SW-MSA 交替计算,每阶段负责提取不同层次特征,捕捉从细粒度到粗粒度的全局强度信息,使得 SwinTransformer 支路逐步实现多尺度、跨窗口的特征提取。该支路输出特征表示为 $S_X^{1:16}$。



(a) 阶段流程
(a) Stage flowchart

(b) 2个连续swin transformer blocks
(b) 2 consecutive swin transformer blocks

图 3　Swintransformer 结构

Fig. 3　Swin transformer structure diagram

将 GRDB 支路的输出特征图 $G_{S_0}^{1:64}$ 和 $G_{DoLP}^{1:64}$ 分别与 SwinTransformer 支路输出特征图 $S_{S_0}^{1:16}$ 与 $S_{DoLP}^{1:16}$ 进行逐元素相加,集成局部和全局信息,得到特征图 $I_G^{1:64}$ 和 $I_S^{1:16}$ 输入融合层。

2) 融合层

融合层由 INN[6] 构成,INN 包括 3 个 Real NVP 模型推广的仿射耦合层(affine coupling layer,ACL),层内输入到输出双向映射,对来自编码器的互补强度与偏振特征通过映射函数增强并耦合,ACL 具体结构如图 4 所示。图 4(a)和(b)分别为正向传播与逆向传播过程,图 4(c)是由卷积层构建的映射函数 $MF_1(\cdot)$ 和 $MF_2(\cdot)$。

特征图 $I_G^{1:64}$ 和 $I_S^{1:16}$ 输入到 INN 后,在 3 层 ACL 中通过映射函数 $MF_1(\cdot)$ 和 $MF_2(\cdot)$ 以交替方式耦合,在确保特征信息无损的基础上,整合局部偏振细节和全局强度特征。正向传播计算过程如式(2)所示。

$$\begin{aligned} I_{(i+1)G}^{1:64} &= I_{(i)G}^{1:64} \times \exp(MF_1(I_{(i)S}^{1:16})) + MF_1(I_{(i)S}^{1:16}) \\ I_{(i+1)S}^{1:16} &= I_{(i)S}^{1:16} \times \exp(MF_2(I_{(i+1)G}^{1:64})) + MF_2(I_{(i+1)G}^{1:64}) \end{aligned} \tag{2}$$

式中:i 为 ACL 层数索引,$i=1,2,3$;$I_{(1)G}^{1:64}=I_G^{1:64}$,$I_{(1)S}^{1:16}=I_S^{1:16}$。

逆向传播计算过程如式(3)所示。

$$\begin{aligned} I_{(i)S}^{1:16} &= (I_{(i+1)S}^{1:16} - MF_2(I_{(i+1)G}^{1:64}))/\exp(MF_2(I_{(i+1)G}^{1:64})) \\ I_{(i)G}^{1:64} &= (I_{(i+1)G}^{1:64} - MF_1(I_{(i)S}^{1:16}))/\exp(MF_1(I_{(i)S}^{1:16})) \end{aligned} \tag{3}$$

经过第 3 层 ACL 后,输出的耦合特征图为 $I_{(4)G}^{1:64}$ 与 $I_{(4)S}^{1:16}$,利用通道级联,构成 80 通道的融合特征 $I_I^{1:80}$,并输入到解码器中。

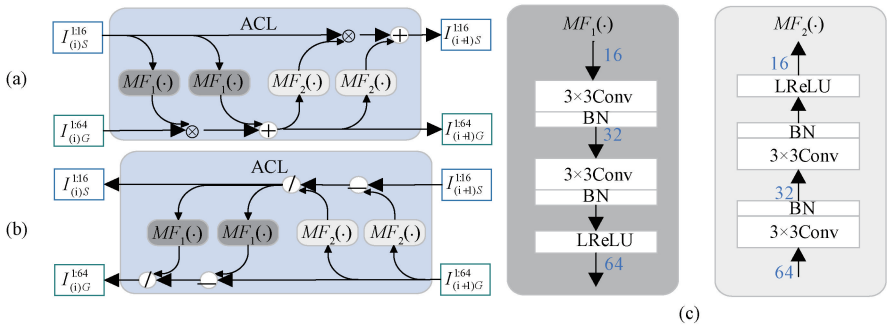


图 4 ACL 具体结构

Fig. 4 The structure of ACL

3) 解码器结构

本文网络的解码器由两个 Restormer 模块串联组成,用以捕捉长距离像素交互,在性能上优于传统 Transformer,具体结构如图 5 所示。

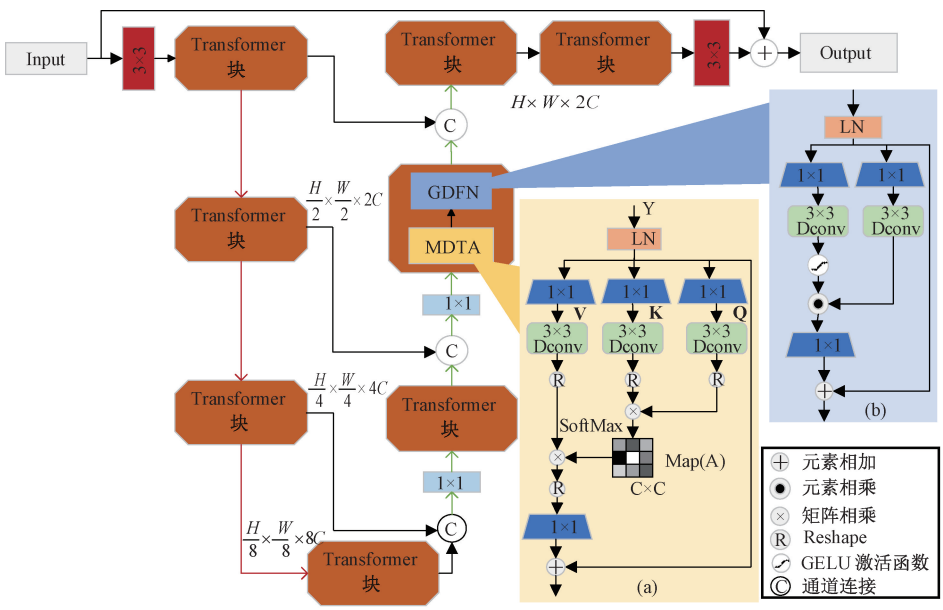


图 5 Restormer 内部结构

Fig. 5 Restormer internal structure

Restormer 采用 Encoder-Decoder 结构,学习高分辨率图像的多尺度特征,每个 Restormer 包含 8 个经过改良的 Transformer 块,块内设有多维卷积注意机制(multi-dconv head transposed attention, MDTA)和动态前馈网络(gated dconv feed forward network, GDFN),可以避免计算复杂度随图像分辨率呈二次增长的问题,提升网络性能。

MDTA 模块通过计算跨通道协方差实现局部和全局像素交互聚合。结构如图 5(a)所示,实现流程如下:首先对来自层归一化张量 $Y \in R^{H\times W\times C}$ 进行 3 个支路的 1×1 和 3×3 深度可分离卷积操作,得到查询(Q)、键(K)和值(V)3 个映射,以增强局部信息;然后对 Q 和 K 分别进行重塑(Reshape)操作,使其尺寸变为 $HW\times C$,再通过矩阵相乘,生成尺寸为 C×C 的转置注意力图 A,与重塑 V 后的特征图做矩阵运算,其结果经过 Reshape 恢复,为 $H\times W\times C$ 特征图;最后通过 1×1 卷积并与原特征逐像素求和,得到最终输出。

GDFN 通过门控机制控制通道中各层级信息流,允许每个级别都专注于其他级别的互补细节,结构如图 5(b)所示,其流程如下:对层归一化的输入特征图进行两个支路的 1×1 和 3×3 深度卷积运算,其中一个支路结果利用 GELU 非线性激活获取权重,与另一支路输出特征图进行点积运算,其结果通过 1×1 卷积层后,再与输入特征图逐像素相加,得到最后输出。

在 Restormer 内部,将多个 Transformer 块排列 4 层获得 4 个尺度,每层包含 1~3 个 Transformer 块,数量从下到上逐步增加且呈左右对称,并通过跳跃连接将左右特征连通。融合特征图 $I_I^{1:80}$ 输入到串联的 Restormer 模块后,得到融合图像 I_F。

1.2　损失函数

为了提高和增强融合图像的视觉质量及语义信息,本文使用联合交互训练策略训练融合网络,并设计了包括内容损失和语义损失在内的联合损失函数 L_{all},定义如式(4)所示。

$$L_{all} = L_{content} + \beta L_{semantic} \tag{4}$$

式中:$L_{content}$ 表示内容损失函数;$L_{semantic}$ 表示语义损失函数。在联合交互训练策略中,β 作为表征语义损失重要性的参数,随着训练次数的增加,β 值逐渐增大。本文将迭代次数设置为 m。在联合损失函数的引导下,利用 Adam 优化器对融合网络中的所有参数进行更新。β 随迭代动态调整,表示为:

$$\beta = \gamma \times (m - 1) \tag{5}$$

式中:γ 作为语义损失和内容损失的平衡参数,本文根据经验设为 0.1。通过联合训练融合与分割网络,随着迭代次数 m 的增加,语义损失将更准确地优化融合网络参数。

1) 内容损失函数

为了使融合网络保留更多的强度和偏振信息,并提升图像的主客观评价,设计了由强度损失和纹理损失组成的内容损失函数 $L_{content}$,如式(6)~(8)所示。

$$L_{content} = L_{\text{int}} + \alpha L_{texture} \tag{6}$$

$$L_{\text{int}} = \frac{1}{HW}\| I_F - max(S_0, DoLP) \|_1 \tag{7}$$

$$L_{texture} = \frac{1}{HW}\| \mid \nabla I_F \mid - max(\mid \nabla S_0 \mid, \mid \nabla DoLP \mid) \|_1 \tag{8}$$

式中:L_{int} 为强度损失函数,用于衡量融合结果和源图像像素间的强度差异;$L_{texture}$ 为纹理损失函数,用于衡量融合结果保留源图像细节和偏振特征的程度;α 是平衡系数,本文设为 10;H 和 W 分别为图像的高和宽;$\| \cdot \|_1$ 表示矩阵的 1-范数;$max(\cdot)$ 表示逐元素最大操作;I_F 为融合图像;S_0 和 $DoLP$ 代表强度图像和线偏振度图像;∇ 为 Sobel 梯度算子;$|\cdot|$ 为取绝对值运算。式(7)和(8)中,通过最大选择策略和范数运算调节 I_F 与源图像的像素亮度一致性,式(8)中,引入 Sobel 梯度算子整合 I_F 和源图像的纹理信息,以保留图像中精细的偏振信息和纹理结构。

2) 内容损失函数

为了增强融合图像的语义信息,训练时引入实时语义分割网络 BiSeNet,利用语义损失函数指导融合网络训练,该损失函数由主语义损失函数和辅助语义损失函数组成,如式(9)所示。

$$L_{semantic} = L_{main} + \lambda L_{aux} \tag{9}$$

式中:$L_{semantic}$ 代表语义损失;L_{main} 为主语义损失函数;L_{aux} 为辅助语义损失函数;λ 为平衡系数,本文设定为 0.1。$L_{semantic}$ 用于优化每个像素的类别预测,L_{aux} 作为主语义损失函数的补充,着重对局部边界的精准分割,如式(10)、(11)所示。

$$L_{main} = \frac{-1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} L_{S0}^{(h,w,c)} \log(I_S^{(h,w,c)}) \tag{10}$$

$$L_{aux} = \frac{-1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C} L_{S0}^{(h,w,c)} \log(I_{sa}^{(h,w,c)}) \tag{11}$$

式中:I_S 是分割结果;I_{sa} 是辅助分割结果;L_{S0} 是由分割标签 L_S 变换而成的二进制向量;H、W、C 分别为图像的高、宽和通道数;h 代表高度方向的行索引;w 代表宽度方向的列索引;c 代表通道方向的索引。

2　实验与分析

2.1　实验设置

训练前,本文为偏振数据集 Rachel Blin[7] 制作了一批语义分割标签,训练时使用 Adam 优化器更新各模块

参数,训练数据集及具体参数设置如表 1 所示。

表 1 训练数据集及具体参数设置

Table 1 Training dataset parameter settings

Item	Fusion(Segement) network
Dataset	70% of Rachel Blin
Image size	640×480
Batch size	1
Epoch	100
Learning rate	0.001(0.01)
Weight decay rate	0.000 2(0.000 5)
Environment	PyTorch1.11.0
Server	AMD EPYC 7T83 CPU、NVIDIA Tesla L40 GPU

在测试阶段,将 Rachel Blin 数据集的 30%和数据集 ZJU-RGB-P[8] 作为测试集。

2.2 融合结果分析

为了验证本文提出的融合网络的有效性,在车辆、建筑物和行人 3 类场景下,将本文融合网络与基于曲波变换(curvelet transform,CVT)[9]、梯度转移融合(gradient transfer-factor,GTF)[10]、MMIF CDD[6]、PAPIF(perceptual-aware patch-based image fusion)[11]、LRRNet[12]、PIAFusion[13]、PF[1] 7 种融合方法进行实验对比,其中 CVT 和 GTF 为典型的传统融合方法,其余为深度学习融合方法。采用信息熵(EN)[14]、标准差(standard deviation,SD)[15]、差异相关性(sum of correlation differnece,SCD)[16]、视觉信息保真度(visual information fidelity,VIF)[17]、融合质量评价因子 Qabf、结构相似度(structural similarity index measure,SSIM)[18] 作为客观评价指标。每种评级指标数值越大,代表融合效果越好,融合方法越优越。

源图像和融合结果如图 6~8 所示,客观评价指标数据如表 2~4 所示。

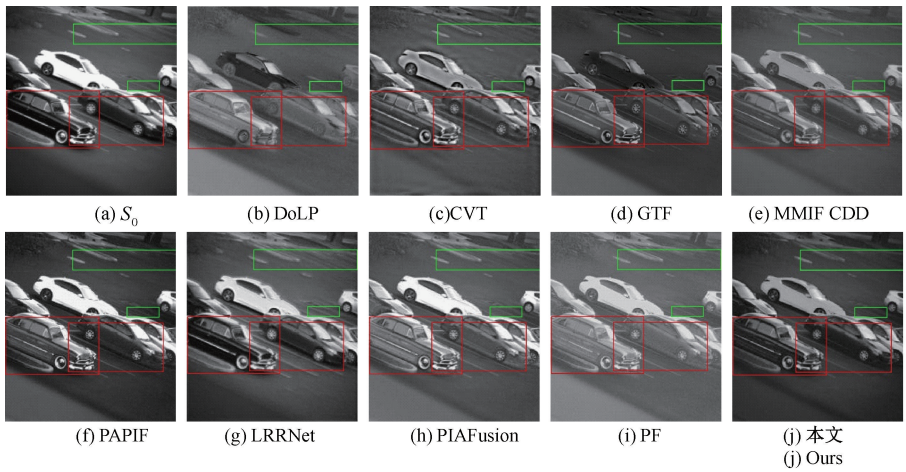


图 6 车辆场景对比实验结果

Fig. 6 Vehicle scene comparison experiment results

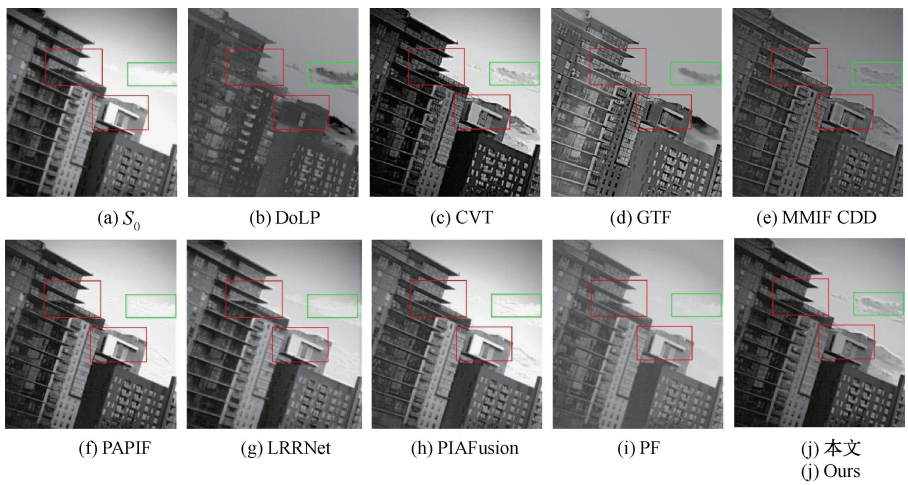


图 7 建筑物场景对比实验结果

Fig. 7 Experimental results of building scene comparison

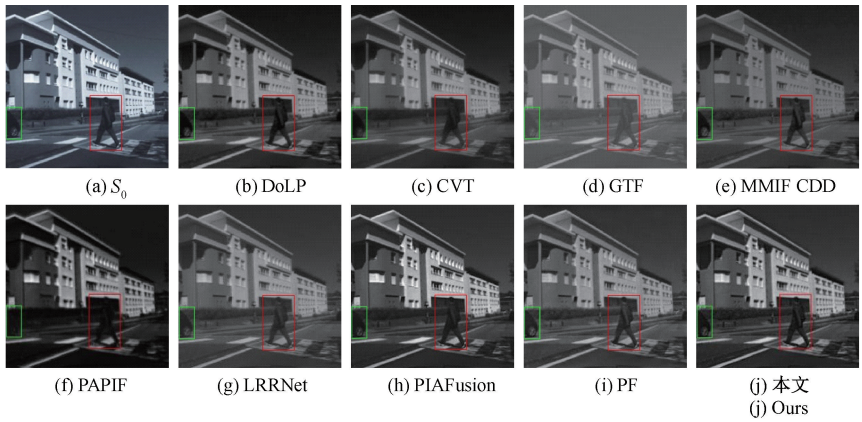


图 8　行人场景对比实验结果

Fig. 8　Pedestrian scene comparison experiment results

图 6 所示为车辆场景下的源图像(出自数据集 ZJU-RGB-P)和融合结果,CVT 和 GTF 的融合结果中,车辆表面出现了强度特征不稳定的现象,不能充分保留亮度的细微变化;MMIF CDD 和 PAPIF 的融合结果图像整体偏暗,无法分辨出车辆的外部细节;LRRNet 融合结果中车身周围存在细微伪影,同时不能正确区分车道线等偏振细节;PIAFusion 和 PF 融合结果中虽然将偏振特征体现得较好,但是强度信息损失严重,导致整体车身表面亮度失衡,并且在 2 种方法的融合结果左上方区域均出现了少许噪音;本文融合方法得到的融合结果在不损失强度信息的基础上整合了关键偏振特征,丰富了偏振图像的偏振细节。对比其他方法结果图像中红框标注的车辆区域,可以看到本文融合结果中汽车的外部材质和车体细节被充分保留,在绿框标注的区域内,周围的“花坛护栏”和“车道线”的边缘细节也更为自然和连贯。客观评价结果如表 2 所示。本文的融合网络在 EN、SD、SCD 这 3 项指标上提升幅度最多,分别为 4.9%、0.9%和 2.3%。而相较于其他方法,本文方法在 VIF、Qabf 和 SSIM 上也均取得了最优值。

表 2　车辆场景下不同融合方法融合结果的客观评价结果

Table 2　Objective evaluation results of fusion results of different fusion methods in vehicle scenario

Method	EN	SD	SCD	VIF	Qabf	SSIM
CVT	7.042	37.437	0.997	0.506	0.503	0.397
GTF	6.729	37.902	1.205	0.593	0.416	0.407
MMIF CDD	6.982	43.860	1.274	0.693	0.521	0.437
PAPIF	7.214	48.309	1.679	0.659	0.531	0.426
LRRNet	7.094	49.795	1.506	0.699	0.509	0.418
PIAFusion	7.042	50.965	1.504	0.729	0.532	0.433
PF	7.157	52.094	1.589	0.768	0.529	0.437
Ours	***7.410***	***54.972***	***1.701***	***0.769***	***0.553***	***0.476***

注:粗斜体为最优值。下同

图 7 所示为建筑物场景下源图像(出自数据集 ZJU-RGB-P)和融合结果,CVT 融合结果中建筑物与周围环境对比度不均衡,且部分区域过暗;GTF 融合结果中,楼的中心区域较为模糊,“楼”边缘细节失真;MMIF CDD 和 PAPIF 融合结果图像在亮度方面保留了 *DoLP* 图像的部分偏振信息,但局部区域丢失了强度信息,降低了整体图像的亮度及清晰度,如图 7 中红框标注的光亮区域;LRRNet 的融合结果图像中保留了较多的偏振细节信息,但仍存在一些建筑细节表达不清等问题,如图 7 中红框内楼宇的“窗户”和“护栏”周围;PIAFusion 融合结果中,建筑物的边缘细节与背景过度不清,丢失少量边缘信息且缺少层次感。PF 融合结果中,设计简约的建筑物中强度图像特征体现较为明显,然而“楼”的外观偏振细节缺失;本文融合结果中,“窗户”等建筑细节中的偏振信息得到了充分表达,同时其周围的“阳台护栏”、“云彩”等亮度信息表达清晰,本文融合方法能够有效地合并源图像的偏振互补信息,在增强亮度等重要特征的同时,也没有削弱建筑物周边微小细节。客观评价结果如表 3 所示,可以看出,本文方法 6 项指标均取得了最优值。

表 3 建筑物场景下不同融合方法融合结果的客观评价结果

Table 3 Objective evaluation results of fusion results of different fusion methods in building scene

Method	EN	SD	SCD	VIF	Qabf	SSIM
CVT	7.014	38.562	0.989	0.486	0.512	0.402
GTF	6.627	38.879	1.351	0.569	0.421	0.413
MMIF CDD	6.783	43.860	1.289	0.702	0.519	0.472
PAPIF	7.251	47.893	1.682	0.663	0.529	0.432
LRRNet	7.190	48.905	1.507	0.685	0.515	0.421
PIAFusion	7.139	51.958	1.506	0.718	0.527	0.419
PF	7.139	52.071	1.590	0.771	0.537	0.441
Ours	***7.441***	***52.973***	***1.811***	***0.774***	***0.531***	***0.479***

图 8 所示为行人场景下源图像(出自数据集 Rachel Blin)和融合结果,CVT 的融合结果背景昏暗,且行人与周围环境的强度对比过弱。GTF 结果中背景与目标区分不够锐利,边缘细节模糊。LRRNet 和 MMIF CDD 的融合结果丢失了大部分强度信息,导致场景中行人姿态及车辆外观等详细信息在图像中表征不清晰;PAPIF 的融合结果在图像的过渡区域出现模糊现象,并且行人与周边环境区分不明显;PF 在内容较少的场景中视觉效果较好,但是在行人、建筑物与车辆都出现的复杂场景下,目标与背景的区分不明显,同时偏振纹理等不够凸显;仅有 PIAFusion 与本文结果相近,但与之相比,本文融合结果强度信息保持良好,行人、车辆的偏振细节刻画精准,亮度与对比度适中,“行人”的服装细节与“车辆”的外观纹理在不同光照条件下均能清晰可见。客观评价结果如表 4 所示。本文方法在图 8 的 6 项客观评价指标中均取得了最高分,尤其在 EN、SD、SSIM 这 3 项中分别提升了 4.7%、2.9%和 3.1%。

表 4 行人场景下不同融合方法融合结果的客观评价结果

Table 4 Objective evaluation results of fusion results of different fusion methods in pedestrian scene

Method	EN	SD	SCD	VIF	Qabf	SSIM
CVT	7.142	39.369	1.089	0.513	0.509	0.397
GTF	6.615	39.809	1.459	0.557	0.439	0.420
MMIF CDD	6.794	43.995	1.253	0.718	0.528	0.412
PAPIF	7.267	48.997	1.705	0.613	0.549	0.424
LRRNet	7.209	50.006	1.239	0.694	0.528	0.419
PIAFusion	7.139	51.963	1.579	0.723	0.539	0.420
PF	7.249	53.051	1.574	0.754	0.510	0.449
Ours	***7.251***	***54.863***	***1.803***	***0.758***	***0.549***	***0.475***

为了验证本文融合方法的稳定性,从 Rachel Blin 数据集中选取 15 对图像进行融合效果对比,6 种评价指标数据折线图如图 9 所示。相比较于其他方法,本文的融合方法得到融合结果图像在 6 种评价指标上均为最优,尤其在 SD、SCD、SSIM 这 3 项上,本文的方法优势明显。

2.3 分割结果分析

为了验证本文融合方法在高级视觉任务应用中的优势,采用经典分割网络 Unet 对不同融合方法的融合结果进行语义分割实验,分割对象为车辆(Car)和行人(Person)结果如图 10 所示。

从图 10 可以看出,当场景中出现多个车辆时,本文融合方法能够准确的将重叠出现的车辆分割出来,体现出车辆的具体轮廓和外观细节;当场景中同时出现行人与车辆时,本文的融合方法能够精准地分割行人,同时将远处的车辆进行细微的区分,行人的衣着和发型边缘分割准确。这得益于本文的融合方法在融合过程中由语义信息引导融合过程,从复杂的道路场景中提取多层次语义特征,提升分割网络 Unet 对偏振融合场景的理解能力,以获得更精准的分割结果。在简单的车道场景中,本文方法能够正确分割。在场景复杂度较高的情况下,本文所提出的融合方法及其优势得到了充分体现,分割效果获得显著提升。

采用准确率(Acc)[19]和交并比(IoU)[20]来衡量分割网络对融合图像的分割性能。Acc 表示分割结果中预测正确的像素数占总像素数的比例,平均准确率(mAcc)是对每个类别准确率的平均值,避免分类不平均对整体准确率的影响;IoU 用于衡量分割结果与 Label 之间的重叠程度,采用预测区域和真实情况交集与并集的比值来表示,平均交并比(mIoU)是对所有类别的交并比取平均值,以评估多类别分割任务中模型的整体性能,特别是对

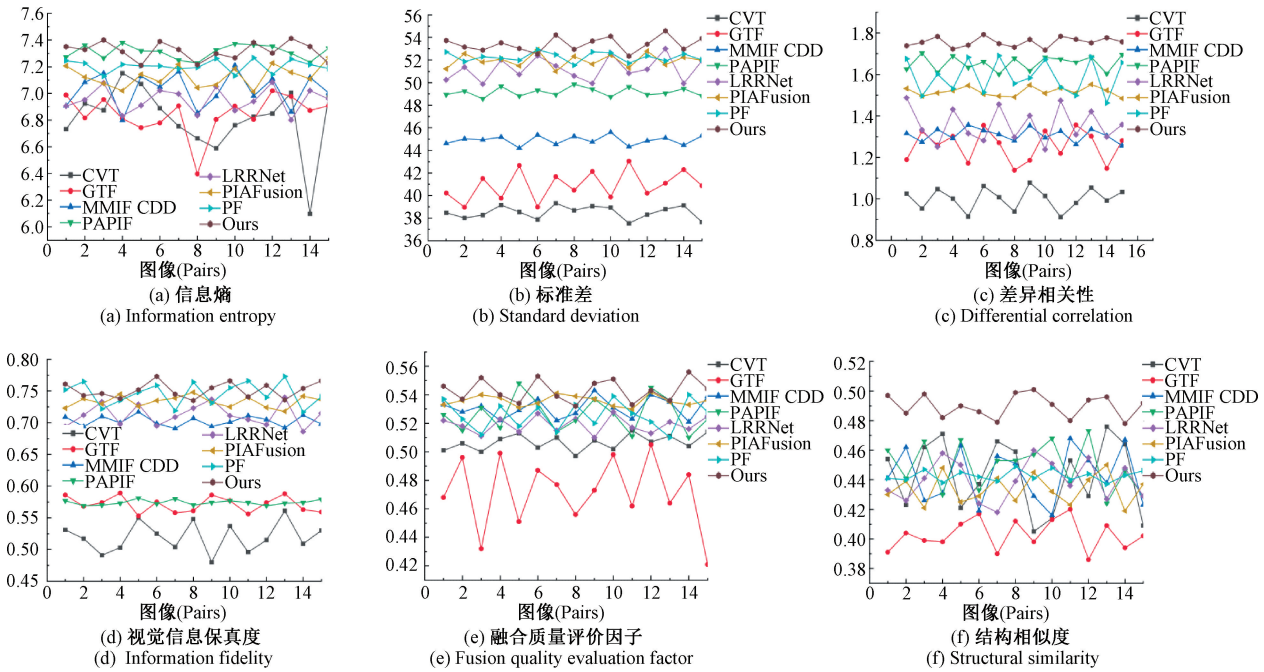


图 9　评价指标点线图

Fig. 9　Evaluation index point chart

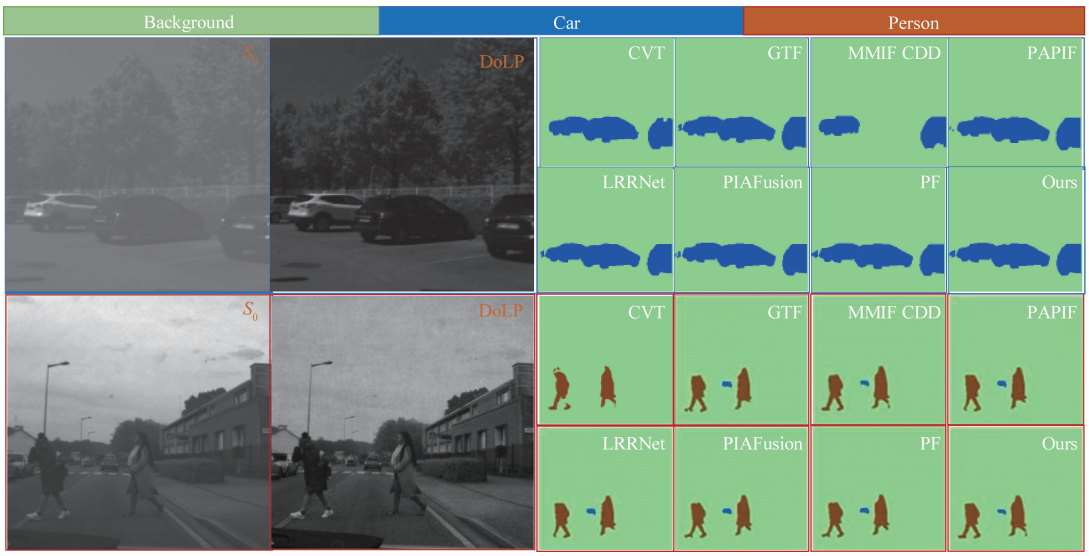


图 10　分割结果

Fig. 10　Segmentation result

于语义分割任务。不同方法在 Rachel Blin 数据集中分割结果的客观评价结果如表 5 所示。本文的方法在各类别分割对象中均获得了最优值。

2.4　消融实验

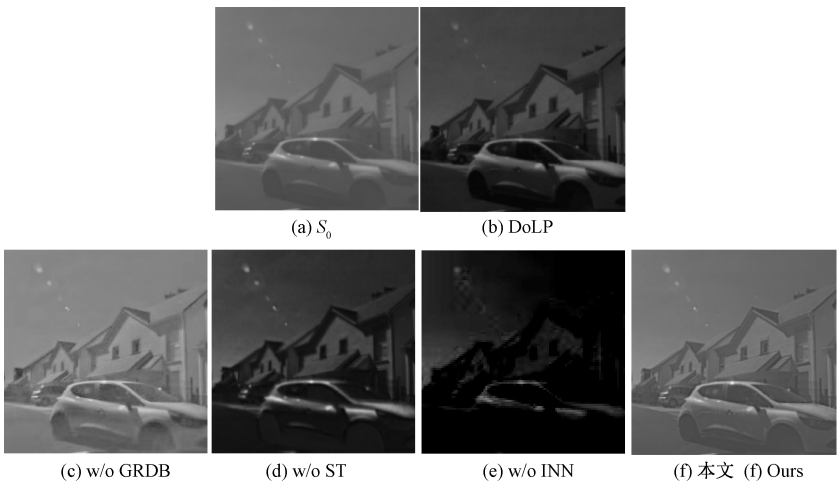
为了验证本文融合网络具有高效的特征信息和语义信息的整合能力，本文设计了网络结构消融实验。设计了 4 组消融对比实验：1）无 GRDB 特征提取支路，记为 w/o GRDB；2）无 SwinTransformer 特征提取支路，记为 w/o ST；3）无 INN 融合层，记为 w/o INN；4）本文网络结构。源图像和融合结果如图 11 所示。客观指标如表 6 所示。

从图 11 可以看出，w/o GRDB 融合结果体现了较多的偏振特征，但整体上缺乏强度细节信息；w/o ST 融合结果体现了较多的强度信息，而缺少了重要偏振特征的表达；w/o INN 融合结果缺少强度和偏振特征，丢失了较多能量信息，图像整体较为模糊，亮度和偏振细节的分布

表 5 不同方法在 Rachel Blin 数据集中分割结果的客观评价结果

Table 5 Objective evaluation results of segmentation results by different methods in the Rachel Blin dataset

Method	Background		Person		Car		mAcc	mIoU
	Acc	IoU	Acc	IoU	Acc	IoU		
CVT	71.72	67.77	73.12	45.99	74.56	71.47	72.80	64.08
GTF	69.58	67.82	72.45	49.47	72.21	72.91	70.08	66.07
MMIF CDD	73.45	72.04	71.89	62.34	73.38	74.23	70.89	72.43
PAPIF	73.31	73.61	73.77	53.58	71.97	73.56	72.35	69.25
LRRNet	71.86	71.23	72.33	65.29	74.72	70.74	71.96	71.44
PIAFusion	72.14	70.39	74.01	57.81	72.67	74.68	71.98	69.96
PF	74.67	71.56	74.56	59.64	75.13	75.35	72.45	70.18
Ours	***78.97***	***75.07***	***77.05***	***68.12***	***79.92***	***78.12***	***78.61***	***75.54***



(a) S_0 (b) DoLP

(c) w/o GRDB (d) w/o ST (e) w/o INN (f) 本文 (f) Ours

图 11 模型结构消融实验结果

Fig. 11 Experimental results of model structure ablation

均较差;本文网络融合结果保留了丰富的偏振细节,纹理和边缘等微小特征,过渡自然,建筑物和车辆的线条边界较为锐利。这一切都得益于本文网络在提取特征阶段充分利用 SwinTransformer 和 GRDB 双支路提取局部偏振细节和全局强度信息。从表 6 中的数据可以看出,本文网络结构得到的融合结果客观评价最优,与主观评价一致。

表 6 模型结构消融实验客观指标

Table 6 Objective index of model structure ablation experiment

Method	EN	SD	SCD	VIF	Qabf	SSIM
w/o GRDB	5.194	45.759	0.632	0.451	0.347	0.228
w/o ST	4.935	47.901	0.759	0.419	0.325	0.274
w/o INN	6.464	51.685	1.239	0.529	0.421	0.319
Ours	***7.197***	***54.779***	***1.795***	***0.741***	***0.547***	***0.471***

3 结 论

本文提出了一种基于双支路特征提取和语义引导的偏振图像融合网络,用于融合强度图像和线偏振度图像。图像首先经过由 GRDB 和 SwinTransformer 构成的双支路特征提取器,充分提取局部偏振特征和全局强度信息,随后利用可逆神经网络将互补特征相互增强并整合出融合特征,最后利用串联 Restormer 模块,将融合特征进行多尺度重构。在训练阶段,将融合网络与分割网络级联,使得语义损失指导融合网络优化参数。与已有的融合方法比较,本文方法仍存在细微的偏振信息失真现象,因此在今后的工作中,本团队将继续探索偏振图像领域,并着重开发特征提取及融合关键步骤,使其更好地应用于高级视觉任务中。

参考文献

[1] ZHANG J C, SHAO J B, CHEN J L, et al. PFNet: An unsupervised deep network for polarization image fusion [J]. Optics Letters, 2020, 45(6):1507-1510.

[2] LI K Y, QI M B, ZHUANG S, et al. TIPFNet: A transformer-based infrared polarization image fusion network [J]. Optics Letters, 2022, 47(16):

4255-4258.

[3] WU Y N, CHANG J, MA N, et al. DBPFNet: A dual-band polarization image fusion network based on the attention mechanism and atrous spatial pyramid pooling[J]. Optics Letters, 2023, 48(19): 5125-5128.

[4] LIU J Y, LI S T, DIAN R W, et al. DT-F Transformer: Dual transpose fusion transformer for polarization image fusion[J]. Information Fusion, 2024, DOI:10.1016/j.inffus.2024.102274.

[5] PENG C L, TIAN T, CHEN C, et al. Bilateral attention decoder: A lightweight decoder for real-time semantic segmentation[J]. Neural Netw, 2021, 137(5): 188-199.

[6] ZHAO Z X, BAI H W, ZHANG J S, et al. Cddfuse: Corre-lation-driven dual-branch feature decomposition for multi-modality image fusion [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023: 5906-5916.

[7] BLIN R, AINOUZ S, CANU S, et al. A new multimodal RGB and polarimetric image dataset for road scenes analysis[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE, 2020. DOI:10.1109/CVPRW50498.2020.00116.

[8] WANG K, XIANG K, YANG K. Polarization-driven semantic segmentation via efficient attention-bridged fusion[J]. Optics Express, 2021, DOI:10.1364/OE.416130.

[9] HUANG S Q, HUANG W Z, ZHANG T, et al. A statistical and Wiener filtering algorithm based on the curvelet transform for SAR images [J]. International Journal of Remote Sensing, 2016, 37(23): 5581-5604.

[10] MA J Y, CHEN C, LI C, et al. Infrared and visible image fusion via gradient transfer and total variation minimization [J]. Information Fusion, 2016, 31: 100-109.

[11] XU H, SUN Y CH, MEI X G, et al. Attention-guided polarization image fusion using salient information distribution[J]. IEEE Transactions on Computational Imaging, 2022, 8:1117-1130.

[12] LI H, XU T Y, WU X J, et al. LRRNet: A novel representation learning guided fusion framework for infrared and visible images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2023, 45(9): 11040-11052.

[13] TANG L F, YUAN J T, 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.

[14] 陈广秋, 代宇航, 段锦, 等. 用于红外与微光图像融合的目标差分注意力和 Transformer 算法[J]. 电子测量与仪器学报, 2025, 39(5): 103-116.
CHEN G Q, DAI Y H, DUAN J, et al. Target differential attention and transformer algorithm for infrared and low-light image fusion [J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(5): 103-116.

[15] ZHOU H B, WU W, ZHANG Y D, et al. Semantic-Supervised infrared and visible image fusion via a dual-discriminator generative adversarial network [J]. IEEE Transactions on Multimedia, 2023, 25: 14.

[16] MA J Y, MA Y, LI C. Infrared and visible image fusion methods and applications: A survey [J]. Information Fusion, 2019, DOI: 10.1016/j.inffus.2018.02.004.

[17] KUO T Y, SU P C, TSAI C M. Improved visual information fidelity based on sensitivity characteristics of digital images[J]. Journal of Visual Communication and Image Representation, 2016, 40: 76-84.

[18] REN L, PAN ZH B, CAO J ZH, et al. Infrared and visible image fusion based on variational auto-encoder and infrared feature compensation [J]. Infrared Physics & Technology, 2021, 117: 103839.

[19] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.

[20] 胡冠华, 张永雷, 申立群. 基于改进 U-Net 的轻量级输电线分割算法[J]. 电子测量与仪器学报, 2024, 38(1): 211-218.
HU G H, ZHANG Y L, SHEN L Q. Lightweight transmission line conductor segmentation algorithm with improved U-Net[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(1): 211-218.

作者简介

陈广秋, 1999 年于吉林大学获得学士学位, 2006 年于吉林大学获得硕士学位, 2015 年于吉林大学获得博士学位, 现为长春理工大学副教授, 主要研究方向为图像处理与机器视觉。
E-mail: gaungqiu_chen@126.com

Chen Guangqiu received his B. Sc. degree from Jilin University in 1999, M. Sc. degree from Jilin University in 2006 and Ph. D. degree from Jilin University in 2015, respectively. Now he is an associate professor in Changchun University of Science and Technology. His main research interests include image processing and machine vision.

代宇航,2022 年于吉林建筑大学获得学士学位,现为长春理工大学硕士研究生,主要研究方向为图像处理与机器视觉。
E-mail:2909119265@qq.com

Dai Yuhang received her B. Sc. degree from Jilin Jianzhu University in 2022. Now she is a M. Sc. candidate at Changchun University of Science and Technology. Her main research interests include image processing and machine vision.

段锦(通信作者),1993 年于北京理工大学获得学士学位,1998 年于沈阳工业学院获得硕士学位,2004 年于吉林大学获得博士学位,现为长春理工大学教授,主要研究方向为偏振成像探测、图像处理与模式识别、数字光学环境仿真。
E-mail:duanjin@vip.sina.com

Duan Jin (Corresponding author) received his B. Sc. degree from Beijing Institute of Technology in 1993, M. Sc. degree from Shenyang Institute of Technology in 1998 and Ph. D. degree from Jilin University in 2004, respectively. Now he is a professor in Changchun University of Science and Technology. His main research interests include polarization imaging detection, image processing and pattern recognition, digital optical environment simulation.

黄丹丹,2007 年于长春理工大学获得学士学位,2009 年于东北大学获得硕士学位,2014 年于大连理工大学获得博士学位,现为长春理工大学讲师,主要研究方向为计算机视觉和机器学习。
E-mail:hdd@cust.edu.cn

Huang Dandan received her B. Sc. degree from Changchun University of Science and Technology in 2007, M. Sc. degree from Northeastern University in 2009 and Ph. D. degree from Dalian University of Technology in 2014, respectively. Now she is a lecturer in Changchun University of Science and Technology. Her main research interests include computer vision and machine learning.