

尹子豪, 刘立群. M-DenseFuse: 一种结合多头注意力机制的异源图像融合模型[J]. 智能计算机与应用, 2025, 15(11): 15-21. DOI:10.20169/j.issn.2095-2163.251103

M-DenseFuse: 一种结合多头注意力机制的异源图像融合模型

尹子豪, 刘立群

(甘肃农业大学 信息科学技术学院, 兰州 730070)

摘要: 为了应对果园图像融合中光照变化与遮挡等复杂环境带来的挑战, 本文提出了一种改进的异源图像融合模型 (M-DenseFuse)。该模型以 DenseFuse 为基础, 引入了 U-Net 网络结构以增强特征提取能力, 同时融合了 M-SENet 通道注意力机制, 并结合 Transformer 中的多头注意力策略, 对通道权重关系进行更精细建模, 从而提升特征表示的准确性与融合图像的质量, 并验证了模型在不同条件下的适应性与优越性能。实验结果表明, 本文所提出模型在图像清晰度、细节保留和融合效果等方面优势明显, 具备良好的实用价值。

关键词: 图像融合; DenseFuse; Transformer; SENet

中图分类号: TP391.41; S225

文献标志码: A

文章编号: 2095-2163(2025)11-0015-07

M-DenseFusion: A heterogeneous image fusion model combining multi-head attention mechanism

YIN Zihao, LIU Liqun

(College of Information Science and Technology, Gansu Agricultural University, Lanzhou 730070, China)

Abstract: To address the challenges posed by complex environments such as illumination changes and occlusions in orchard image fusion, this paper proposes an improved heterogeneous image fusion model (M-DenseFuse). Based on DenseFuse, this model introduces the U-Net network structure to enhance feature extraction capability, and integrates the M-SENet channel attention mechanism, which incorporates the multi-head attention strategy from Transformer to more precisely model the channel weight relationships. This thereby improves the accuracy of feature representation and the quality of fused images, and verifies the model's adaptability and superior performance under different conditions. Experimental results demonstrate that the proposed model has significant advantages in terms of image clarity, detail preservation, and fusion effect, and possesses good practical value.

Key words: image fusion; DenseFuse; Transformer; Squeeze-and-Excitation Networks

0 引言

国内自动化果蔬采摘设备正处于起步阶段, 目前真实的种植环境中光照和水果及叶片间遮挡对苹果的分割、识别、定位、分级带来极大的挑战。

目前, 国内外学者在图像融合领域进行了广泛的研究^[1]。图像融合方法分为 4 种: 基于多尺度分解的方法、基于稀疏表示的方法、基于子空间的方法和基于深度学习的方法。

1 方法

1.1 多尺度分解方法

多尺度分解图像融合方法可以将输入图像分解为多个尺度的子图像, 然后对这些子图像进行融合, 最终得到融合后的图像。在多尺度分解中, 不同尺度的子图像包含了不同的图像细节信息, 对这些子图像进行融合, 可以得到更加全面、准确的图像信息。基于多尺度分解方法的图像融合可以应用于许多场景, 如医学影像、遥感图像、安防监控等, 以提高

基金项目: 甘肃省高校教师创新基金项目 (2023A-051); 甘肃农业大学青年导师基金资助项目 (GAU-QDFC-2020-08); 甘肃省科技计划资助 (20JR5RA032)。

作者简介: 尹子豪 (1999—), 男, 硕士研究生, 主要研究方向: 深度学习与图像处理。

通信作者: 刘立群 (1982—), 女, 硕士, 教授, 硕士生导师, 主要研究方向: 智能计算, 深度学习。Email: llqhjy@126.com。

收稿日期: 2024-02-29

图像的质量和信量,从而更好地满足实际应用需求。多尺度分解方法主要有2种:拉普拉斯金字塔多尺度分解方法以及离散小波变换方法。

(1) 拉普拉斯金字塔多尺度分解的图像融合方法

首先,对原始图像进行高斯金字塔分解,得到多个尺度的子图像;针对每个尺度的子图像,进行拉普拉斯金字塔分解,分离高频和低频信息;对高频子图像进行加权平均处理,以生成融合后的高频子图;最后,将合并后的高频子图像与低频子图像重新组合,生成最终融合的图像。

该算法可以在不同尺度上有效的保存图像中的丰富细节,但可能会出现过度平滑、计算量较大、对噪声敏感、需要合理选择融合权重。

(2) 离散小波变换方法

将多幅图像分别进行离散小波变换,得到多个尺度的小波系数;对两幅图像的小波系数执行加权平均,从而获得融合后的小波系数;通过逆变换这些系数,生成最终融合图像。

通过离散小波变换,可以在不同尺度上实现图像融合,从而实现对图像的局部分分析和全局分析,保留原始图像的细节信息和整体特征,得到更加准确、自然的融合图像。

1.2 基于稀疏表示的方法

首先,通过大量自然图像样本训练,构建一个完备字典,该字典可视为包含丰富图像特征的基础集合,用于有效表示图像的局部结构信息;其次,将待融合图像划分为多个重叠的图像块(patch),在同一完备字典中对每个图像块进行稀疏分解,提取其稀疏表示系数;最后,通过融合不同模态图像的稀疏系数,并结合原字典进行重建,最终生成融合图像。

1.3 基于子空间方法

基于子空间的图像融合方法是一种将图像分解到不同特征子空间中进行处理的融合策略。文献[2]提出了一种基于潜在低秩表示的红外与可见光图像融合方法,首先将红外图像和可见光图像分别分解为低秩子空间与稀疏子空间,用以表示图像中的共性结构和个性细节;其次,对不同模态的低秩子空间进行加权平均,得到融合后的低秩成分,再将稀疏子空间同样加权融合;最后,将低秩和稀疏融合结果进行线性组合,生成最终融合图像。

1.4 基于深度学习的方法

深度学习因其强大的特征提取能力和端到端建模特性,在图像融合领域展现出显著优势,其主要优

点包括可自动学习图像特征、简化融合流程以及良好的数据适应性。当然深度学习的方法也存在一定局限,如对大规模数据的依赖、高计算资源需求以及较差的模型可解释性。尽管如此,深度学习在图像融合研究中仍然占据重要地位。

基于深度学习的图像融合方法主要可以分为基于神经网络的图像融合方法、基于生成对抗网络的图像融合方法、基于自编码器的图像融合方法。

基于神经网络的图像融合方法因其卓越的特征提取能力和优异的特征鲁棒性,已成为计算机视觉领域的研究热点,如基于卷积神经网络的图像融合框架及基于卷积稀疏表示的图像融合方法。文献[3]提出了一种基于卷积稀疏表示的图像融合方法,首先,将可见光图像和红外图像分别输入两个卷积神经网络,以获得对应的稀疏表示;其次,对这两个稀疏表示进行加权融合,得到融合后的稀疏表示;最后,采用逆卷积操作重构出融合图像。

基于生成对抗网络的图像融合方法核心思想是通过合并红外强度信息和可见光梯度,生成初始融合图像的生成器,通过判别器的不断学习来区分融合图像和可见光图像,以增加图像的细节信息^[4]。然而,这种方法存在一些问题,如生成的融合图像清晰度较低、保存的细节信息不足、可能引入一些不必要的噪声等。

基于自编码器的图像融合方法主要由编码器、融合层和解码器3部分组成。编码器负责提取输入图像的特征,提取的特征经过融合层处理后传递给解码器,最终生成融合图像。文献[5]提出的DeepFuse方法,通过自编码器实现图像融合,但其仅利用了编码器最后一层的输出特征,未充分利用所有层次的信息,导致部分图像细节信息的丢失;文献[6]提出DenseFuse方法,利用密集连接的DenseNet结构,以有效学习两幅图像中的特征并实现融合。DenseFuse的显著优势在于能够自动提取图像特征,避免了繁琐的人工特征设计过程。在此基础上,文献[7]提出了一种多层次分解与融合策略NestFuse,在结构上采用了改进的U-Net++网络,将图像分解为多个不同尺度的子图像,并对各尺度子图像分别进行融合处理,引入了跳跃连接机制以提升解码阶段的效率,同时结合了通道注意力机制与空间注意力机制,从而进一步增强了融合效果;文献[8]提出RFN-Nest方法,结合了残差融合网络与NestFuse算法,实现了图像的深度特征提取与多尺度融合,将两幅图像输入RFN网络,获得多尺度

的特征图;其次,利用 NestFuse 网络对各尺度特征图进行融合,得到融合后的特征表示;最后,将融合特征输入解码器,经过解码器的处理生成融合图像,实现了特征的有效融合。

近年来,深度学习因其强大的特征提取能力和端到端的学习特性,在图像融合领域取得了广泛关注。本文以真实种植环境为背景,采集可见光图像与飞行时间(ToF)图像并制作数据集,ToF 图像虽能精准、快速测量物体与传感器的距离且不受不良环境因素影响,但存在输出图像分辨率较低、难以提取足够特征的不足,将二者融合可有效提升图像质量。本文提出一种改进的图像融合网络模型 M-DenseFuse,该模型在 DenseFuse 的基础上进行结构优化,引入 U-net 网络架构以增强特征提取与重建能力,同时对 SENet 通道注意力机制加以改进并融入了 Transformer 中的多头注意力机制,从而在保持结构信息的同时进一步提升了融合图像的质量与细节表现能力。

2 本文模型

本文针对 DenseFuse 模型的特征提取性能瓶颈,引入 U-net 网络架构中的级联特征融合机制,通过构建多尺度特征层级交互结构,强化模型对不同分辨率图像特征的捕获与整合能力,有效提升特征表征的完整性与判别性;同时在模型编码器模块中创新性嵌入 M-SENet 通道注意力机制,旨在实现对图像关键信息区域的自适应聚焦,通过动态分配权重参数引导模型优先关注高信息密度的特征通道,从而在有限计算资源约束下实现计算资源的精准分配与高效利用,最终达成图像融合质量的显著提升。但 M-SENet 机制在经典 SENet 的基础上引入了 Transformer 架构中的多头注意力核心思想,通过构建多组并行的通道注意力头,使模型能够从不同维度学习输入特征通道间的复杂关联权重,进一步增强通道注意力的表达能力,显著提升特征表示的鲁棒性与区分度;从机制本质而言,通道注意力机制通过抑制冗余通道特征、强化关键通道信息的方式,有效提升模型对有用特征的挖掘效率,为模型性能优化提供关键技术支撑。

2.1 模型框架

训练阶段的模型如图 1 所示,在训练阶段中只使用了编码器以及解码器。由于训练的核心是提升解码器恢复图像的能力,无需进行图像融合操作,因此未设置融合层,Im1 和 Im2 代表原始图像,Im 则为融合后的输出图像。训练阶段的模型能够提取各

个尺度的特征,其中浅层特征会保留更多细节信息。

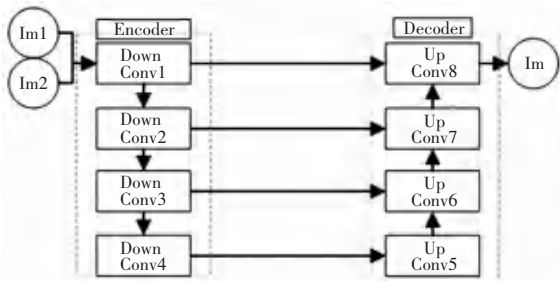


图 1 训练阶段
Fig. 1 Training phase

融合阶段的模型如图 2 所示。在融合阶段中使用了 3 个模块,编码器,解码器,融合层(FN)。Im1 和 Im2 分别输入模型中,编码器对其进行提取特征后,放入融合层(FN)中对其进行相加操作,融合后放入解码器中进行恢复图像。

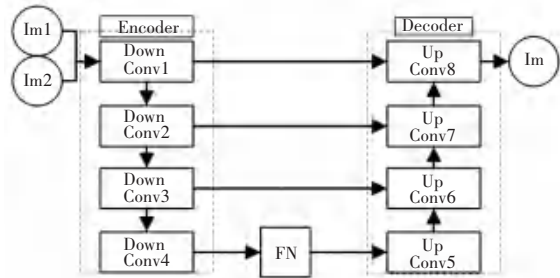


图 2 融合阶段
Fig. 2 Integration stage

2.2 下卷积架构

卷积操作可以提取图像的特征,同时也能够通过卷积运算来增强图像中的某些特征并减少噪声。

编码器中的下卷积是实现特征下采样与关键信息强化的核心模块,其内部整合最大池化(Maxpooling)、批量归一化(BatchNorm)、线性整流激活函数(ReLU)、M-SENet(改进型挤压-激励通道注意力机制)及卷积层(Conv)等关键操作,其架构如图 3 所示。

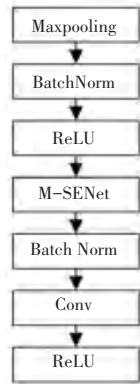


图 3 下卷积架构
Fig. 3 Down Conv architecture

Maxpooling 操作可以从图像中提取深度特征,计算滑动窗口内的数据的最大值,这样可以减少一些特征,节省计算资源。

BatchNorm 操作是对输入数据进行归一化处理,可以提高模型精准度,还能防止过拟合。

ReLU 是一种分段线性激活函数,当输入值为正时,输出与输入保持一致;当输入值为负时,输出恒为 0。该激活函数能有效缓解深度神经网络训练过程中的梯度消失问题,加速模型收敛速度,同时通过引入稀疏性特征提升模型的表达能力与泛化性能,从而在整体上优化模型的训练效率与最终性能表现。

M-SENet 作为基于 SENet 改进的多注意力头通道注意力模型,目的是增强输入特征图的有效通道,以改善特征表示的质量,每个头学习不同的通道权重,从而提高了模型的表达能力,其核心特点围绕多注意力头的并行计算与特征优化展开,具体架构如图 4 所示。M-SENET 模型中的多个注意力头,每个头都执行一遍注意力操作,全局平均池化以捕获每个通道的全局信息;通过两个线性层进行通道注意力的计算,包括降维和升维;使用 ReLU 激活函数进行非线性变换;最后,利用 Sigmoid 激活函数将通道注意力权重归一化到 0~1 的范围内,每个注意力头的输出经过 Sigmoid 激活后与输入特征图进行逐元素相乘,以强化感兴趣的通道,最终输出的是来自所有注意力头的结果。

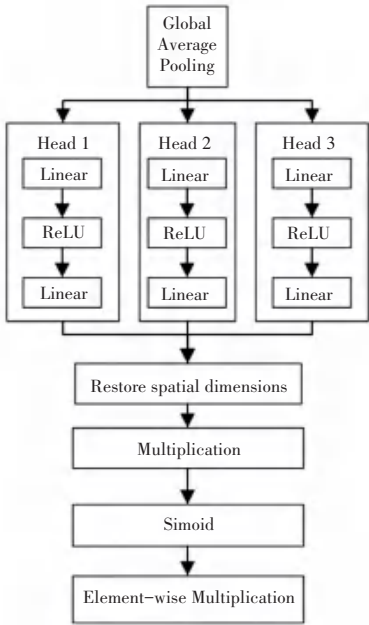


图 4 M-SENet 结构示意图
Fig. 4 M-SENet structure diagram

2.3 M-DenseFuse 模型参数

M-DenseFuse 模型结构见表 1,训练中输入通道,输出通道的整体结构,其中卷积参数表示的是每个卷积层中卷积核的大小和通道数。例如 1×1, 8→16 表示该卷积层使用 1×1 大小的卷积核,输入通道数为 8,输出通道数为 16。

表 1 M-DenseFuse 模型结构表
Table 1 M-Dense Fuse model structure table

卷积层	输入大小	输出大小	卷积参数
输入层	1, 256, 256	16, 256, 256	1×1, 8→16; 3×3, 16→16
下卷积 1	16, 256, 256	32, 128, 128	1×1, 16→8; 3×3, 8→32
下卷积 2	32, 128, 128	64, 64, 64	1×1, 32→16; 3×3, 16→64
下卷积 3	64, 64, 64	128, 32, 32	1×1, 64→32; 3×3, 32→128
下卷积 4	128, 32, 32	128, 16, 16	1×1, 128→64; 3×3, 64→128
上卷积 1	128, 16, 16	64, 32, 32	1×1, 256→128; 3×3, 128→64
上卷积 2	64, 32, 32	32, 64, 64	1×1, 128→64; 3×3, 64→32
上卷积 3	32, 64, 64	16, 128, 128	1×1, 64→32; 3×3, 32→16
上卷积 4	16, 128, 128	16, 256, 256	1×1, 32→16; 3×3, 16→16
输出层	16, 256, 256	1, 256, 256	1×1, 16→1

3 实验

3.1 数据集

本文采用的数据集包括公共数据集与自建数据集两类,Kfuji 公共数据集的样本采集自西班牙加泰

罗尼亚地区的一处商业富士苹果种植园,提供了具有区域代表性的公开数据支撑。自建数据集通过在真实苹果种植环境中开展系统性的苹果图像拍摄与数据采集构建而成,可针对性满足本文对特定场景下数据的需求,确保实验数据与实际应用场景的匹

配度,该数据集包括了 RGB 图像和 ToF 图像,其中 RGB 图像和 ToF 图像各 665 张,共计 1 330 张。此外,本文还收集了 70 张高曝光图像,为苹果检测和分类研究提供有价值的数据。

3.2 软件与环境

软件环境:Windows 11 操作系统,pycharm 社区版作为开发工具,Python 版本为 3. 7,torch 版本为 1. 311,cuda 版本为 12. 0. 133。硬件环境见表 2。

表 2 硬件环境	
Table 2 Hardware Environment Diagram	
配件	规格
CPU	11th Gen Intel(R) Core(TM) i5-11400H @ 2. 70 GHz 2. 69 GHz
硬盘	Samsung SSD 980
内存	Samsung DDR4 3200 MHz 16 G
显卡	NVIDIA GeForce RTX 3050 Laptop GPU

3.3 实验与分析

3.3.1 评价指标

本文选取 5 个指标对实验结果进行评价:

(1)信息熵(EN):描述图像中像素值分布的不确定性和规则性。一个熵值较低的图像,可以被视作是一个规律性非常高的图像,比如大片区域的完全相同或者规律的图案;一个熵值较高的图像,则代表出现了更多的像素值,不同的图案和纹理,即图像信息熵数值越大,说明图像更为多样化。

(2)平均梯度(AG):通过计算梯度来表征图像清晰程度,反映融合图像的细节和纹理信息。平均梯度越大,融合图像信息越丰富,融合性能越好。

(3)边缘强度(EIN):评价边缘检测效果,数值越大,图像边缘保持得越清晰、连续。

(4)边缘信息(QAB/F):图像 A、B 融合成 F 后,评价 F 的边缘情况,常用 Sobel 算子计算。

(5)图像清晰度(FD):通过计算每个像素与相邻像素的差值平均值,衡量图像清晰度和细节丰富度。

3.3.2 在公共数据集上实验与分析

将本文提出的模型与 DenseFuse、CBF^[9]、CNN^[10]、IVIF^[11]、FusionGAN 模型进行对比实验,实验结果如图 5 和表 3 所示。

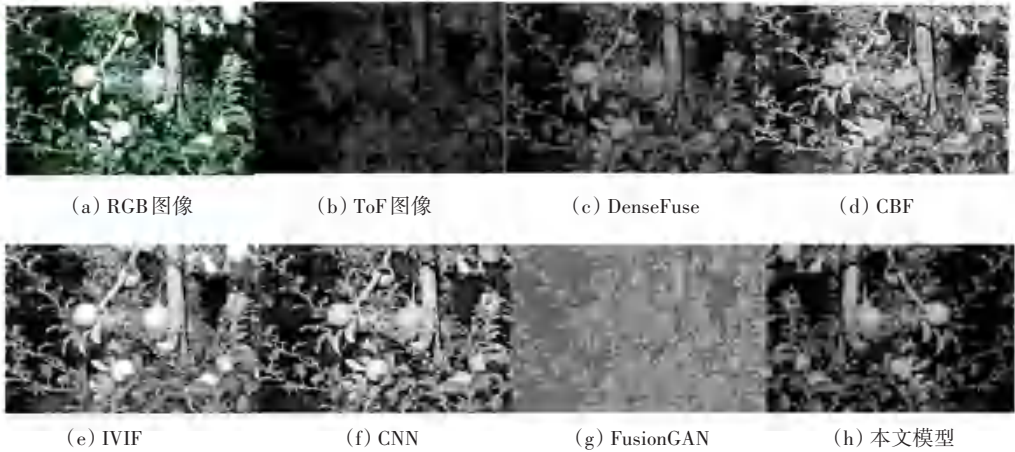


图 5 公共数据集对比图
Fig. 5 Comparison of Public Datasets

表 3 公共数据集指标对比					
Table 3 Comparison of public dataset indicators					
模型	EN	AG	EIN	QAB/F	FD
DenseFuse	7. 229 1	6. 319 4	68. 406 2	0. 432 4	6. 814 6
CBF	5. 481 0	6. 309 8	68. 511 4	0. 690 9	7. 267 4
CNN	5. 320 5	5. 808 0	63. 222 5	0. 745 0	6. 576 6
IVIF	5. 513 5	5. 458 6	56. 828 0	0. 728 5	5. 878 9
FusionGAN	6. 475 9	4. 937 2	53. 240 0	0. 158 6	5. 313 2
本文模型	7. 563 5	8. 034 1	75. 936 7	0. 429 0	7. 619 3

实验输入图像包含 2 类:RGB 图像和 ToF 图像。可见本文模型生成的融合图像优势显著,相较于其他对比模型,其在细节完整性、纹理清晰度及场景整体还原度上表现更优,能更精准地保留原始输入图像的关键信息,呈现出更接近真实场景的视觉效果,融合质量明显优于其余对比模型。

3.3.3 自建数据集实验与分析

曝光图像对比图如图 6 所示。从视觉效果来看,原始 RGB 曝光图像虽具备基础亮度与色彩信息,但存在曝光相关的细节压缩;ToF 图像则侧重空间结构与深度特征;融合图像通过有效整合两类输入图像的核心信息,不仅保留了 RGB 曝光图像的亮度特征与 ToF 图像的结构细节,还额外补充了更多

曝光场景下的关键特征信息,画面清晰度与细节丰富度显著增强,原始图像的边缘信息如物体轮廓、结构边界得以完整保留,整体视觉效果更符合人眼观察习惯,让曝光场景下的图像内容更易清晰识别。

非曝光图像对比图如图 7 所示。在非曝光场景中,原始 RGB 图像易因光线不足导致暗部细节缺失、画面层次感较弱;ToF 图像虽能提供基础空间结构,但视觉信息维度相对单一;融合图像不仅补充了非曝光场景下缺失的暗部特征信息,还进一步增强了图像的细节清晰度与整体画面通透感,完整保留了原始 RGB 图像与 ToF 图像的边缘信息,使得非曝光场景下的图像内容从“模糊难辨”转变为“清晰可察”,大幅提升了非曝光图像的识别效率。



图 6 曝光图像对比图

Fig. 6 Exposure image comparison chart

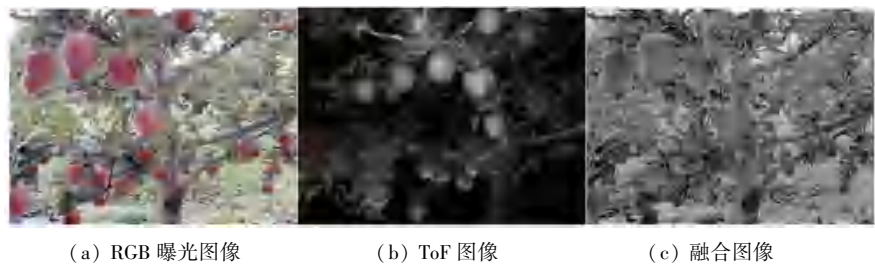


图 7 非曝光图像对比图

Fig. 7 Non exposure image comparison chart

在非曝光场景和曝光场景下在自建数据集的指标对比见表 4。可见在非曝光场景下本文模型的综合性能显著优于 DenseFuse 模型,各项评价指标均呈现明显优化趋势,其生成的融合图像像素值分布更具多样性,能够保留更多非曝光场景下的特征信息;本文模型 AG 值显著高于 DenseFuse 模型,说明融合图像的清晰度与细节丰富度更优,可有效改善非曝光场景下常见的暗部细节缺失问题;本文模型 QAB/F 值大于 DenseFuse 模型,证明其对非曝光图像边缘的保留与还原能力更强,能更完整地维持原始图像的结构边界特征。

针对曝光场景,本文模型虽在信息熵(EN)指标上与 DenseFuse 存在微小差异,但差值较小,对整体融合质量影响有限,且在关键性能指标上优势突出。

本文模型 AG 值显著高于 DenseFuse 模型,能有效缓解曝光场景下图像易出现的模糊问题,提升融合图像的细节锐利度;本文模型 QAB/F 值远高于 DenseFuse 模型,说明其可更好地抑制曝光区域的过亮干扰,精准保留目标物体的边缘轮廓与关键细节,进一步凸显曝光场景下的整体融合效果优势。

表 4 自建数据集指标对比

Table 4 Comparison of indicators for self built datasets			
模型	EN	AG	QAB/F
DenseFuse 非曝光图	6.80	8.30	0.06
本文模型非曝光图	7.37	12.85	0.55
DenseFuse 曝光图	6.20	3.60	0.04
本文模型曝光图像	6.10	6.42	0.43

3.4 消融实验

针对模型中的注意力机制进行消融,以验证其对融合结果的影响,消融实验结果见表 5,可见注意

力机制对本模型在 EN、SD、SF、AG、QAB/F 都有不小的提升。

表 5 消融实验结果
Table 5 Results of ablation experiment

模型	EN	SD	SF	AG	QAB/F
DenseFuse	7.229 1	10.2	0.053 9	6.319 4	0.432
DenseFuse+U-Net	6.053 7	54.5	0.089 0	5.263 0	0.073
DenseFuse+U-Net+SENet(本文模型)	7.563 5	59.9	0.141 3	12.856 0	0.551

从公共数据集、自建数据集的验证实验及消融实验结果来看,本文所提模型在图像融合任务中优势显著。相较于现有主流模型,该模型在维持融合图像清晰度的基础上,能够更高效地捕获多源输入图像的细节特征,并实现特征的深度融合,最终生成细节信息更丰富、具有视觉质量更优的融合结果。

4 结束语

为解决果园异源图像融合中光照动态变化、目标遮挡等复杂环境对融合质量的制约,本文提出并验证了一种结合多头注意力机制的异源图像融合模型。该模型以 DenseFuse 为基础框架,引入 U-Net 网络结构,强化多尺度特征提取能力,有效捕捉果园场景中苹果目标与枝叶背景的细粒度特征,同时创新性融合 SENet 通道注意力机制并结合 Transformer 多头注意力策略,通过对注意力计算流程的模块化重构与参数协同优化,构建适用于果园异源图像特征筛选的 M-SENet 通道注意力模块,实现对通道间权重关系的精细化建模,显著提升特征表示准确性,减少光照干扰与遮挡区域对融合过程的负面影响。在公共数据集、自建果园数据集的实验及消融实验结果表明,M-DenseFuse 模型在融合图像清晰度、苹果表皮纹理与边缘轮廓等目标细节保留度上优势显著,且在强光过曝、枝叶遮挡等复杂子场景中仍能稳定生成视觉质量优、信息完整性高的融合结果,也充分证明该模型对果园复杂环境的强适应性,为果园目标检测、病虫害识别等下游任务提供高质量特征输入,也为农业场景下异源图像融合技术的优化提供了有价值的参考。

参考文献

[1] 唐霖峰, 张浩, 徐涵, 等. 基于深度学习的图像融合方法综述 [J]. 中国图象图形学报, 2023, 28(1): 3-36.

[2] LI H, WU X J. Infrared and visible image fusion using latent low-rank representation [J]. arXiv preprint arXiv, 1804. 08992, 2018.

[3] YANG B, LI S. Multifocus image fusion and restoration with sparse representation [J]. IEEE Transactions on Instrumentation & Measurement, 2010, 59(4): 884-892.

[4] 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.

[5] PRABHAKAR K R, SRIKAR V S, BABU R V. Deepfuse: A deep unsupervised approach for exposure fusion with extreme exposure image pairs [C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway, NJ: IEEE, 2017: 4714-4722.

[6] LI H, WU X J. DenseFuse: A fusion approach to infrared and visible images [J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614-2623.

[7] LI H, WU X J, DURRANI T. NestFuse: An infrared and visible image fusion architecture based on nest connection and spatial/channel attention models [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12): 9645-9656.

[8] 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.

[9] KUMAR K B S. Image fusion based on pixel significance using cross bilateral filter [J]. Signal, Image and Video Processing, 2015, 9(5): 1193-1204.

[10] LIU Y, CHEN X, PENG H, et al. Multi-focus image fusion with a deep convolutional neural network [J]. Information Fusion, 2017, 36: 191-207.

[11] ZHANG Y, ZHANG L, BAI X, et al. Infrared and visual image fusion through infrared feature extraction and visual information preservation [J]. Infrared Physics & Technology, 2017, 83: 227-237.