

武汉大学学报(信息科学版)

Geomatics and Information Science of Wuhan University

ISSN 1671-8860, CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目：一种全天候道路场景下的红外和可见光图像融合方法
作者：甘文霞，潘俊杰，耿晶，王慧妮，胡小弟
DOI：10.13203/j.whugis20240173
收稿日期：2024-12-08
网络首发日期：2024-12-30
引用格式：甘文霞，潘俊杰，耿晶，王慧妮，胡小弟. 一种全天候道路场景下的红外和可见光图像融合方法[J/OL]. 武汉大学学报(信息科学版).
<https://doi.org/10.13203/j.whugis20240173>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20240173

引用格式：

甘文霞, 潘俊杰, 耿晶, 等. 一种全天候道路场景下的红外和可见光图像融合方法[J]. 武汉大学学报(信息科学版), 2024, DOI:10.13203/J.whugis20240173 (Gan Wenxia, Pan Junjie, Geng Jing, et al. A Fusion Method for Infrared and Visible Images in All-weather Road Scenes[J]. Geomatics and Information Science of Wuhan University, 2024, DOI:10.13203/J.whugis20240173)

一种全天候道路场景下的红外和可见光图像融合方法

甘文霞¹, 潘俊杰¹, 耿晶², 王慧妮^{1*}, 胡小弟¹

1 武汉工程大学土木工程与建筑学院, 湖北 武汉, 430074

2 北京理工大学计算机学院, 北京, 100081

摘要:结合红外图像的高亮目标信息及可见光图像丰富的纹理细节可以生成同时包含丰富纹理和显著目标的融合图像, 但受限于光照条件、恶劣天气的复杂多变性约束, 现有的红外和可见光图像融合算法在全天候道路应用场景中难以达到理想的效果。针对此问题, 提出一种适用于全天候道路场景的红外与可见光图像融合网络。首先通过光照损失函数设计, 指导融合网络自适应地保持背景纹理细节和显著目标的强度分布。此外, 纹理损失函数引入 Scharrr 梯度算子提高梯度计算精度, 从而约束融合图像保留更细腻的纹理细节。最后, 基于 Resblock 结构设计了二阶梯度残差块, 以增强融合网络提取图像强弱纹理特征的能力。在公开数据集 MFNet、Roadscene 和 TNO 数据上对该算法进行了对比实验和泛化实验, 结果表明算法在主观视觉效果和客观指标评价上均优于其它先进算法, 得到的融合图像纹理清晰且红外目标显著。

关键词: 全天候道路场景; 红外和可见光图像; 图像融合; 光照损失

A Fusion Method for Infrared and Visible Images in All-weather Road Scenes

Gan Wenxia¹, Pan Junjie¹, Geng Jing², Wang Huini^{1*}, Hu Xiaodi¹

1 School of Civil Engineering and Architecture, Wuhan Institute of Technology, Wuhan
430074 China

2 School of Computer Science & Technology, Beijing Institute of Technology, Beijing
100081 China

Abstract: The purpose of infrared and visible image fusion in the all-weather road scenes is to generate fusion images that contain rich textures and significant goals. However, due to the influence of light conditions and harsh weather factors, the

收稿日期: 2024-12-08

基金项目: 国家自然科学基金项目(42471388,42201461),湖北省科技厅科研项目(2023-121-3-4)。

第一作者: 甘文霞, 博士, 副教授, 从事遥感图像处理与应用、计算机视觉、深度学习等工作。charlottegan@whu.edu.cn

通信作者: 王慧妮, 博士, 副教授。wanghuini@wit.edu.cn

existing infrared and visible image fusion algorithm is difficult to achieve the ideal effect. In response to this issue, an infrared and visible image fusion network suitable for all-weather road scenes was proposed. First of all, we designed an illumination loss function that can guide the intensity distribution of the background texture and the intensity of significant goals in the adaptive network. In addition, the introduction of the Scharr gradient to improve the calculation accuracy of the gradient of texture loss, and restrict the fusion image to retain more delicate texture details. Finally, we designed the second order residual block based on the Resblock structure to enhance the ability to integrate network extraction of strong and weak texture features of images. Comparison and generalization experiments on MFNet, RoadScene, and TNO data show that the algorithm is better than other advanced algorithms in subjective visual effects and objective indicators.

Keywords: all-weather road scenes; infrared and visible image; image fusion; illumination loss

道路场景中的动态目标感知是自动驾驶、智能交通管理等的重要基础，实时、准确地对场景中的各类目标，如行人、车辆进行监测是应用的关键^[1-4]。但道路场景中，昼夜交替带来的光照差异及多变的天气环境等多因素耦合使得场景异常复杂，受限于感知设备的理论和技术，单一传感器获取的图像难以有效和全面地描述场景内容^[5]。

可见光传感器能够捕获光反射信息生成纹理丰富的图像，但对环境很敏感，易受光照、天气变化、遮挡等因素影响^[6-8]。红外传感器能够捕获物体的热辐射信息，受天气和光照条件变化影响较小，在低照度环境下能够清晰区分目标物体和非目标物体，但描述场景细节能力不足^[9-11]。通过图像融合技术综合利用两种传感器的互补性和相关性，在突出目标的同时更好地呈现背景纹理细节，生成具有丰富信息，有益于下游应用^[12-14]，也因而被高级视觉任务，如对象检测、目标跟踪等广泛采用。对象检测^[15-17]、目标跟踪^[18-20]、语义分割^[21-23]等。

随着图像融合技术的不断迭代发展，出现了许多不同的融合算法。传统方法如基于多尺度变换的方法^[24-27]、基于稀疏表示的方法^[28-30]、基于子空间分析的方法^[31-32]、基于优化的方法^[33]和混合方法^[34]等经过多年发展，方法复杂限制了实时性能，对权重配比和规则设计的依赖限制了方法在复杂场景中的适用性，进入了发展瓶颈期。近年来基于深度学习的融合方法有了迅猛的发展，大致可以归纳为基于自动编码器的框架^[35-37]，基于生成对抗性网络的框架^[38-40]和基于卷积神经网络的框架^[41-43]这三大类框架的方法。其中，基于自动编码器的融合框架通过训练自动编码器来实现特征提取及重构，2021年 Jian 等人设计了一种具有残差块的对称编解码器网络^[35]，通过残差块和卷积层结合基于注意力的融合策略提取并融合图像特征，同时通过复用补偿特征减少下采样损失对融合图像的影响，最后重构生成融合图像。在2019年，Ma 等人首次将 GAN 网络引入图像融合领域，提出一种基于对抗网络的融合模型^[38]，利用鉴别器约束生成器生成具有丰富纹理的融合图像。2022年，

Tang 等人将高级视觉任务和图像融合联系并提出了一种语义感知的实时图像融合网络 (SeAFusion, A semantic-aware real-time infrared and visible image fusion network) [43], 级联融合网络和语义分割模块并利用语义损失引导高级语义信息回流, 有效地提高了高级视觉任务在融合图像上的性能。另外, 梯度残差密集块及其轻量化的设计能够提高网络对图像细节的描述能力和融合效率。但是, SeAFusion 未考虑到光照条件及恶劣天气的影响, 在复杂环境下融合图像效果不够理想。

在全天候道路场景下, 昼夜光照不一以及恶劣天气会带来源图像信息分布差异以及遮挡缺失问题。因而导致融合图像出现纹理缺失问题。此外, 道路场景的自动驾驶或者交通监控摄像头中, 目标与传感器之间的距离存在较大差异, 近处物体在图像中像素占比大, 纹理更加清晰, 远处物体在图像中像素占比小, 纹理相对模糊, 即源图像中存在强弱纹理现象。因而导致图像融合中弱纹理目标信息丢失。

针对上述现象和不足, 本文提出一种适用于全天候道路场景的图像融合网络 (ARSFusion, All-weather Road Scenes-SeAFusion), 以 SeAFusion 实时融合网络为基础, 考虑光照差异和强弱纹理问题, 提升网络集成有效强度纹理的能力。本文研究的贡献总结如下: 首先提出一种基于光照判别加权的强度损失项构建内容损失函数, 能够根据场景中的不同光照条件自适应集成源图像中的强度信息; 并在纹理损失中引入 Scharr 梯度算子以保留更细腻的纹理细节; 此外, 在特征提取环节, 本文设计了一种二阶梯度残差模块 (SGRB, Second order Gradient Residual Block), 能够增强纹理细节特征提取能力, 通过二阶梯度残差流实现图像的强弱纹理特征提取。后文将对本文提出的模型进行介绍, 并以道路场景 MFNet、Roadscene 数据集以及 TNO 红外和可见光数据集为基础, 通过消融实验、对比实验以及泛化实验对本文方法进行综合分析及评价。

1 ARSFusion

1.1 ARSFusion 结构

ARSFusion 网络由特征提取器和图像重建器两部分组成。

特征提取器采用两个并行的特征提取流分别用来提取红外特征和可见光特征, 如图 1 (a) 所示, 每一个特征提取流由卷积层和二阶梯度残差块 SGRB 组成, 其中, 卷积层中卷积核大小为 3×3 , 激活函数为 LReLU, 用于提取图像细粒度特征。SGRB 是 Resblock^[44] 的一个变体, 结构如图 1 (a) 所示, 由主流和两个残差流构成。通过逐元素相加将输入与第二残差流的输出相加实现第一阶段的纹理增强, 然后将主流和第一残差流的输出在通道维度上进行级联处理, 以得到更深层次的细粒度特征, 实现第二阶段的纹理增强。该模块主流中部署了两个核大小 3×3 、激活函数为 LReLU 的通用卷积层和一个大小 1×1 的卷积层。第一个残差流采用 Sobel 算子提取图像强纹理特征, 并衔接 1×1 卷积层消除通道维度

的差异。第二个残差流采用 Laplacian 算子提取图像的弱纹理特征。

图像重建器由两个串联的 3×3 卷积层和一个 1×1 卷积层构成，其中， 3×3 卷积层采用 LReLU 作为激活函数， 1×1 卷积层采用 Tanh 激活函数。最后，采用级联策略将红外和可见光特征提取流提取的特征进行集成，最后将结果送入图像重建器，实现特征聚合和图像重建。

此外，SeAFusion 网络结构如图 1 (b) 所示，特征提取器中特征提取流由卷积层和两个梯度残差密集块 (GRDB, Gradient Residual Dense Block) 组成，图像重建器则由三个串联的 3×3 卷积层和一个 1×1 卷积层构成。简单来说，ARSFusion 结构在特征提取模块部分用 SGRB 替换了基础模型 SeAFusion 的两个串联 GRDB，而在图像重建模块较比基础模型少了一个 3×3 卷积层。

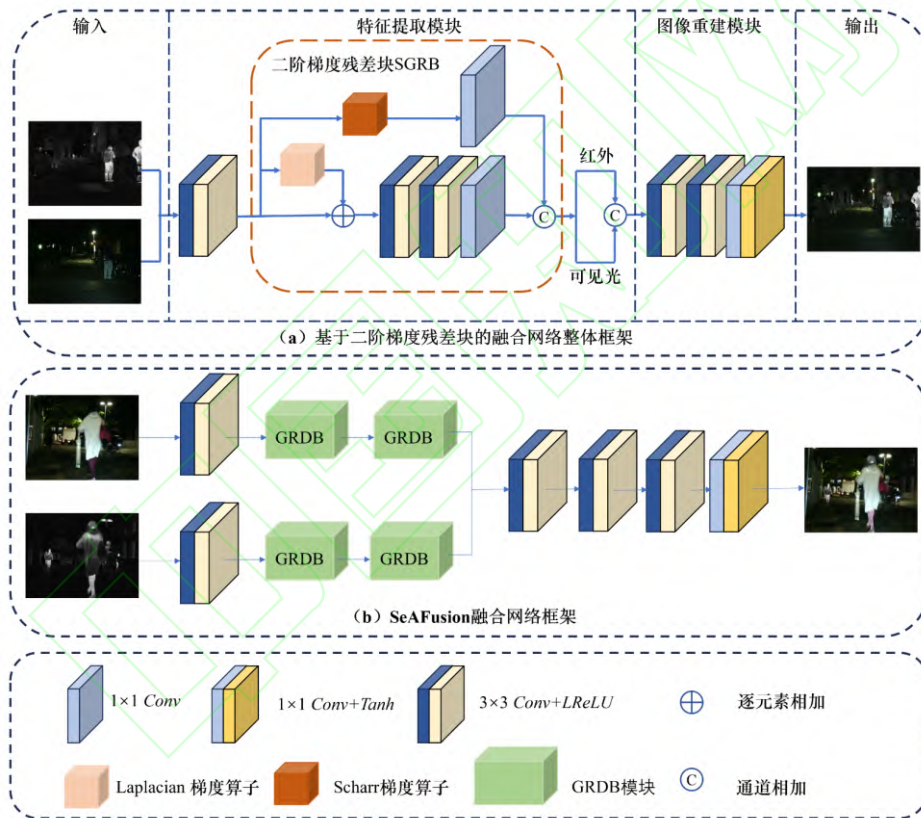


图 1 图像融合网络框架

Fig.1 The framework of the image fusion network

1.2 损失函数

ARSfusion 的损失函数由内容损失 ($\mathcal{L}_{content}$) 和语义损失 ($\mathcal{L}_{semantic}$) 方面联合构成，内容损失用于保留更多源图像信息，提升图像视觉质量；语义损失用于实现融合图像语义信息的增强，如公式 (1) 所示。

$$\mathcal{L}_{joint} = \mathcal{L}_{content} + \beta \mathcal{L}_{semantic} \quad (1)$$

式中超参数 β 用于平衡内容损失和语义损失，随训练迭代自适应变化。

1.2.1 内容损失

内容损失函数由强度损失、纹理损失和光照损失三部分组成。具体如公示(2)所示：

$$\mathcal{L}_{content} = \lambda_1 \mathcal{L}_{illum} + \lambda_2 \mathcal{L}_{int} + \lambda_3 \mathcal{L}_{texture} \quad (2)$$

光照损失 \mathcal{L}_{illum} 表达光照判别权重加持的像素级别上融合图像和源图像之间的差异，用来约束融合图像的强度分布，应该根据照明差异和不同的源图像保持一致，定义如下：

$$\mathcal{L}_{illum} = W_{vi} \cdot \mathcal{L}_{int}^{vi} + W_{ir} \cdot \mathcal{L}_{int}^{ir} \quad (3)$$

$$\mathcal{L}_{int}^{vi} = \frac{1}{HW} \| I_f - I_{vi} \|_1 \quad (4)$$

$$\mathcal{L}_{int}^{ir} = \frac{1}{HW} \| I_f - I_{ir} \|_1 \quad (5)$$

式中 \mathcal{L}_{int}^{vi} 和 \mathcal{L}_{int}^{ir} 分别为可见光图像和红外图像的强度损失， W_{vi} 和 W_{ir} 为光照权重，用于调整融合图像的强度约束， H 和 W 分别是输入图像的高和宽， $\|\cdot\|_1$ 为L1范数， I_f 为融合图像的像素值， I_{vi} 和 I_{ir} 分别为可见光图像和红外图像的像素值。其中，光照权重采用黑色像素判别法动态调整，设置像素阈值为60，低于阈值判定为黑色像素，然后计算黑色像素在可见光图像中的占比 P_n ，设定图片是夜晚的概率为 P_n ，白天的概率为 $1 - P_n$ 。由于白天大部分信息集中在可见光图像上，而红外图像在夜间包含更多场景描述信息，光照判别概率从侧面反映了源图像信息的丰富程度，因此设定光照权重 $W_{vi} = 1 - P_n$ ， $W_{ir} = P_n$ 。图像场景是白天的概率越高，融合图像中来自可见光图像强度信息越多，图像场景是夜晚的概率越高，融合图像中来自红外图像的强度信息越多。

强度损失 \mathcal{L}_{int} 表达像素级别上融合图像和源图像之间的差异，用来约束融合图像保持最佳强度分布，定义如下：

$$\mathcal{L}_{int} = \frac{1}{HW} \| I_f - \max(I_{vi}, I_{ir}) \|_1 \quad (6)$$

式中 $\max(\cdot)$ 表示元素中的最大项。

纹理损失 $\mathcal{L}_{texture}$ 表达融合图像和源图像之间的梯度差异，用来使融合图像保留更多纹理细节，定义如下：

$$\mathcal{L}_{texture} = \frac{1}{HW} \| |\nabla I_f| - \max(|\nabla I_{vi}|, |\nabla I_{ir}|) \|_1 \quad (7)$$

式中 ∇ 为图像纹理信息计算算子 Scharr，替换了 SeaFusion 网络中 Sobel 算子的改进，能够更好地提升梯度计算效率和精度。

1.2.2 语义损失

语义损失包括主要语义损失和辅助语义损失两个要素。主语义损失和辅助语义损失从

不同角度反映了融合图像中包含的语义信息，其表示如下：

$$\mathcal{L}_{semantic} = \mathcal{L}_{main} + \alpha \mathcal{L}_{aux} \quad (8)$$

式中 α 为平衡主语义损失和辅助语义损失的常数^[45]。

计算时首先由实时分割模型^[45]对融合图像分割，得到分割结果 $I_s \in \mathbb{R}^{H \times W \times C}$ 和辅助分割结果 $I_{sa} \in \mathbb{R}^{H \times W \times C}$ 后分别计算主要语义损失和辅助语义损失，定义如下：

$$\mathcal{L}_{main} = \frac{-1}{HW} \sum_{h=1}^H \sum_{w=1}^W \sum_{c=1}^C L_{so}^{(h,w,c)} \log(I_s^{(h,w,c)}) \quad (9)$$

$$\mathcal{L}_{aux} = \frac{-1}{HW} \sum_{h=1}^H \sum_{w=1}^W \sum_{c=1}^C L_{so}^{(h,w,c)} \log(I_{sa}^{(h,w,c)}) \quad (10)$$

式中 $L_{so} \in \mathbb{R}^{H \times W \times C}$ 表示从分割标签 $L_s \in \mathbb{R}^{H \times W \times C}$ 转换得到的独热向量。

1.3 训练策略

本文采用联合迭代训练的策略来训练融合网络和分割网络，并设置迭代轮次为 M 。联合损失的超参数 β 随训练迭代自适应调整，具体如下：

$$\beta = \gamma \times (m - 1) \quad (11)$$

式中 m 表示当前迭代轮次。可以看出，第一轮迭代语义损失未参与图像融合过程，所以 β 值为 0 ，但随着迭代次数的增加，语义损失参与融合过程且能够更准确的指导融合网络的训练，所以 β 也随之增加。 γ 是平衡语义损失和内容损失的常数。

2 实验验证与结果分析

2.1 数据集

实验使用武汉大学唐霖峰等人基于 MFNet^[21]构建的数据集进行模型训练及测试。MFnet 数据集包含 1569 对红外和可见光图像（820 个白天图像对和 749 个夜间图像对），空间分辨率为 480×640，但数据集中大部分红外图像数据的对比度和信噪比都较低，且存在部分图像对未经配准对齐处理。为此，唐等人首先筛选出配准对齐的 715 对白天图像和 729 对夜间图像数据，然后采取基于暗通道先验的图像增强算法处理数据中的红外图像，提高其对比度和信噪比，最终得到 1444 对高质量且完全配准的红外和可见光图像，数据集提供了 9 个对象的语义标签（包含车、行人、自行车、弯道、车站、护栏、警戒锥以及背景），并按 3:1 比例划分数据作为训练集和测试集。泛化实验数据集为 Roadscene^[41]和 TNO^[46]两种开源数据集，RoadScene 有 221 对已对齐的道路场景红外和可见光图像，数据集中图像均源于 20 世纪 80 年代极具代表性的场景前视红外视频，包含丰富的车辆、行人对象，图像分布场景包括白天、强光以及夜晚场景；TNO 数据集，具有 261 对红外和可见

光图像，图像数据均在半黑暗环境下采集，图像分布场景包括半黑暗及烟雾场景。

本文在以上三个数据集上开展了实验，并与九种其他方法进行了比较，包括两种传统方法：MDLatLRR^[26]和 GTF^[33]，两种基于 AE 自动编码器的方法：SEDRFuse^[35]和 CSF^[36]，两种基于 GAN 模型的方法：FusionGAN^[38]和 GANMcC^[40]，以及三种基于卷积神经网络的方法：SDNet^[42]、U2Fusion^[41]和 SeAFusion^[43]。本文选取的九种对比算法均按照作者原始论文设置参数进行模型训练。

2.2 实验设置

2.2.1 评价指标

本文选取互信息 (MI, Mutual Information)^[47]、标准差 (SD, Standard Deviation)、视觉信息保真度 (VIF, The Visual Information Fidelity for fusion)^[48]和 Qabf^[49]等 4 个评价指标进行融合结果的定量评估分析。MI 指标从基于信息论的角度反映出从源图像传递至融合图像的信息量，SD 指标从统计学的角度反映出融合图像的灰度级分布情况，VIF 指标基于人类视觉系统评估融合图像的信息保真度，Qabf 指标利用局部度量估计源图像的显著信息在融合图像中的表现程度。

2.2.2 训练参数

本文根据联合训练策略迭代训练融合网络和分割网络，针对融合模型设置不同 Batch size 和 Epoch 组合的模型实验，并在 MFNet 数据集上进行性能验证，随着两种参数值增加，训练耗时加长，且模型性能降低甚至出现融合图像色彩严重失真。综合训练耗时和模型性能两方面考虑，设置训练参数如下：迭代轮次 M 为 4，每轮迭代训练轮次(Epoch)为 10，参数更新次数分别为 2700 和 20000。并使用 ADAM 优化器更新参数，批处理大小 (Batch size)为 4，初始学习率为 0.0001，权重衰减系数为 0.0002。

另外，采用随机梯度下降法优化分割模型，并按照原始论文设置批处理大小(Batch size)为 16，动量因子为 0.9，初始学习率为 0.01，权重衰减系数为 0.0005。本文模型实验平台为 Windows 10 操作系统，GPU 型号为 NVIDIA GeForce RTX 2080 Ti，CPU 型号为 Intel Core i9-10980XE，运行内存为 128GB。编程环境为 Python 和 PyTorch 平台^[50]。

2.3 消融研究

为了验证不同改进模块对图像融合性能的影响，本文设计了一系列消融实验，在相同数据集和训练参数的条件下对不同改进模块组合的模型精度进行了定性和定量比较，结果如图 2 和表 1 所示。

如图 2 所示，红框中为热目标信息，右下角绿框为图中小绿框区域等比放大后的纹理信息，原始模型所生成的融合结果出现了部分红外亮度信息的丢失以及纹理模糊的情况。相比之下，各改进模型均能够保留显著的红外亮度信息，且 Scharr 和 SGRB 模型能够有

效解决纹理模糊的问题。总的来说，ARSFusion 融合模型保留了更显著的红外亮度目标以及更清晰的纹理细节，生成视觉效果更佳的融合图像。

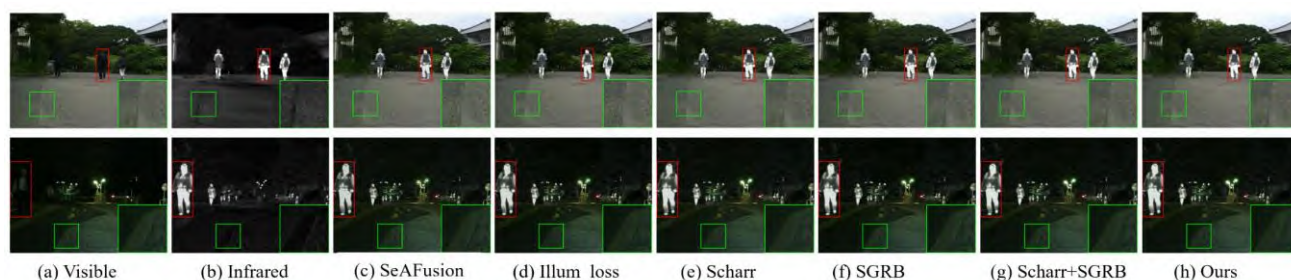


图 2 消融研究的可视化结果
Fig.2 Visualized results of ablation studies

如表 1 所示，融合模型仅添加光照判断损失函数在指标 Q_{abf} 取得最优值，表明此损失函数能够有效提升源图像的显著信息在融合图像中的表现程度。而仅使用纹理损失梯度算子 Scharr 较比原始模型在四项指标上均有提升。此外，融合模型单独添加 SGRB 模块在指标 SD 和 VIF 上均取得最优值，证明此模块能够有效保留源图像的特征信息，提升融合图像质量。总体而言，本文提出的 ARSFusion 模型综合了各模块的优点，模型指标较比基础 SeAFusion 有较大提升，互信息 MI 提升 1.2336，标准差 SD 提升 0.9755，融合质量 Q_{abf} 提升 1.97%，视觉保真度 VIF 提升 0.0751。在融合质量评价和视觉效果两方面都表现较好。

表 1 各模块在 MFNet 测试集上的消融研究定量比较（红色代表最优结果，蓝色代表次优结果）
Tab.1 Quantitative comparison of ablation studies on various modules on the MFNet dataset
(RED indicates the best result and BLUE represents the second best result)

| 模型 | Illum_Loss | Scharr | SGRB | MI | SD | Q_{abf} | VIF |
|-----------|------------|--------|------|--------|---------|-----------|---------|
| SeAFusion | | | | 3.8874 | 41.2680 | 0.6657 | 0.9402 |
| 模型 A | ✓ | | | 4.1454 | 42.2784 | 0.7001 | 0.9973 |
| 模型 B | | ✓ | | 4.5507 | 42.2156 | 0.6781 | 0.9746 |
| 模型 C | | | ✓ | 4.8943 | 42.3887 | 0.6820 | 1.0289 |
| 模型 D | | ✓ | ✓ | 5.0089 | 42.3174 | 0.6830 | 1.00603 |
| 本文方法 | ✓ | ✓ | ✓ | 5.1210 | 42.2435 | 0.6854 | 1.0153 |

2.4 对比实验

如图 3 所示，在白天场景中，红外图像的热辐射信息可以作为可见光图像的补充信息。GTF、CSF 和 FusionGAN 未能保留可见光图像中的纹理信息。而 MDLatLRR、SEDRFuse、GANMcC、SDNet 和 U2Fusion 能够结合可见光纹理信息和红外热辐射信息，但其背景均

存在不同程度的阴影和模糊问题，导致视觉效果较差。另外，SEDRFuse 和 U2Fusion 对红外目标信息有所削弱。而本文方法和 SeAFusion 融合的图像保留了纹理细节和亮度目标。

如图 4 所示，在夜晚场景中，可见光图像包含的纹理信息较少，红外图像的亮度信息和丰富纹理信息可以作为补充信息。所有方法均能够保留显著的亮度信息，GTF、CSF、GANMcC、SDNet 和 U2Fusion 无法清楚地显示护栏和树干。除了本文方法和 SeAFusion，其他方法均存在程度不一的背景纹理模糊情况。

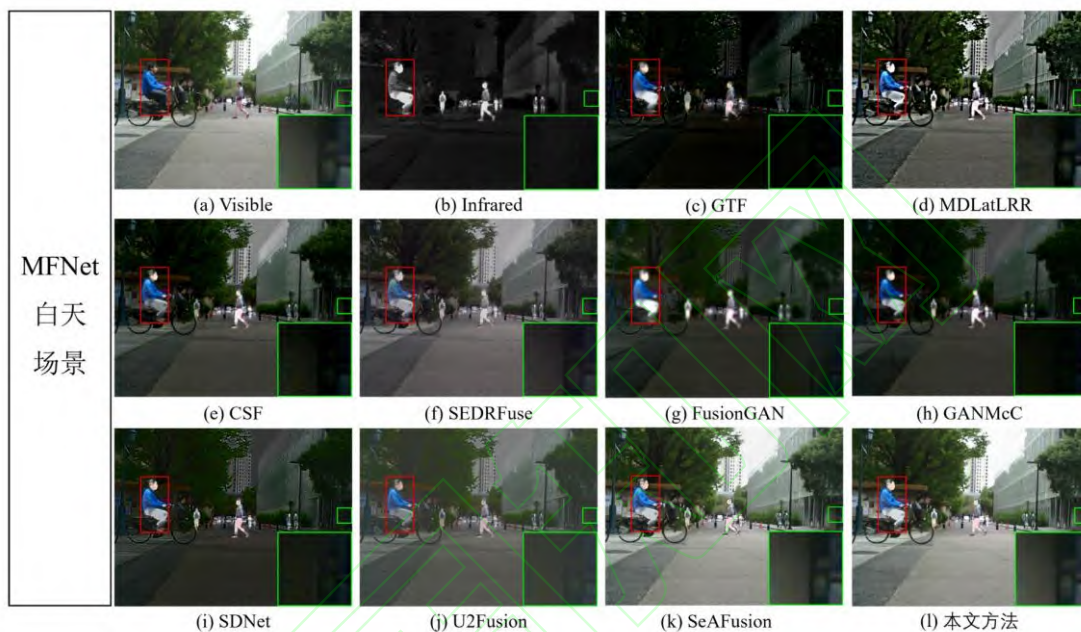


图 3 MFNet 数据集上本文方法和 9 种最先进方法的白天场景定性比较

Fig.3 Qualitative comparison of the proposed method with 9 state-of-the-art methods on the MFNet dataset in day

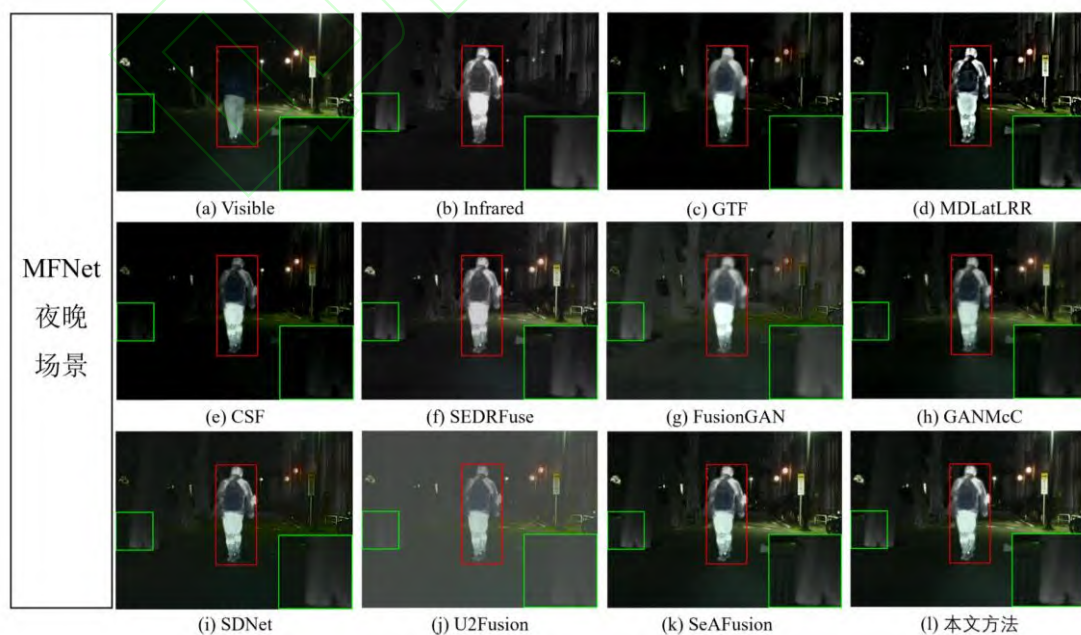


图 4 MFNet 数据集上本文方法和 9 种最先进方法的夜晚场景定性比较

Fig.4 Qualitative comparison of the proposed method with 9 state-of-the-art methods on the MFNet dataset in night

图 5 展示了不同方法在 MFNet 数据集中的定量指标评价结果，可以看出本文方法在 MI、SD、VIF 和 Qabf 四项指标上都表现出显著的优势。最佳 MI 指标 5.121 代表本文方法能够从源图像传递更多的信息至融合图像。最佳 SD 指标 42.2435 和 VIF 指标 1.0153 表明本文融合图像具有更高对比度且有最好的视觉效果，更符合人类视觉系统。此外，本文方法在 Qabf 指标上也取得最优值 0.6854，这表明本文融合图像中保留了更多源图像的显著信息。

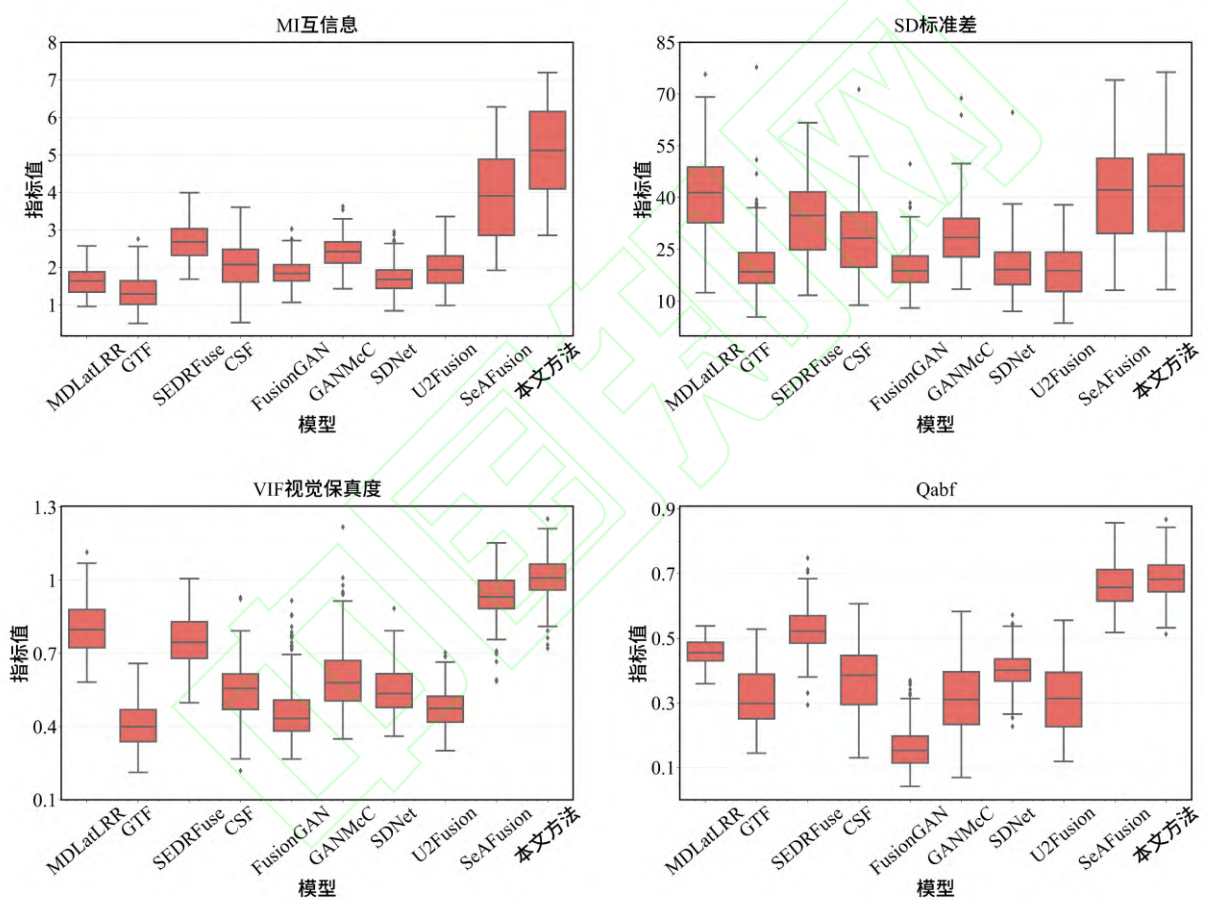


图 5 MFNet 数据集 361 对图像上不同融合方法指标 MI、SD、VIF、Qabf 的定量比较
 Fig.5 Quantitative comparisons of MI,SD,VIF,Qabf of different fusion methods on 361 image pairs from the MFNet dataset

2.5 泛化实验

图 6 和图 7 展示了方法在 Roadscene 数据集白天和强炫光场景下的两副图像上不同算法的融合结果。从图中可以观察到，除了本文方法和 SeAFusion，其它方法结果的背景纹理区域均受到了热辐射的影响，具体表现为背景灰暗、纹理模糊。而 GTF、FusionGAN 和

GANMcC 无法维持显著目标的锐利边缘，SEDRFuse 和 U2Fusion 无法保持显著目标的强度。另外，在强炫光图像中可以看到，SeAFusion 背景区域也出现一定程度上的纹理模糊现象，而本文方法仍然能够保持清晰的纹理细节，生成视觉效果良好的融合图像。

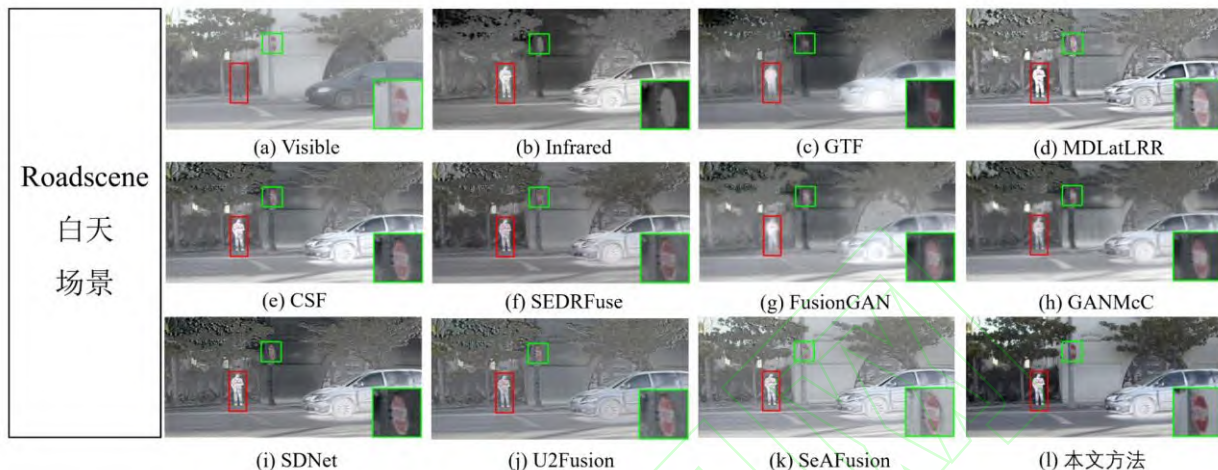


图 6 Roadscene 数据集上本文方法和 9 种最先进方法的白天场景定性比较

Fig.6 Qualitative comparison of the proposed method with 9 state-of-the-art methods on the Roadscene dataset in day

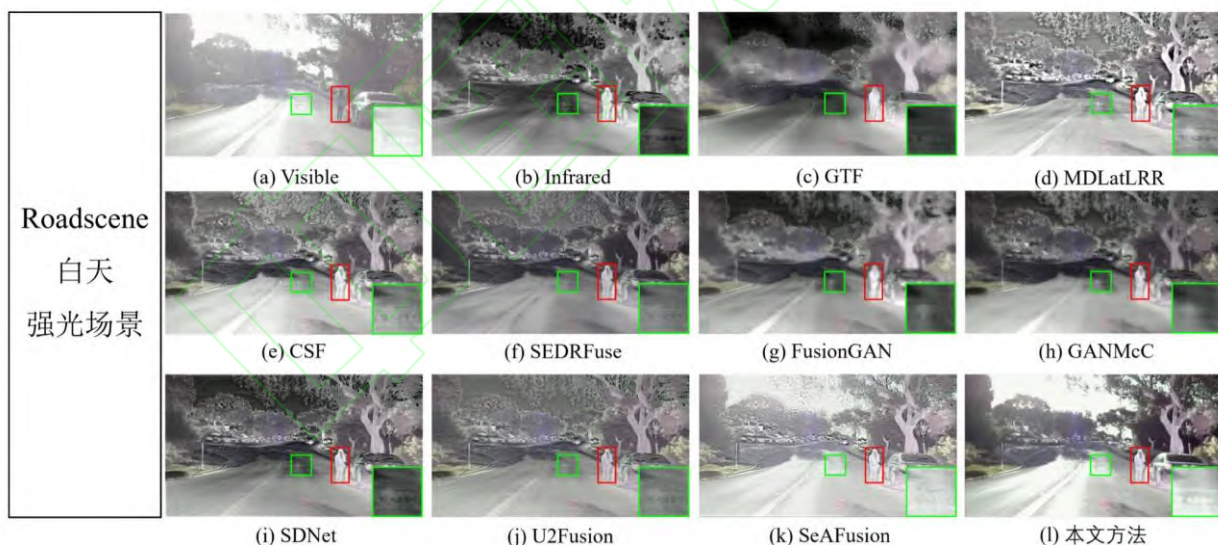


图 7 Roadscene 数据集上本文方法和 9 种最先进方法的强光场景定性比较

Fig.7 Qualitative comparison of the proposed method with 9 state-of-the-art methods on the Roadscene dataset in strong light scene

图 8 和图 9 展示了方法在 TNO 数据集夜晚和浓雾场景下的两副图像上不同算法的融合结果。从红框区域可以看出，GTF、MDLatLRR、CSF、SEDRFuse 和 U2Fusion 削弱了显著目标的强度，GTF、FusionGAN 和 GANMcC 模糊了显著目标的边缘。此外，GTF、

FusionGAN、GANMcC 和 SDNet 均出现出一定程度的纹理模糊，只有本文方法能够同时保持可见光的纹理细节和显著目标的强度。

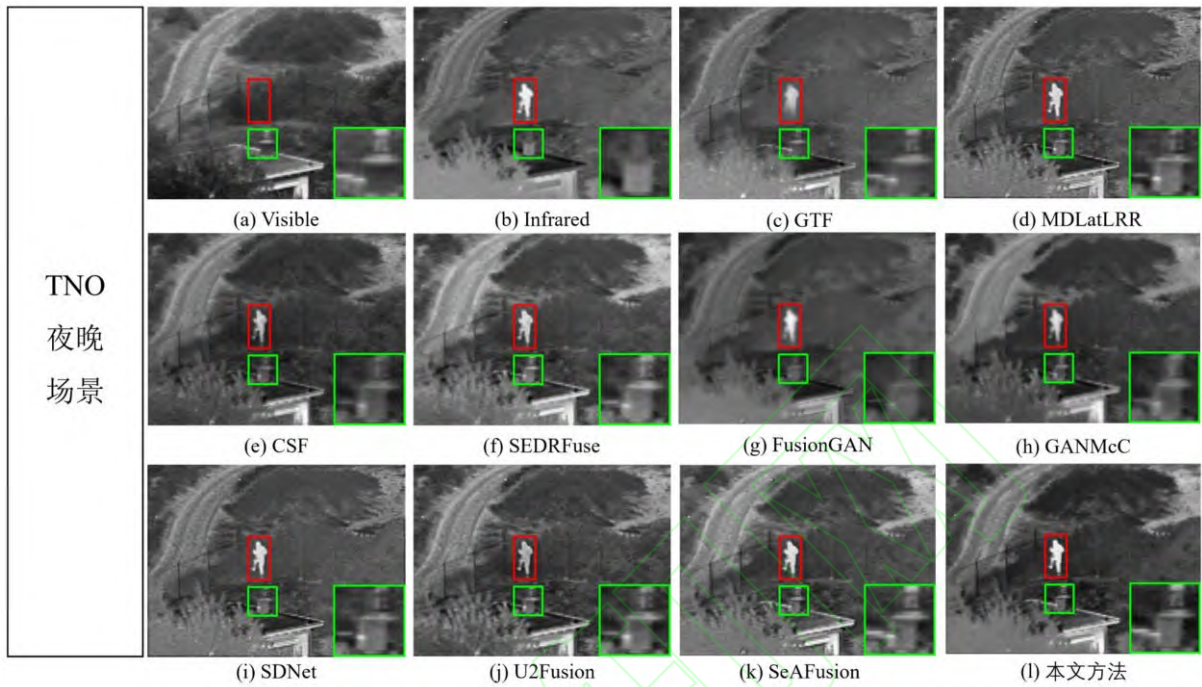


图 8 TNO 数据集上本文方法和 9 种最先进方法的夜晚场景定性比较

Fig.8 Qualitative comparison of the proposed method with 9 state-of-the-art methods on the TNO dataset in night

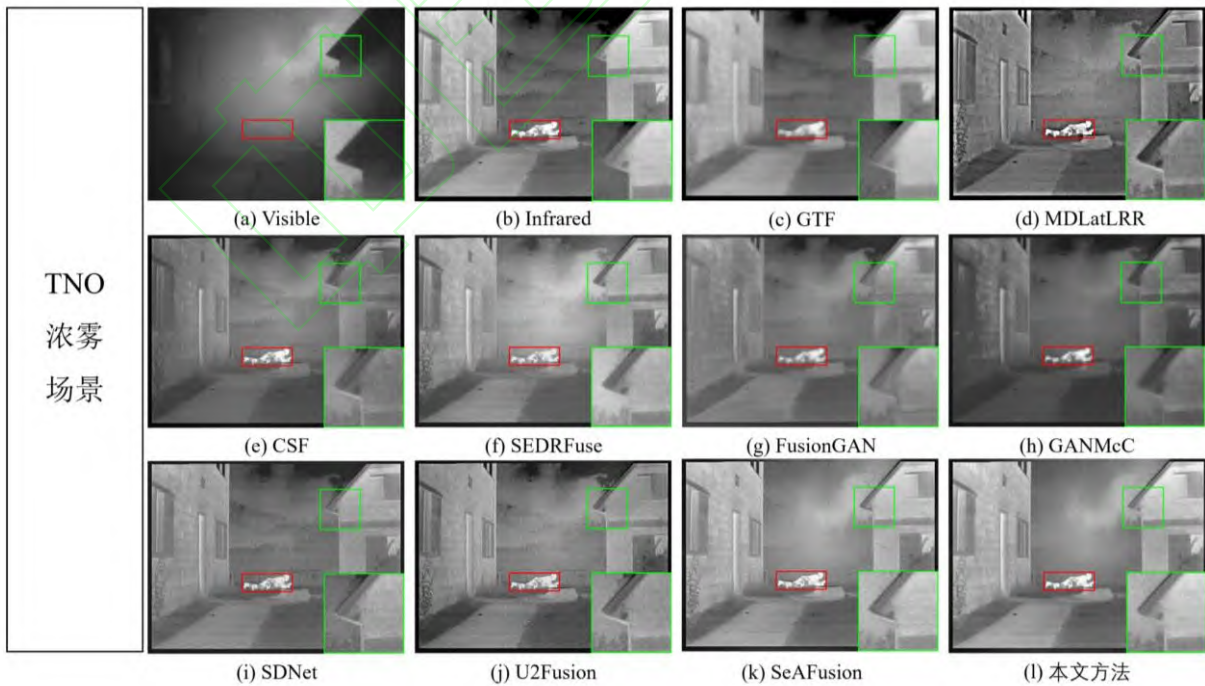


图 9 TNO 数据集上本文方法和 9 种最先进方法的浓雾场景定性比较

Fig.9 Qualitative comparison of the proposed method with 9 state-of-the-art methods on the TNO dataset in

thick fog scene

方法在 Roadscene 数据集和 TNO 数据集中上的结果定量评价结果如图 10 所示，在 Roadscene 数据集上，ARSFusion 在 MI、VIF 和 Qabf 指标上表现出较大的优势，在 SD 指标上，本文方法仅次于 GTF。在 TNO 数据集上，ARSFusion 在 MI、SD、VIF 和 Qabf 指标上均表现出显著优势。表明本文方法能够从源图像传递更多的信息至融合图像，生成纹理清晰、视觉效果最好的融合图像。

在泛化数据集上结果表明本文方法在保持背景区域的纹理细节以及目标区域的强度分布方面具有优势，归因于以下三个原因：1) 光照损失函数的设计能够指导融合模型自适应地保持显著目标的强度分布以及背景区域的纹理细节。2) 由光照损失、强度损失和纹理损失组成的内容损失能够约束融合网络从强度分布和纹理细节方面集成有意义的信息；3) 二阶梯度残差块能够增强融合网络对源图像细粒度特征的提取能力。

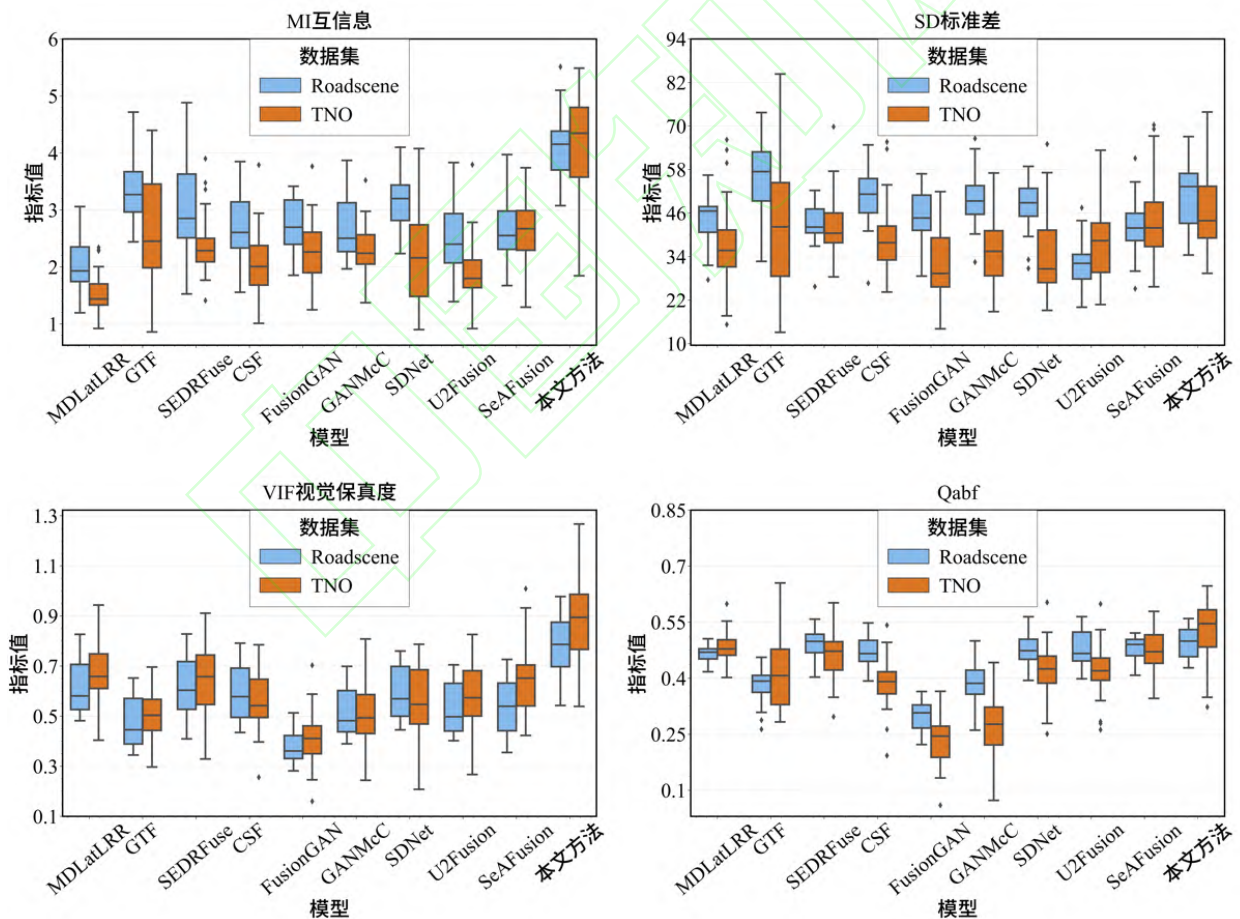


图 10 Roadscene 数据集和 TNO 数据集上不同融合方法指标 MI、SD、VIF、Qabf 对比结果

Fig.10 Quantitative comparisons of MI,SD,VIF,Qabf of different fusion methods on the Roadscene dataset and the TNO dataset

2.6 效率比较

为了评估不同算法的融合效率, 本文对不同算法的运行效率进行了比较, 如表 2 所示, 传统方法在融合图像方面耗时更多, 深度学习的方法的运行效率明显更高, 其中本文方法在 Roadscene, TNO 两个数据集上的运行时间分别为 0.13 秒和 0.16 秒, 优于其他方法, 在 MFNet 数据上为 0.19 秒, 仅次于 FusionGAN 和 SDNet。

表 2 不同方法运行效率 (单位: 秒, 红色代表最优结果, 蓝色代表次优结果)

Tab.2 Mean of the running times of all methods on all datasets (unit: second, RED indicates the best result and BLUE represents the second best result)

| 模型 | MFNet | Roadscene | TNO |
|-----------|----------|-----------|----------|
| MDLatLRR | 153.7452 | 53.3292 | 103.1227 |
| GTF | 5.6339 | 1.7974 | 3.0739 |
| SEDRFuse | 10.1583 | 5.6843 | 9.7324 |
| CSF | 2.8760 | 2.5481 | 3.1670 |
| FusionGAN | 0.0863 | 0.4735 | 0.3175 |
| GANMcC | 0.1499 | 0.3175 | 0.5165 |
| SDNet | 0.0242 | 0.2042 | 1.2235 |
| U2Fusion | 0.8802 | 0.4309 | 0.7424 |
| SeAFusion | 0.2083 | 0.1414 | 0.1729 |
| 本文方法 | 0.1959 | 0.1329 | 0.1635 |

3 结论

针对全天候道路场景下光照不一和恶劣天气所带来的源图像信息分布差异和遮挡缺失问题, 本文在语义感知融合网络 SeAFusion 的基础上, 设计了二阶梯度残差块来提高融合网络对图像强弱纹理特征的提取能力; 引入光照判别损失增强融合网络根据光照条件自适应地集成有效信息的能力; 在纹理损失中采用 Scharr 梯度算子提高网络梯度计算精度。而光照损失、强度损失和纹理损失组成的内容损失能够有效约束融合网络生成包含更多纹理细节和红外热信息的图像。实验结果表明, 本文融合方法 ARSFusion 在主观效果均优于其它主流方法, 改进模型在 MI、SD、Qabf 和 VIF 四项定量指标上远优于其他九种主流模型, 较比基础模型在 MFNet 数据集上的四项指标提升百分比分别为 31.73%、2.36%、2.96%和 7.98%, 且在三种数据集上单对图像的平均融合速度分别为 0.1959 秒、0.1329 秒和 0.1635 秒, 能够在道路全天候场景下快速有效地生成主观视觉良好、客观指标均佳的红外和可见光融合图像。

参考文献

- [1]. 应申,蒋跃文,顾江岩等.面向自动驾驶的高精地图模型及关键技术[J/OL].武汉大学学报(信息科学版):1-12[2023-11-23].<https://doi.org/10.13203/j.whugis20230227>.
- [2]. 申彩英,朱思瑶,黄兴驰.双目视觉的智能汽车目标检测算法研究[J/OL].重庆理工大学学报(自然科学):1-8[2023-11-06].<http://kns.cnki.net/kcms/detail/50.1205.T.20231011.1117.004.html>.
- [3]. 胡淼,姜麟,陶友凤等.改进 YOLOv7 的自动驾驶目标检测算法[J/OL].计算机工程与应用:1-11[2023-11-06].<http://kns.cnki.net/kcms/detail/11.2127.TP.20230922.1630.004.html>.
- [4]. 叶语同,李必军,付黎明.智能驾驶中点云目标快速检测与跟踪[J].武汉大学学报(信息科学版),2019,44(01):139-144+152.DOI:10.13203/j.whugis20170146.
- [5]. Zhang H, Xu H, Tian X, et al. Image fusion meets deep learning: A survey and perspective[J]. Information Fusion, 2021, 76: 323-336.
- [6]. Ma J, Ma Y, Li C. Infrared and visible image fusion methods and applications: A survey[J]. Information fusion, 2019, 45: 153-178.
- [7]. Zhang X, Demiris Y. Visible and Infrared Image Fusion Using Deep Learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.
- [8]. Ma W, Wang K, Li J, et al. Infrared and Visible Image Fusion Technology and Application: A Review[J]. Sensors, 2023, 23(2): 599.
- [9]. 韩林凯,姚江伟,王坤峰.保留梯度和轮廓的可见光与红外图像融合 [J/OL].计算机应用, 1-7[2023-11-07] <http://kns.cnki.net/kcms/detail/51.1307.TP.20230417.1032.002.html>.
- [10]. Shuai H, Tian H, Beiyi A, et al. VDFEFuse: A novel fusion approach to infrared and visible images[J]. Infrared Physics and Technology, 2022, 121.
- [11]. Sun C, Zhang C, Xiong N. Infrared and visible image fusion techniques based on deep learning: A review[J]. Electronics, 2020, 9(12): 2162.
- [12]. 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(03): 266-275.
- [13]. 向天烛,高熔溶,闫利等.一种顾及区域特征差异的热红外与可见光图像多尺度融合方法[J].武汉大学学报(信息科学版), 2017, 42(07): 911-917. DOI: 10.13203/j.whugis20141007.
- [14]. Li S, Kang X, Fang L, et al. Pixel-level image fusion: A survey of the state of the art[J]. information Fusion, 2017, 33: 100-112.
- [15]. 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. Li C, Song D, Tong R, et al. Illumination-aware faster R-CNN for robust multispectral pedestrian detection[J]. Pattern Recognition, 2019, 85: 161-171.

- [16]. 苑朝,赵亚冬,张耀等.基于 YOLO 轻量化的多模态行人检测算法[J/OL].图学学报:1-12[2023-11-07].<http://kns.cnki.net/kcms/detail/10.1034.T.20231026.1644.002.html>.
- [17]. 孙文财,胡旭歌,杨志发等.基于红外及可见光图像融合的道路目标检测优化方法[J/OL].吉林大学学报(工学版):1-8[2023-11-07].<https://doi.org/10.13229/j.cnki.jdxbgxb.20230474>.
- [18]. Geng X, Li M, Liu W, et al. Person tracking by detection using dual visible-infrared cameras[J]. IEEE Internet of Things Journal, 2022, 9(22): 23241-23251.
- [19]. 刘万军,梁林林,曲海成.利用 Transformer 的多模态目标跟踪算法[J/OL].计算机工程与应用, 1-11[2023-11-07]<http://kns.cnki.net/kcms/detail/11.2127.TP.20230508.1642.022.html>.
- [20]. Liu W, Liu W, Sun Y. Visible-Infrared Dual-Sensor Fusion for Single-Object Tracking[J]. IEEE Sensors Journal, 2023, 23(4): 4118-4128.
- [21]. Ha Q, Watanabe K, Karasawa T, et al. MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes[C]//2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017: 5108-5115.
- [22]. Hou J, Zhang D, Wu W, et al. A generative adversarial network for infrared and visible image fusion based on semantic segmentation[J]. Entropy, 2021, 23(3): 376.
- [23]. 朱浩,谷小婧,蓝鑫等.基于多尺度轮廓增强的 RGB-IR 双波段图像语义分割算法[J].激光与红外,2022,52(04):543-551.
- [24]. Wang Z, Xu J, Jiang X, et al. Infrared and visible image fusion via hybrid decomposition of NSCT and morphological sequential toggle operator[J]. Optik, 2020, 201: 163497.
- [25]. Chen J,Li X,Luo L, et al. Infrared and visible image fusion based on target-enhanced multiscale transform decomposition[J]. Information Sciences,2020,508.
- [26]. Hui L,Xiao-Jun W,Josef K. MDLatLRR: A novel decomposition method for infrared and visible image fusion.[J]. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society,2020,29.
- [27]. 赵庆典,杨德宏.基于图像增强和二次 NSCT 的红外与可见光图像融合[J/OL].激光与光电子学进展:1-17[2023-11-07].<http://kns.cnki.net/kcms/detail/31.1690.TN.20230714.1054.202.html>.
- [28]. 武凌霄,康家银,姬云翔.NSST 域下基于引导滤波与稀疏表示的红外与可见光图像融合[J].红外技术,2023,45(09):915-924.
- [29]. Liu Y,Chen X,Ward K R, et al. Image Fusion With Convolutional Sparse Representation.[J]. IEEE Signal Process. Lett.,2016,23(12).
- [30]. Liu Y, Yang X, Zhang R, et al. Entropy-based image fusion with joint sparse representation and rolling guidance filter[J]. Entropy, 2020, 22(1): 118.
- [31]. Cvejic N, Bull D, Canagarajah N. Region-based multimodal image fusion using ICA bases[J]. IEEE Sensors Journal, 2007, 7(5): 743-751.

- [32]. Bavirisetti D P, Xiao G, Liu G. Multi-sensor image fusion based on fourth order partial differential equations[C]//2017 20th International conference on information fusion (Fusion). IEEE, 2017: 1-9.
- [33]. Ma J, Chen C, Li C, et al. Infrared and visible image fusion via gradient transfer and total variation minimization[J]. *Information Fusion*, 2016, 31.
- [34]. Ma J, Zhou Z, Wang B, et al. Infrared and visible image fusion based on visual saliency map and weighted least square optimization[J]. *Infrared Physics and Technology*, 2017, 82.
- [35]. Lihua J, Xiaomin Y, Zheng L, et al. SEDRFuse: A Symmetric Encoder-Decoder With Residual Block Network for Infrared and Visible Image Fusion[J]. *IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT*, 2021, 70.
- [36]. Han X, Hao Z, Jiayi M. Classification Saliency-Based Rule for Visible and Infrared Image Fusion[J]. *IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING*, 2021, 7.
- [37]. 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.
- [38]. Ma J, Yu W, Liang P, et al. FusionGAN: A generative adversarial network for infrared and visible image fusion[J]. *Information Fusion*, 2018, 48.
- [39]. Yang Y, Liu J, Huang S, et al. Infrared and visible image fusion via texture conditional generative adversarial network[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 31(12): 4771-4783.
- [40]. Ma J, Zhang H, Shao Z, et al. GANMcC: A generative adversarial network with multiclassification constraints for infrared and visible image fusion[J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 70: 1-14.
- [41]. 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.
- [42]. Zhang H, Ma J. SDNet: A versatile squeeze-and-decomposition network for real-time image fusion[J]. *International Journal of Computer Vision*, 2021, 129: 2761-2785.
- [43]. 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.
- [44]. Ma J, Tang L, Xu M, et al. STDFusionNet: An infrared and visible image fusion network based on salient target detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 1-13.
- [45]. Peng C, Tian T, Chen C, et al. Bilateral attention decoder: A lightweight decoder for real-time semantic segmentation[J]. *Neural Networks*, 2021, 137: 188-199.
- [46]. 4Toet A. TNO Image Fusion Dataset [Internet]. figshare; 2014, <https://doi.org/10.6084/m9.figshare.1008029.v2>

- [47]. Qu G,Zhang D,Yan P. Information measure for performance of image fusion[J]. Electronics Letters,2002,38(7).
- [48]. 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.
- [49]. Piella G, Heijmans H. A new quality metric for image fusion[C]//Proceedings 2003 international conference on image processing (Cat. No. 03CH37429). IEEE, 2003, 3: III-173.
- [50]. Paszke A, Gross S, Massa F, et al. Pytorch: An imperative style, high-performance deep learning library[J]. Advances in neural information processing systems, 2019, 32.

网络首发：

标题：一种全天候道路场景下的红外和可见光图像融合方法

作者：甘文霞, 潘俊杰, 耿晶, 王慧妮, 胡小弟

收稿日期：2024-12-08

DOI:10.13203/j.whugis20240173