引文格式: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),2025,50(7):1346-1358.DOI:10.13203/j.whugis20240173

Citation: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,2025,50(7):1346-1358.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[2] 耿 晶[3] 王慧妮[2] 胡小弟[2]

1 武汉工程大学计算机科学与工程学院(人工智能学院),湖北 武汉,430205

2 武汉工程大学土木工程与建筑学院,湖北 武汉,430074

3 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与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,但受限于光照条件和恶劣天气,现有的红外和可见光图像融合方法在全天候道路应用场景中难以得到理想的效果。针对此问题,提出了一种适用于全天候道路场景的红外和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;然后,在纹理损失函数中引入Scharr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上对该方法进行了对比实验和泛化实验,结果表明,所提方法在主观视觉效果和客观指标评价上均优于其他先进方法,得到的融合图像纹理清晰且红外目标显著。因此,所提出的3种优化设计可以提高可见光和红外图像的融合效果。

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

中图分类号:P237 **文献标识码**:A **收稿日期**:2024-04-01

DOI:10.13203/j.whugis20240173 **文章编号**:1671-8860(2025)07-1346-13

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

GAN Wenxia[1] *PAN Junjie*[2] *GENG Jing*[3] *WANG Huini*[2] *HU Xiaodi*[2]

1 School of Computer Science and Engineering (School of Artificial Intelligence), Wuhan Institute of Technology, Wuhan 430205, China

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

3 School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China

Abstract: **Objectives**: The objective 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 existing infrared and visible image fusion algorithm is difficult to achieve the ideal effect. **Methods**: In response to this issue, an infrared and visible image fusion network suitable for all -weather road scenes was proposed. First, 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. Second, 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. **Results**: Comparison and generalization experiments on MFNet, RoadScene, and TNO data sets show that the proposed method is better than other ad-

基金项目:国家自然科学基金(42471388,42201461);湖北省交通厅科研项目(2023-121-3-4);湖北省教育厅科学研究计划指导性项目(B2023042)。

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、深度学习等研究。charlottegan@whu.edu.cn

通信作者:王慧妮,博士,副教授。wanghuini@wit.edu.cn

vanced methods in subjective visual effects and objective indicators. **Conclusions**: Therefore, the three optimized designs the proposed can improve the fusion effect of visible and infrared images.

Key words: 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，但描述场景细节的能力不足[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[15-17]、目标跟踪[18-20]、语义分割[21-23]等广泛采用。

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方法。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，文献[35]设计了一种具有残差块的对称编解码器网络，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，重构生成融合图像。文献[38]首次将生成对抗网络(generative adversarial network，GAN)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。文献[43]将高级视觉任务和图像融合联系，提出了一种语义感知的实时图像融合网络(a semantic-aware real-time infrared and visible image fusion network，SeAFusion)，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，复杂环境下的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，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，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，本文提出了一种适用于全天候道路场景的图像融合网络(all-weather road scenes-SeAFusion，ARSFusion)，以 SeAFusion为基础模型，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。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；在特征提取环节，设计了一种二阶梯度残差模块(second order gradient residual block，SGRB)，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。

1　ARSFusion

1.1　ARSFusion 结构

ARSFusion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SGRB组成，其中，卷积层中卷积核大小为3×3，激活函数为带泄漏的修正线性单元(leaky rectified linear unit，LReLU)，用于提取图像细粒度特征。SGRB是Resblock[44]的一个变体，由主流和

两个残差流构成,首先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为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。

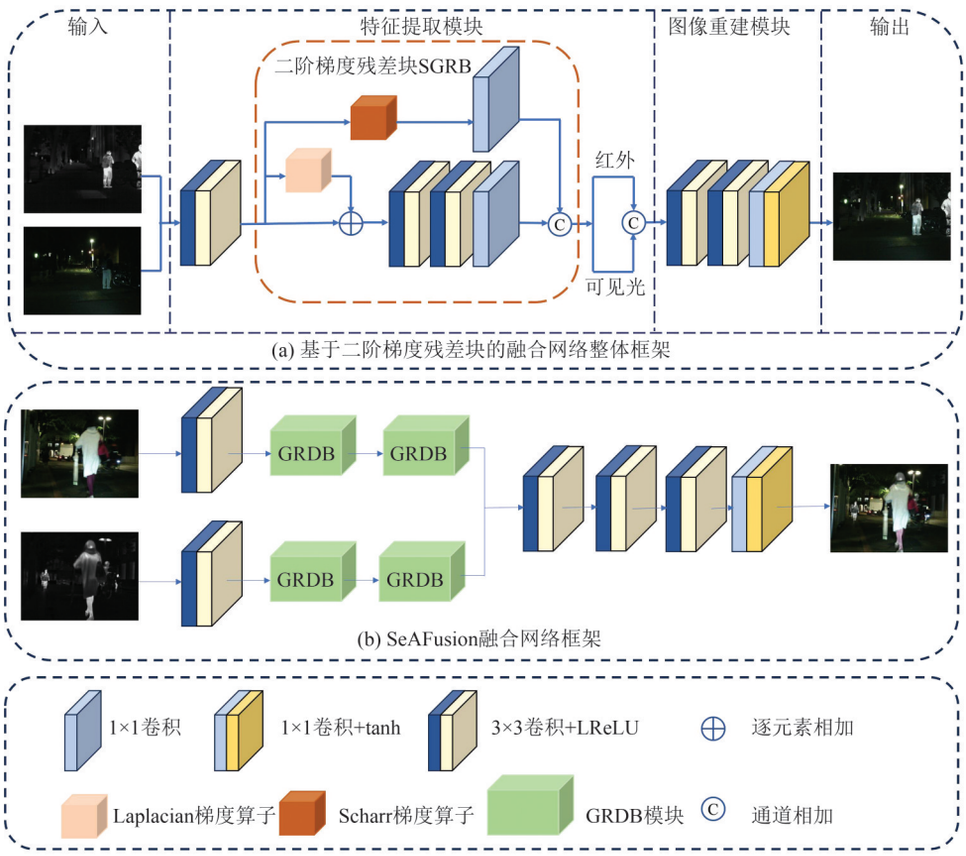


图1 图像融合网络框架

Fig. 1 Framework of Image Fusion Network

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。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,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(gradient residual dense block, GRDB)组成,图像重建器则由3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,而在图像重建模块比基础模型少了一个3×3卷积层。

1.2 损失函数

ARSfusion的损失函数L_{joint}由内容损失函数L_{content}和语义损失函数L_{semantic}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,计算公式为:

$$L_{\text{joint}}=L_{\text{content}}+\beta L_{\text{semantic}} \tag{1}$$

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L_{int}、纹理损失L_{texture}和光照损失L_{illum}3部分组成,具体计算公式为:

$$L_{\text{content}}=\lambda_1 L_{\text{illum}}+\lambda_2 L_{\text{int}}+\lambda_3 L_{\text{texture}} \tag{2}$$

光照损失L_{illum}表示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,计算公式为:

$$L_{\text{illum}}=W_{vi}L_{\text{int}}^{vi}+W_{ir}L_{\text{int}}^{ir} \tag{3}$$

$$L_{\text{int}}^{vi}=\frac{1}{HW}\left\| \boldsymbol{I}_f-\boldsymbol{I}_{vi}\right\|_1 \tag{4}$$

$$L_{\text{int}}^{ir}=\frac{1}{HW}\left\| \boldsymbol{I}_f-\boldsymbol{I}_{ir}\right\|_1 \tag{5}$$

式中,L_{int}^{vi}和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}为对应的光照权重,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;$\| \ \|_1$为L1范数;I_f为融合图像像素值矩阵;$\boldsymbol{I}_{vi}$和$\boldsymbol{I}_{ir}$分别为可见光图像和红外图像像素值矩阵。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P_n为图片是夜晚的概率,$1-P_n$为白天的概率。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L_{int}表示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,计算公式为:

$$L_{\text{int}}=\frac{1}{HW}\left\| \boldsymbol{I}_f-\max(\boldsymbol{I}_{vi},\boldsymbol{I}_{ir})\right\|_1 \tag{6}$$

式中,max()表示对两个图像相同位置的像素值进行比较取最大项。

纹理损失L_{texture}表示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,计算公式为:

$$L_{\text{texture}}=\frac{1}{HW}\left\| \left|\nabla \boldsymbol{I}_f\right|-\max\left(\left|\nabla \boldsymbol{I}_{vi}\right|,\left|\nabla \boldsymbol{I}_{ir}\right|\right)\right\|_1 \tag{7}$$

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

语义损失包括主语义损失L_{main}和辅助语义损失L_{aux}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,计算公式为:

$$L_{\text{semantic}}=L_{\text{main}}+\alpha L_{\text{aux}} \tag{8}$$

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]对融合的图像分割,得到分割结果概率矩阵$\boldsymbol{I}_s\in\mathbf{R}^{H\times W\times C}$和辅助分割结果概率矩阵$\boldsymbol{I}_{sa}\in\mathbf{R}^{H\times W\times C}$后,分别计算主语义损失和辅助语义损失,计算公式为:

$$L_{\text{main}}=\frac{-1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C}L_{so}^{(h,w,c)}\lg I_s^{(h,w,c)} \tag{9}$$

$$L_{\text{aux}}=\frac{-1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\sum_{c=1}^{C}L_{so}^{(h,w,c)}\lg I_{sa}^{(h,w,c)} \tag{10}$$

式中,$\boldsymbol{L}_{so}\in\mathbf{R}^{H\times W\times C}$表示从分割标签$\boldsymbol{L}_s\in\mathbf{R}^{H\times W\times C}$转换得到的独热向量,$L_{so}^{(h,w,c)}$为向量$\boldsymbol{L}_{so}$的元素;$I_s^{(h,w,c)}$、$I_{sa}^{(h,w,c)}$分别为$\boldsymbol{I}_s$、$\boldsymbol{I}_{sa}$的元素。

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,计算公式为:

$$\beta=\gamma\times(m-1) \tag{11}$$

式中,m表示当前迭代轮次;γ是平衡语义损失和内容损失的常数。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地指导融合网络的训练,所以β也随之增加。

2　数据处理与结果分析

2.1　数据集

实验使用武汉大学唐霖峰等基于MFNet[21]构建的数据集进行模型训练及测试。MFNet数据集包含1 569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像素,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,唐霖峰等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方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1 444对高质量且完全配准的红外和可见光图像,数据集提供了8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3个数据集上开展了实验,并与9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3种基于卷积神经网络的方法SDNet[42]、U2Fusion[41]和SeAFusion[43]。本文选取的9种对比方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(mutual information, MI)[47]、标准差(standard deviation, SD)、视觉信息保真度(visual information fidelity for fusion, VIF)[48]和Q_{abf}[49]4个评价指标进行融合结果的定量评估分析。MI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从统计学的角度反映出融合图像的灰度级分布情况,VIF基于人类视觉系统评估融合图像的信息保真度,Q_{abf}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批大小和训练轮次组合的模型实验,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,每轮迭代训练轮次为10,参数更新次数分别为2 700和20 000。使用Adam优化器更新参数,批处理大小为4,初始学习率为0.000 1,权重衰减系数为0.000 2。

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0.000 5。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128 GB。编程环境为Python和PyTorch平台[50-52]。

2.3 消融实验

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,不同改进模块组合分别命名为模块A、B、C、D。其中,Illum-loss指添加了光照损失L_{illum}的损失函数,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定性和定量比较,结果如图2和表1所示。

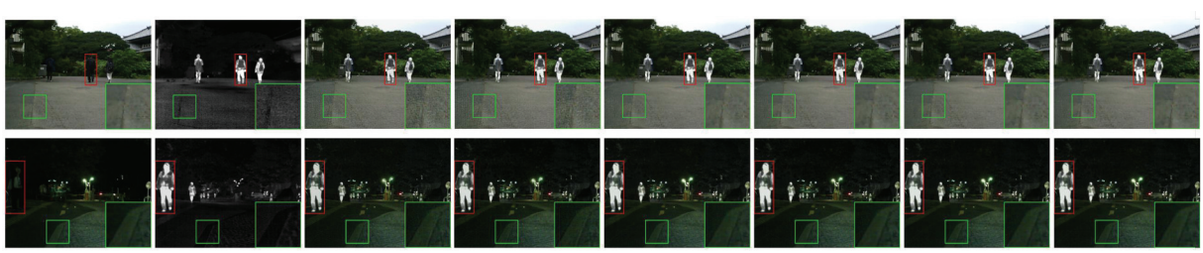


图2 消融研究可视化结果

Fig. 2 Visualized Results of Ablation Studies

表1 各模块在MFNet测试集上的消融研究定量比较

Table 1 Quantitative Comparison of Ablation Studies on Various Modules on MFNet Dataset

模型	Illum_loss	Scharr	SGRB	MI	SD	Q_{abf}	VIF
SeAFusion				3.887 4	41.268 0	0.665 7	0.940 2
模型A	√			4.145 4	42.278 4	0.700 1	0.997 3
模型B		√		4.550 7	42.215 6	0.678 1	0.974 6
模型C			√	4.894 3	42.388 7	0.682 0	1.028 9
模型D		√	√	5.008 9	42.317 4	0.683 0	1.006 3
ARSFusion	√	√	√	5.121 0	42.243 5	0.685 4	1.015 3

如图2所示,红框中为热目标信息,右下角绿框为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,且模型B和模型C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。

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比原始模型在4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,模型指标比SeAFusion有较大提升,MI提升1.233 6,SD提升0.975 5,Q_{abf}提升0.019 7,VIF提升0.075 1,在融合质量评价和视觉效果两方面都表现较好。

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ARSFusion和SeAFusion融合的图像保留了

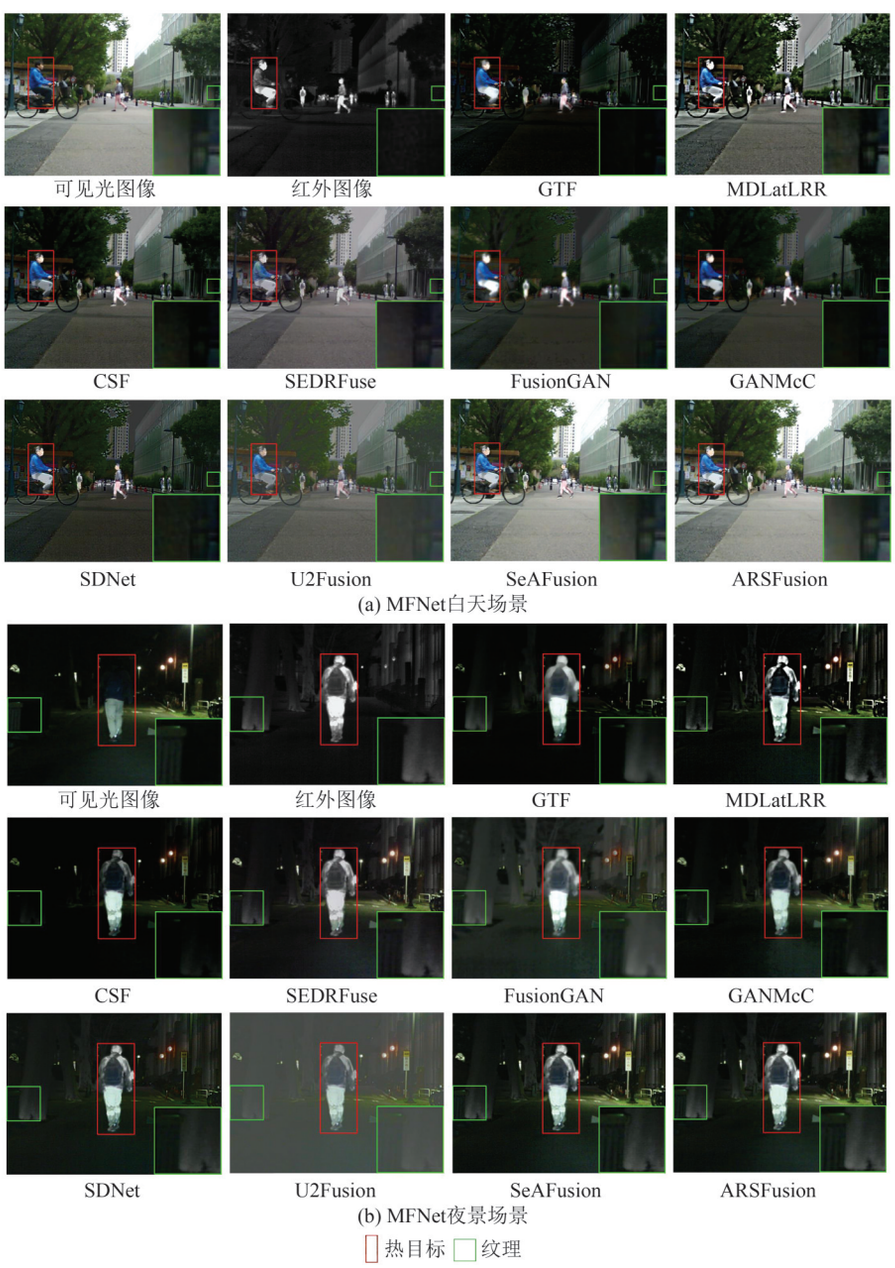


图3　MFNet数据集上ARSFusion和9种先进方法的定性比较

Fig. 3　Qualitative Comparison of the Proposed Method with 9 State-of-the-Art Methods on MFNet Dataset

纹理细节和亮度目标。

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ARSFusion和SeAFusion,其他方法均存在程度不一的背景纹理模糊情况。

图4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,ARSFusion在MI、SD、VIF和Q_{abf} 4项指标上都表现出显著的优势。最佳MI为5.121代表ARSFusion能够从源图像传递更多的信息至融合图像。最佳SD为42.243 5和VIF为1.015 3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,ARSFusion在Q_{abf}指标上也取得最优值0.685 4,这表明本文融合图像中保留了更多源图像的显著信息。

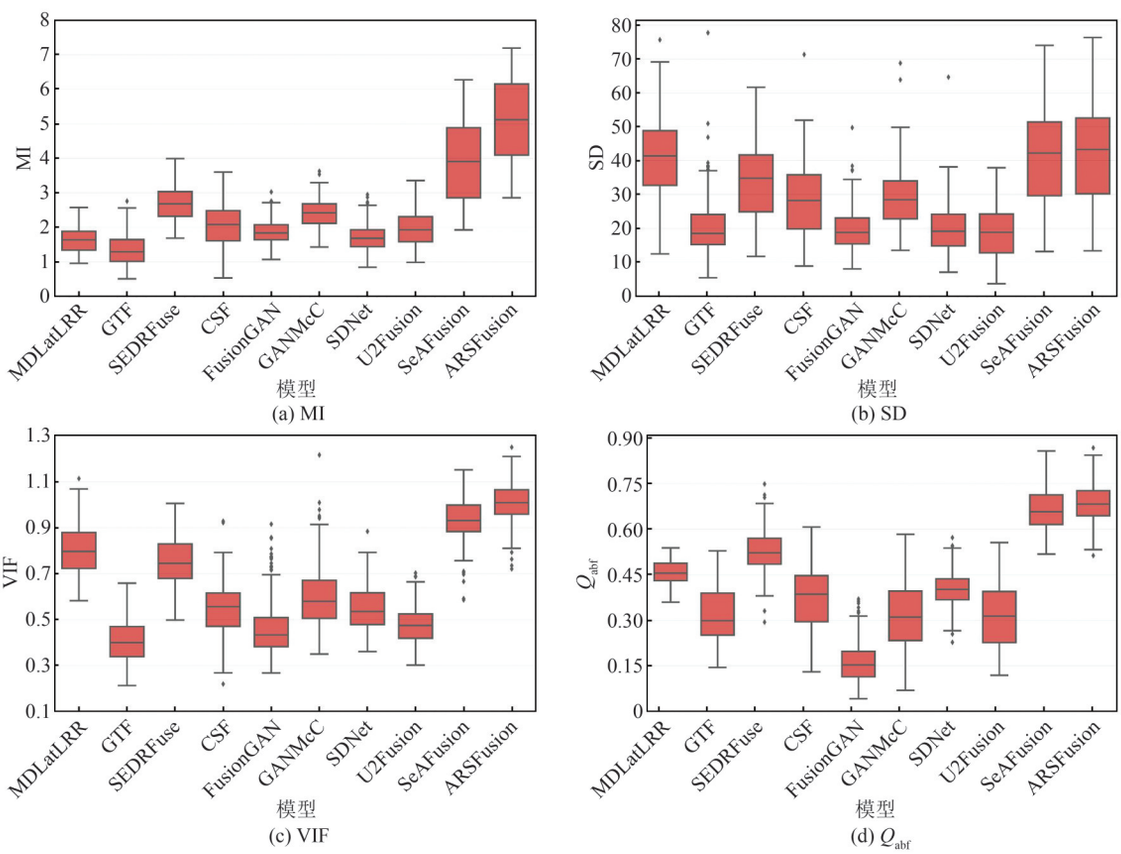


图4 MFNet数据集361对图像上不同融合方法MI、SD、VIF、Q_{abf}的定量比较

Fig. 4 Quantitative Comparison of MI, SD, VIF, Q_{abf} of Different Fusion Methods on 361 Image Pairs from MFNet Dataset

2.5 泛化实验

图5展示了在RoadScene数据集白天和强炫光场景下的两幅图像上9种对比方法和ARSFusion的融合结果。可以看出,除了ARSFusion和SeAFusion,其他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ARSFusion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展示了在TNO数据集夜晚和浓雾场景下的两幅图像上9种对比方法和ARSFusion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ARSFusion能够同时保持可见光的纹理细节和显著目标的强度。

9种对比方法和ARSFusion在RoadScene数据集和TNO数据集上的定量结果评价结果如

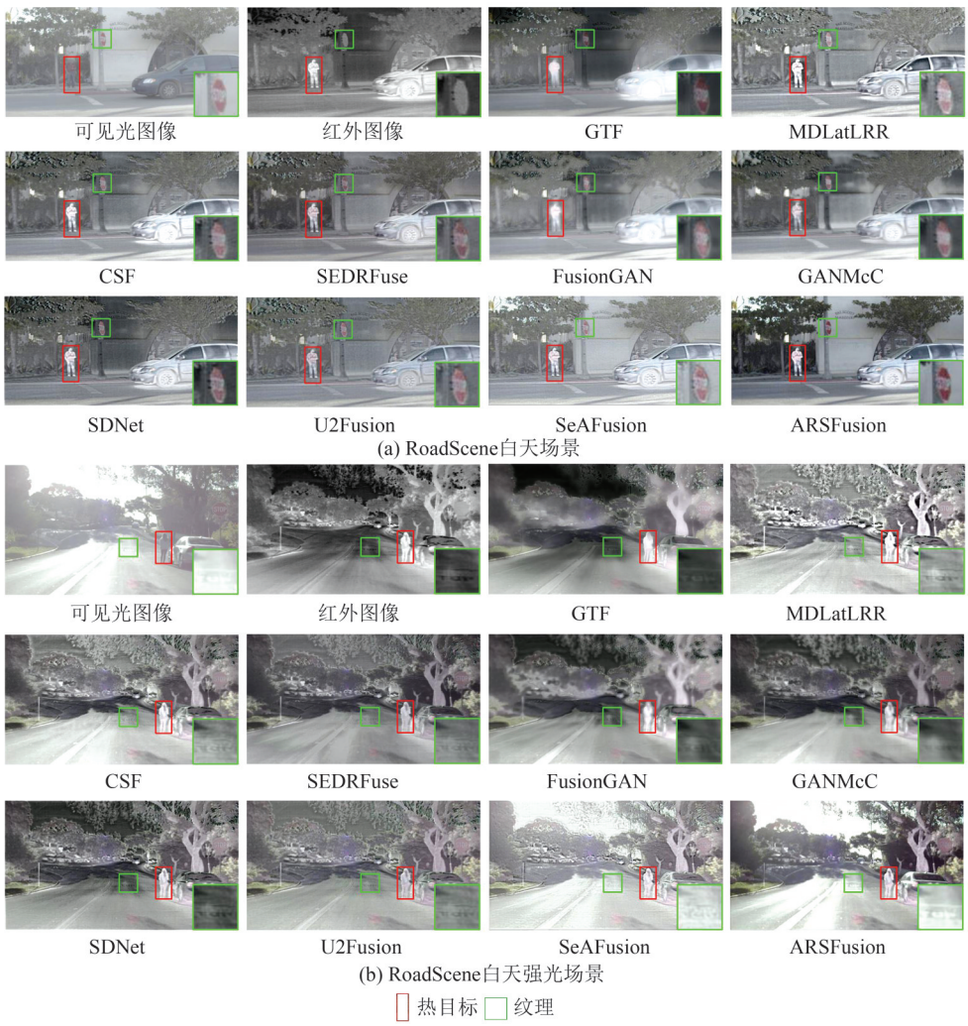


图 5　RoadScene 数据集上 ARSFusion 和 9 种先进方法的定性比较

Fig. 5　Qualitative Comparison of ARSFusion with 9 State-of-the-Art Methods on RoadScene Dataset

图 7 所示。在 Roadscene 数据集上,ARSFusion 在 MI、VIF 和 Q_{abf} 上表现出较大的优势;在 SD 上,ARSFusion 仅次于 GTF;在 TNO 数据集上,ARSFusion 在 MI、SD、VIF 和 Q_{abf} 上均表现出显著优势,表明 ARSFusion 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。

在泛化数据集上的结果表明,ARSFusion 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,原因有:(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。

2.6　效率比较

为了评估不同方法的融合效率,本文对不同方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,深度学习方法的运行效率明显更高,其中 ARSFusion 在 RoadScene、TNO 两个数据集上的运行时间分别为 0.132 9 s 和 0.163 5 s,优于其他方法,在 MFNet 数据上为 0.195 9 s,仅次于 FusionGAN 和 SDNet。

3　结　语

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 的基础上进行了以下工作:

1)基于 Resblock 设计了二阶梯度残差模块,利用 Scharr 和拉普拉斯两种梯度算子结合形成两组残差流对图像纹理特性信息进行补充,从而提高融合网络对图像强弱纹理特征的提取能力;

2)引入光照判别损失,在判断图像处于白天

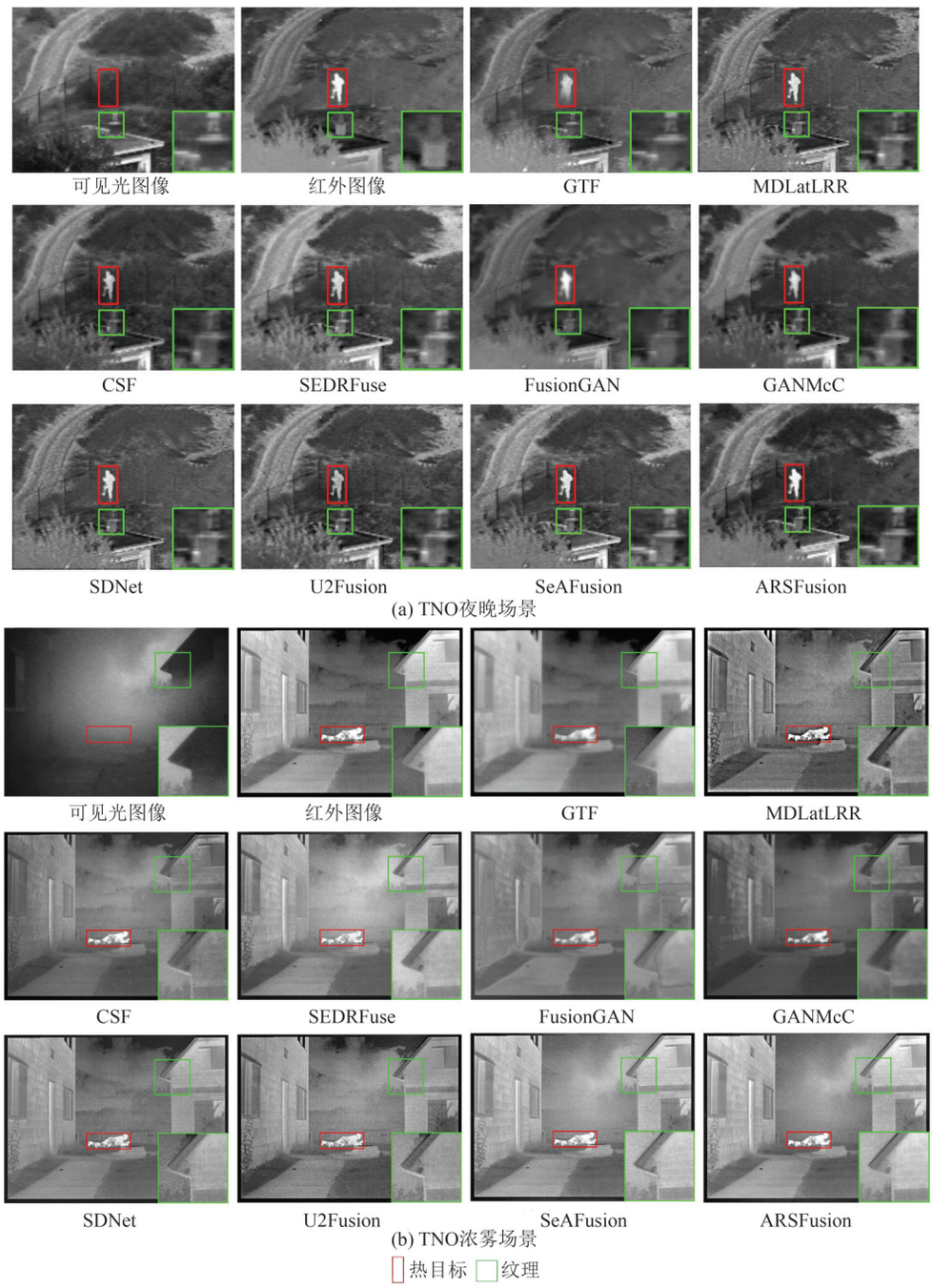


图6 TNO数据集上ARSFusion和9种先进方法的定性比较

Fig. 6 Qualitative Comparison of ARSFusion with 9 State-of-the-Art Methods on TNO Dataset

或者夜晚后,有选择性地从可见光图像和红外图像中提取有用的信息,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;

3)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热红外信息的图像。

实验结果表明,ARSFusion主观效果优于其他主流方法,视觉上保留了更显著的红外高亮目标信息以及更清晰的细节纹理,在MI、SD、Q_{abf}和VIF 4项指标上远优于MDLatLRR、GTF、SEDRFuse、CSF、FusionGAN、GANMcC、SDNet、U2Fusion、SeAFusion这9种主流模型,较SeAFusion模型在MFNet数据集上的4项指标分别提升约31.73%、2.36%、2.96%和7.98%,且在3种数据集上单对图像的平均融合速度分别为0.195 9 s、0.132 9 s和0.163 5 s,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。

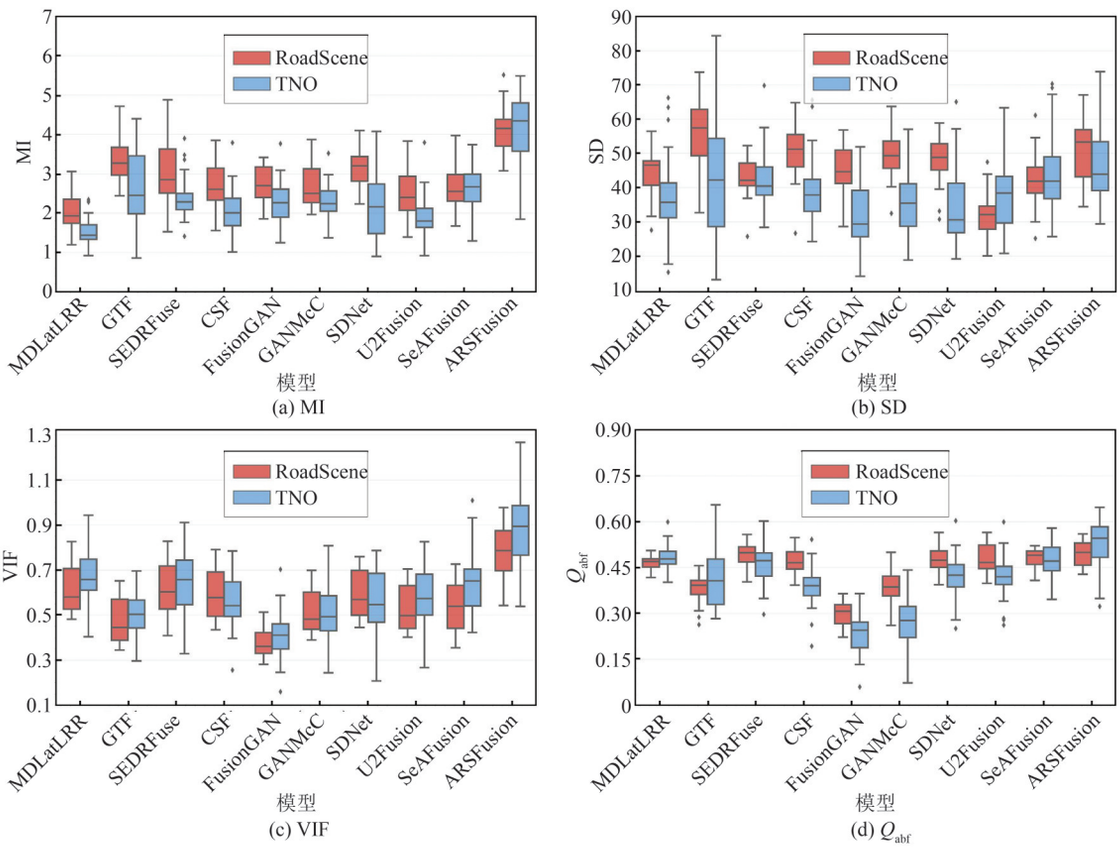


图 7　RoadScene 数据集和 TNO 数据集上不同融合方法 MI、SD、VIF、Q_{abf} 的定量对比

Fig. 7　Quantitative Comparison of MI, SD, VIF, Q_{abf} of Different Fusion Methods on RoadScene Dataset and TNO Dataset

表 2　不同方法在不同数据集上的运行效率/s

Table 2　Running Time of All Methods on All Datasets/s

融合方法	MFNet	RoadScene	TNO
MDLatLRR	153.745 2	53.329 2	103.122 7
GTF	5.633 9	1.797 4	3.073 9
SEDRFuse	10.158 3	5.684 3	9.732 4
CSF	2.876 0	2.548 1	3.167 0
FusionGAN	0.086 3	0.473 5	0.317 5
GANMcC	0.149 9	0.317 5	0.516 5
SDNet	0.024 2	0.204 2	1.223 5
U2Fusion	0.880 2	0.430 9	0.742 4
SeAFusion	0.208 3	0.141 4	0.172 9
ARSFusion	0.195 9	0.132 9	0.163 5

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]. 武汉大学学报(信息科学版)，2024，49(4)：506-515.
YING Shen, JIANG Yuewen, GU Jiangyan, et al. High Definition Map Model for Autonomous Driving and Key Technologies[J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(4): 506-515.

[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]. 重庆理工大学学报(自然科学)，2023，37(11)：11-19.
SHEN Caiying, ZHU Siyao, HUANG Xingchi. Research on Intelligent Vehicle Target Detection Algorithm Based on Binocular Vision[J]. *Journal of Chongqing University of Technology (Natural Science)*, 2023, 37(11): 11-19.

[3] 胡淼，姜麟，陶友凤，等. 改进 YOLOv7 的自动驾驶目标检测算法[J]. 计算机工程与应用，2024，60(11)：165-172.
HU Miao, JIANG Lin, TAO Youfeng, et al. Improved Automatic Driving Target Detection Algorithm of YOLOv7[J]. *Computer Engineering and Applications*, 2024, 60(11): 165-172.

[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(1)：139-144.
YE Yutong, LI Bijun, FU Liming. Fast Object Detection and Tracking in Laser Data for Autonomous Driving[J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(1): 139-144.

[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 Y, MA Y, LI C. Infrared and Visible Image Fusion Methods and Applications: A Survey[J]. *Information Fusion*, 2019, 45: 153-178.

[7] ZHANG X C, DEMIRIS Y. Visible and Infrared Image Fusion Using Deep Learning [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023, 45(8): 10535-10554.

[8] MA W H, WANG K, LI J W, 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]. 计算机应用,2023,43(11): 3574-3578.
HAN Linkai, YAO Jiangwei, WANG Kunfeng. Fusion of Visible and Infrared Images Preserving Gradient and Contour[J]. *Journal of Computer Applications*, 2023,43(11):3574-3578.

[10] HAO S,HE T,AN B Y,et al. VDFEFuse: A Novel Fusion Approach to Infrared and Visible Images[J]. *Infrared Physics & Technology*, 2022, 121: 104048.

[11] SUN C Q, ZHANG C, XIONG N X. 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(3):266-275.
CHEN Yanlin, WANG Zhishe, SHAO Wenyu, et al. Multi-scale Transformer Fusion Method for Infrared and Visible Images[J]. *Infrared Technology*, 2023, 45(3): 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(7): 911-917.
XIANG Tianzhu, GAO Rongrong, YAN Li, et al. Region Feature Based Multi-Scale Fusion Method for Thermal Infrared and Visible Images[J]. *Geomatics and Information Science of Wuhan University*, 2017, 42(7): 911-917.

[14] LI S T, KANG X D, FANG L Y, et al. Pixel-Level Image Fusion: A Survey of the State of the Art[J]. *Information Fusion*, 2017, 33: 100-112.

[15] LIU J Y, FAN X, HUANG Z B, et al. Target-Aware Dual Adversarial Learning and a Multi-Scenario Multi-modality Benchmark to Fuse Infrared and Visible for Object Detection[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 2022.

[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]. 图学学报, 2024, 45(1): 35-46.
YUAN Chao, ZHAO Yadong, ZHANG Yao, et al. Multimodal Pedestrian Detection Algorithm Based on YOLO Lightweight[J]. *Journal of Graphics*, 2024, 45(1): 35-46.

[17] 孙文财,胡旭歌,杨志发,等. 融合GPNet与图像多尺度特性的红外-可见光道路目标检测优化方法[J]. 吉林大学学报(工学版),2024,54(10):2799-2806.
SUN Wencai, HU Xuge, YANG Zhifa, et al. Optimization of Infrared-Visible Road Target Detection by Fusing GPNet and Image Multiscale Features [J]. *Journal of Jilin University (Engineering and Technology Edition)*, 2024, 54(10): 2799-2806.

[18] GENG X W, LI M L, LIU W P, 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]. 计算机工程与应用, 2024, 60(11): 84-94.
LIU Wanjun, LIANG Linlin, QU Haicheng. Multimodal Target Tracking Algorithm Using Transformer [J]. *Computer Engineering and Applications*, 2024, 60(11): 84-94.

[20] LIU W C, LIU W B, SUN Y X. Visible-Infrared Dual-Sensor Fusion for Single-Object Tracking[J]. *IEEE Sensors Journal*, 2023, 23(4): 4118-4128.

[21] HA Q S, WATANABE K, KARASAWA T, et al. MFNet: Towards Real-Time Semantic Segmentation for Autonomous Vehicles with Multi-spectral Scenes[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 2017.

[22] HOU J L,ZHANG D Z,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(4):543-551.
ZHU Hao, GU Xiaojing, LAN Xin, et al. RGB IR Semantic Segmentation Based Onmulti Scale Edge Enhancement[J]. *Laser & Infrared*, 2022, 52(4): 543-551.

[24] WANG Z S, XU J W, JIANG X L, et al. Infrared and Visible Image Fusion via Hybrid Decomposition of NSCT and Morphological Sequential Toggle Ope-

rator[J]. *Optik*, 2020, 201: 163497.

[25] CHEN J, LI X J, LUO L B, et al. Infrared and Visible Image Fusion Based on Target-Enhanced Multiscale Transform Decomposition[J]. *Information Sciences*, 2020, 508: 64-78.

[26] LI H, WU X J, KITTLER J. MDLatLRR: A Novel Decomposition Method for Infrared and Visible Image Fusion[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4733-4746.

[27] 赵庆典，杨德宏. 基于图像增强和二次非下采样轮廓波变换的红外与可见光图像融合[J]. 激光与光电子学进展，2024，61(4)：0437001.
ZHAO Qingdian, YANG Dehong. Infrared and Visible Light Image Fusion Based on Image Enhancement and Secondary Nonsubsampled Contourlet Transform [J]. *Laser & Optoelectronics Progress*, 2024, 61(4): 0437001.

[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(9)：915-924.
WU Lingxiao, KANG Jiayin, JI Yunxiang. Infrared and Visible Image Fusion Based on Guided Filter and Sparse Representation in NSST Domain[J]. *Infrared Technology*, 2023, 45(9): 915-924.

[29] LIU Y, CHEN X, WARD R K, et al. Image Fusion with Convolutional Sparse Representation [J]. *IEEE Signal Processing Letters*, 2016, 23(12): 1882-1886.

[30] LIU Y D, YANG X M, ZHANG R Z, 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]//The 20th International Conference on Information Fusion (Fusion), Xi'an, China, 2017.

[33] 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.

[34] MA J L, ZHOU Z Q, WANG B, et al. Infrared and Visible Image Fusion Based on Visual Saliency Map and Weighted Least Square Optimization [J]. *Infrared Physics & Technology*, 2017, 82: 8-17.

[35] JIAN L H, YANG X M, LIU Z, et al. SEDRFuse: A Symmetric Encoder-Decoder with Residual Block Network for Infrared and Visible Image Fusion [J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 70: 5002215.

[36] XU H, ZHANG H, MA J Y. Classification Saliency-Based Rule for Visible and Infrared Image Fusion [J]. *IEEE Transactions on Computational Imaging*, 2021, 7: 824-836.

[37] LIU J Y, 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 Y, YU W, LIANG P W, et al. FusionGAN: A Generative Adversarial Network for Infrared and Visible Image Fusion [J]. *Information Fusion*, 2019, 48: 11-26.

[39] YANG Y, LIU J X, HUANG S Y, 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 Y, ZHANG H, SHAO Z F, 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: 5005014.

[41] XU H, MA J Y, JIANG J J, et al. U2Fusion: A Unified Unsupervised Image Fusion Network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, 44(1): 502-518.

[42] ZHANG H, MA J Y. SDNet: A Versatile Squeeze-and-Decomposition Network for Real-Time Image Fusion [J]. *International Journal of Computer Vision*, 2021, 129(10): 2761-2785.

[43] TANG L F, YUAN J T, MA J Y. 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 Y, TANG L F, XU M L, et al. STDFusionNet: An Infrared and Visible Image Fusion Network Based on Salient Target Detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, 70: 5009513.

[45] PENG C L, 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] Alexander T. TNO Image Fusion Dataset [EB/

OL]. (2014-07-08)[2024-02-01]. https://doi.org/10.6084/m9.figshare.1008029.v1.

[47] QU G H, ZHANG D L, YAN P F. Information Measure for Performance of Image Fusion[J]. *Electronics Letters*, 2002, 38(7): 313.

[48] HAN Y, CAI Y Z, 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]//International Conference on Image Processing, Barcelona, Spain, 2003.

[50] Paszke A, Gross S, Massa F, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library [EB/OL]. (2019-07-07)[2024-02-01]. https://doi.org/10.48550/arXiv.1912.01703.

[51] GENG J, GAN W X, XU J J, et al. Support Vector Machine Regression (SVR)-Based Nonlinear Modeling of Radiometric Transforming Relation for the Coarse-Resolution Data-Referenced Relative Radiometric Normalization (RRN)[J]. *Geo-spatial Information Science*, 2020, 23(3): 12.

[52] GAN W X, Albanwan H, Qin R. Radiometric Normalization of Multi-temporal Landsat and Sentinel-2 Images Using a Reference MODIS Product Through Spatiotemporal Filtering[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2021, 99: 1.