http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0077

基于嵌套连接的多尺度红外与可见光图像融合

吴楠，穆成坡*，贺杨，刘太晔
(北京理工大学 机电学院，北京 100080)

摘 要：为解决红外与可见光图像融合方法中依赖于人工设计的融合策略、运行时间过长以及无法提取多尺度深度特征的问题，提出了一种实时的基于嵌套连接的多尺度图像融合模型。特征提取器提取多尺度深层次特征，通过融合网络生成具有多尺度深度特征的特征图像，通过图像重构器得到融合图像。在公开数据集上与其他算法进行对比，在主观定性评价中，该算法在保持图像强度的同时能够包含锐利的图像边缘，在过度曝光、目标遮挡、细节模糊等复杂情况下具有更好的融合效果。客观指标对比中，在基于信息论、基于图像特征、基于图像结构相似性和人类感知度量4类9个指标上取得了5个第一2个第二的结果，另外2个指标的结果也处于较高水平，且融合时间大幅缩短。实验证明，该模型能够有效解决当前图像融合方法中存在的问题，具有较高的实用性。

关 键 词：图像融合；深度学习；嵌套连接；多尺度；实时
中图分类号：V231.1；TP391.4
文献标志码：A **文章编号**：1001-5965(2025)02-0683-09

由于技术上的限制，单一模态的传感器只能从一个方面获取信息，无法提供全面或必需的信息，但某些复杂应用的出现需要提供某些情况的全面信息，以便加深对各种场景的理解。而图像融合技术的目标在于，把非同类传感器所获得的数据收集起来，通过某种算法结合以生成更具信息量的图像[1]。可见光、红外线、核磁共振成像等都是可以用于融合的源图像，在这一些源图像的融合当中[2]，红外和可见光融合的应用相对更加广泛[3]，相较于其他传感方式所获得的信息，可见光和红外图像融合在许多方面的性能都更加优越[4]。

现有的融合算法可以分为两类，一类是传统算法，包括多尺度变换融合算法、稀疏表示融合算法、基于子空间方法和其他算法等[5]，该类方法需要通过人工设计的数学变换方法来进行特征提取，使得融合方法日益复杂，并且只能针对特定任务进行设计以适应不同的融合情况，限制了融合的性能。另一类是基于深度学习的算法，该类方法能够很好地弥补传统方法的不足，借助神经网络独特的特征提取表示能力，从而能够非线性拟合模型，根据网络架构的不同主要分为3类[6]，基于自动编码器(auto-encoder, AE)的方法、基于生成对抗网络(generative adversarial networks, GAN)的方法和基于卷积神经网络(convolutional neural networks, CNN)的方法[7]。基于AE的方法需要先在大规模数据集上训练自动编码器作为编码层和解码层，然后利用编码层提取的图像特征训练融合层，经处理后输入解码层从而获得最终图像，但融合方法中使用了传统手工制作的特定融合规则[8]来合并特征，训练策略不可学习。基于GAN的方法在图像融合过程引入生成器和鉴别器的对抗，通过二者之间的博弈使得判别器难以判别生成器所生成的融合图像是否真实，融合过程逐渐分别保留住可见光图的梯度信息以及红外图的热辐射信息，但该方法生成的图像清晰度较低，细节纹理信息容易丢失且目标边缘信息模糊，视觉保真度仍有待提高[9]。基于CNN的融

收稿日期：2023-02-25；录用日期：2023-03-31；网络出版时间：2023-04-19 17:03
网络出版地址：link.cnki.net/urlid/11.2625.V.20230419.1544.001
*通信作者. E-mail：muchengpo@bit.edu.cn

引用格式：吴楠，穆成坡，贺杨，等. 基于嵌套连接的多尺度红外与可见光图像融合[J]. 北京航空航天大学学报，2025，51(2)：683-691.
WU N，MU C P，HE Y，et al. Multi-scale infrared and visible image fusion based on nest connection[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(2)：683-691 (in Chinese).

合方法依靠优越的网络架构和损失函数来确保融合性能，但在设计网络架构和损失函数时容易对两种源图像产生偏向，从而导致特征丢失的问题存在[10-12]。

融合方法能够生成让人满意的融合图像，但是上述方法各自还存在一些值得改进的地方。总体上存在一些普遍问题：1）现有的融合算法倾向于追求更好的视觉效果和更高的评价指标，网络层次越来越深、融合方法越来越复杂，从而导致了融合的速度越来越慢，在方法的实用性和时效性上考虑的较少；2）在深度特征提取完成后，许多算法依赖于人工设计的融合策略，这导致了模型无法通用其他融合任务；3）因两种源图成像方式的不同，基本都是采用孪生网络分别对红外图像和可见光图像进行特征提取，融合框架大致如图1所示，对图像的特征提取存在偏向，存在特征丢失的问题；4）大部分网络没有采用下采样算子，因此无法提取多尺度特征，图像的深度特征没有得到充分利用。

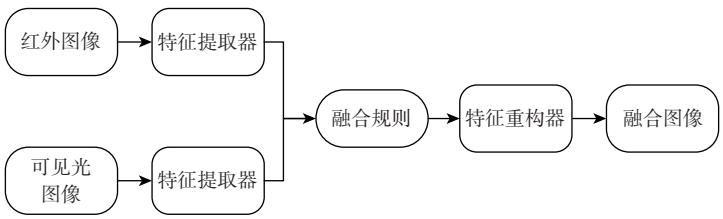


图1 算法融合框架

Fig. 1 The architecture of algorithm Fusion

针对上述问题，本文提出一种基于嵌套连接的多尺度实时图像融合模型。先将可见光图像从RGB空间通过转换为等效的YCbCr空间[13]，提取Y通道作为可见光源图，经特征提取、融合、重构后得到最终的融合图像。

1 基于嵌套连接的多尺度实时图像融合模型

1.1 相关工作

主流的融合框架主要有3个流程[14]，如图1所示。首先以孪生网络分别提取出两种源图像特征，采取某种融合策略将特征融合，最后对特征图重构输出最终融合结果。如何利用损失函数更好地对两种源图像分别提取特征和采用何种融合策略为关键的两个因素，通过修改这两个过程可以针对不同的任务进行相应的设计，并且针对这两个过程提出了大量的方法。对于如何提取图像特征，主流的算法基本都是利用恰当的损失函数，采用孪生网络将红外和可见光图像分别分解为包含目标信息和环境背景信息的特征图，对前者着重于显著目标而对环境背景信息有所忽略，对后者着重于图像纹理细节而对目标特征信息有所忽略，因此在出现两种源图像包含信息与应提取的信息特征不同的情况时，会导致融合图像不能更多地将两张源图像信息包含在内的情况出现。对于采用何种融合策略，常用的规则有最大值、最小值、加法、L1-范数等，然而，这些手动设计的融合规则的选择也产生了不同的融合性能。

由论文[15]针对医学图像分割任务提出了嵌套连接结构，其中使用的跳跃连接用于保留来自前几层的信息。在网络架构中如果使用远距离的跳跃连接，由于存在语义差距，则会对后层产生意想不到的结果，对此提出了一种混合上采样和短跳连接的结构来避免长跳连接可能带来的不良后果，采用这种嵌套连接结构可以有效约束语义间隙的影响，从而能够保留更多的信息最终获得更好的分割结果。在此基础之上，这种体系结构被NestFuse[8]引入。

在融合处理过程中引入下采样操作，增大感受野的同时增加对图像一些小扰动的鲁棒性，且能有效减少模型的运算量，尺寸大的特征图保留小目标特征，尺寸小的特征图保留深度特征信息，这种结构能够产生并有效利用图像的多尺度特征，但融合策略仍然是需要手工设计，在一定程度上限制了融合性能，并且运行速度仍然达不到实时性的要求，图像整体亮度偏低[7]。受此启发，在满足提取并融合多尺度图像特征的要求之上，对该网络结构的编码器进行优化，将解码器优化成特征融合器，并另外增加图像重构器进行图像重构，提出了一种新的基于嵌套连接的多尺度实时图像融合模型。

1.2 网络结构

基于嵌套连接的多尺度实时图像融合模型网络框架如图2所示，模型由3个部分组成：特征提取器、特征融合器和图像重构器。

特征提取器包含1个卷积层和3个下采样卷积（down-sampling convolutional block，DPCB）模块。卷积层的卷积核大小为3×3，激活函数为LReLU，用于提取浅层特征。紧接着3个DPCB模块[8]，用于提取多尺度深度特征，由1个最大池化层

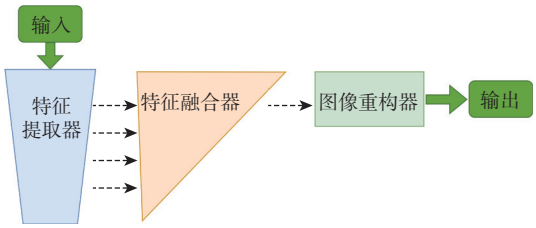


图 2　网络框架

Fig. 2　Network framework

(Pooling)和 2 个卷积层(3×3 Conv)组成。在 NestFuse 中的 4 个 ECB 模块包含两个卷积层, 第 1 次卷积操作通道数降为输入的一半, 第 2 次卷积操作通道数大幅提升为设置的输出通道, 如第 1 次编码操作的第一次卷积通道数由 16 变为 8, 第 2 次卷积通道数由 8 变为 64, 此过程会导致部分特征信息丢失的同时提升运算量。结合图像融合任务对网络层次的深度需求[16], 将卷积通道数逐步提升最高至 64, 另外在图像融合任务中, 考虑到信息丢失这一灾难性问题, 因此只采用了 3 个包含下采样操作的 DPCB 模块。在特征提取过程中, 将源图像输入特征提取器,通过损失函数和融合策略对融合过程进行约束, 避免融合规则影响和对源图像信息提取偏差。通过特征提取器后, 浅层特征保留更多的细节信息, 再通过 DPCB 中的最大池化操作, 可以增大感受野。从源图像中提取多尺度深度特征, 不仅能够有效减少运算量, 还能增加对图像噪声干扰的影响, 从而能在多个尺度上将提取的多模态深度特征进行融合。

特征融合器为一个基于嵌套连接的网络, 将源图像经过特征提取器后提取出的多尺度特征在每一行中通过短跳连接的方式级联起来输入下一个卷积层, 保证每一个层级的特征都能被引入模型之中。在跨层链路中通过上采样卷积并通过 1×1 的卷积减少通道数后连接特征提取器中的多尺度深度特征, 把模型中采集到的抽象特征还原到原尺寸大小, 最终输出具有多尺度深度特征的特征图实现特征聚合。

图像重构器接收特征融合器输出的多尺度特征图像, 聚合的特征进一步重构实现图像重建。图像重构器由串联的 3 个 3×3 卷积层和一个 1×1 卷积层组成。3×3 卷积层都采用 LReLU 作为激活函数, 而 1×1 卷积层的激活函数是 Tanh。与其他方法不同, 本文单独采用了图像重构器进行图像重建, 通过渐进式减小特征图通道数减少了特征的丢失, 将具有多尺度特征的图像内容整合进行还原输出。

如图 3 所示, 左侧蓝色虚线框内为特征提取器, 中间黄色虚线框内为特征融合器, 右边绿色虚线框内为图像重构器。融合网络设置的卷积核大小、DPCB 模块参数、通道数和激活函数等参数如表 1 所示。

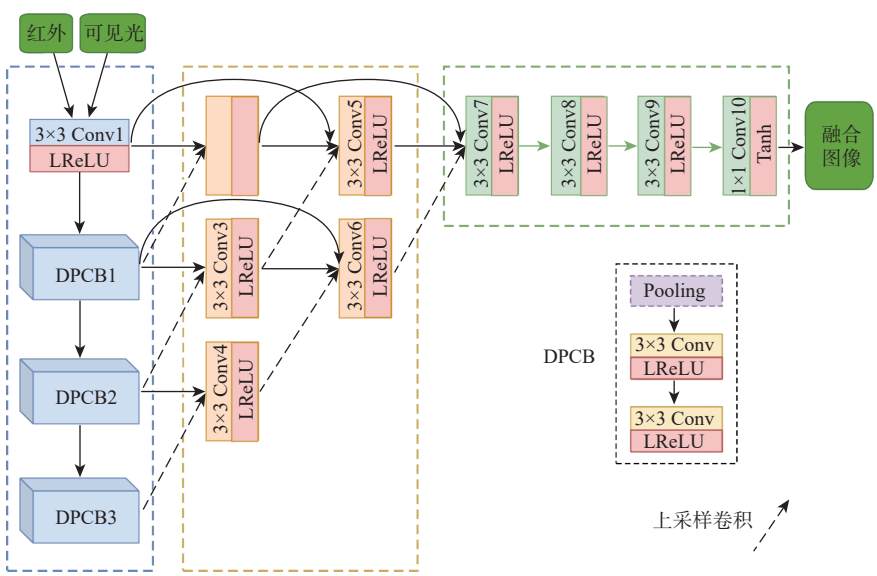


图 3　基于嵌套连接的多尺度实时图像融合模型网络结构

Fig. 3　The architecture of multi-scale real-time infrared and visible image fusion based on Nest Connection

1.3　损失函数

为了促进融合模型整合更多有意义的信息, 提高视觉质量和量化指标, 设计了 3 个方面的损失函数, 即强度损失、纹理损失[16]、相似性损失[17], 表示如下:

$$L = L_{\text{int}} + \alpha L_{\text{texture}} + \beta L_{\text{ssim}} \tag{1}$$

其中: α和β为平衡参数; L_{int}为强度损失, 约束融合图像的整体表观强度; L_{texture}为纹理损失, 促进融合

表1 网络结构参数

Table 1 Network structure parameters

模块	层	卷积核	输入	输出	步长	填充	激活
特征提取器	Conv1	3×3	2	16	1	1	leaky ReLU
	DPCB1		16	32			
	DPCB2		32	48			
	DPCB3		48	64			
特征融合器	Conv2	3×3	32	16	1	1	leaky ReLU
	Conv3	3×3	64	32	1	1	leaky ReLU
	Conv4	3×3	96	48	1	1	leaky ReLU
	Conv5	3×3	48	32	1	1	leaky ReLU
	Conv6	3×3	96	32	1	1	leaky ReLU
图像重构器	Conv7	3×3	64	48	1	1	leaky ReLU
	Conv8	3×3	48	32	1	1	leaky ReLU
	Conv9	3×3	32	16	1	1	leaky ReLU
	Conv10	1×1	16	1	1	0	Tanh
DPCB	Max-pooling						
	Conv	3×3	N_{in}	N_{in}	1	1	leaky ReLU
	Conv	3×3	N_{in}	N_{out}	1	1	leaky ReLU
上采样卷积	UpSamping						
	Conv	1×1	N	N-16	1	0	

图像包含更细粒度的纹理细节；L_{ssim}为相似性损失，提高源图像和融合图像的相似度。

强度损失函数在像素级测量，减少融合图像和源图像之间的强度差异，定义为

$$L_{int}=\frac{1}{HW}\|I_o-\max(I_{vi},I_{ir})\|_1 \tag{2}$$

式中，H和W为图像高和宽；$\|\ \|_1$为L1范数；I_{vi}和I_{ir}分别为可见光和红外图像；I_o为融合后的输出图像。强度损失为模型学习提供了粗粒度的分布约束，为更好地保留源图像的纹理信息与强度损失互补，因此引入纹理损失来保持住更细粒度的纹理信息：

$$L_{texture}=\frac{1}{HW}\|\|\nabla I_o|-\max(|\nabla I_{vi}|,|\nabla I_{ir}|)\|_1 \tag{3}$$

其中，∇为索贝尔算子，检测图像的边缘信息，用于计算图像的细粒度纹理信息；$|\ |$为求取绝对值，将源图像的最大纹理定义为图像的最佳纹理。另外，通过引入结构相似性损失为强度损失和纹理损失补充在对比度和结构等方面的相似性：

$$L_{ssim}=1-\max(\mathrm{SSIM}(I_{vi},I_o),\mathrm{SSIM}(I_{ir},I_o)) \tag{4}$$

其中，SSIM为结构相似性度量值。

在强度损失、纹理损失和结构相似性损失函数的指导下，模型能够同时获得最佳的亮度分布和丰富的细节信息，不仅可以提高融合图像的视觉质量，还可以提高图像的统计学质量，使融合效果达到最优。

1.4 训练策略

采用MSRS数据集，该数据集包含了1 569幅图片（820幅在白天拍摄，749幅在夜间拍摄）并已经进行了预处理，包括了道路、建筑、树木、车辆以及路人等场景，从中挑选1 000张图片进行训练，在训练过程中先将可见光图像转为YCbCr通道，Y通道作为可见光与红外图像输入进行训练。

训练在NVIDIA GeForce GTX 1080Ti和Intel(R) Core(TM) i5-6500 CPU @ 3.20GHz硬件上使用Pytorch深度学习框架进行。训练参数设置，batch size设为4，迭代20次，初始学习率设置为0.001，在每次迭代后更新为上一次的0.75，纹理损失的α设置为10，相似性损失的β设置为1，另使用Adam优化器进行优化。

2 结果与分析

为了能够全面评估模型，将本文模型与当前8种先进的方法进行定性和定量比较，其中包括1种传统方法GTF[18]，1种基于AE的方法NestFuse[8]，1种基于GAN的方法FusionGAN[9]，5种基于CNN的方法SeAFusion[16]、PIAFusion[12]、STDFusionNet[10]、DIDFuse[11]、U2Fusion[14]，除了GTF为2016年和FusionGAN为2019年，其余算法皆为2020年之后的新算法。

2.1 主观评价

为直观地对各种算法进行比较，并确保对比的公平性，从RoadScene、TNO、VIFB中各选择1张图片进行主观对比，所选图片为近年论文所对比的图片，融合结果如图4～图6所示。

图4显示了来自VIFB图像对[19]的融合结果，为了直观地展示融合图像，绿色框中放大了汽车周围的行人，红色框为具有丰富纹理细节的区域。过度曝光情况下，可见光图片中汽车周围的人不可见，在绿框中可以观察到，PIAFusion、NestFuse、STDFusionNet产生的融合图像中，汽车周围的人仍然不可见或不清晰，显著目标弱化，DIDFuse、SeAFusion人物纹理遭受热辐射污染，FusionGAN、GTF模糊了行人边缘轮廓，不能够保持目标的锐利边缘。在红框区域中，DIDFuse、FusionGAN、GTF、NestFuse的融合结果显示未能融入纹理信息；SeAFusion、STDFusionNet的纹理信息不够丰富，细节信息不够清晰。本文融合模型不仅能够有效地突出目标，并且在保持图像强度的同时还能包含锐利的图像边缘信息。

图5显示了来自RoadScene图像对[10]的融合结果。在红框中可以注意到FusionGAN和GTF两种

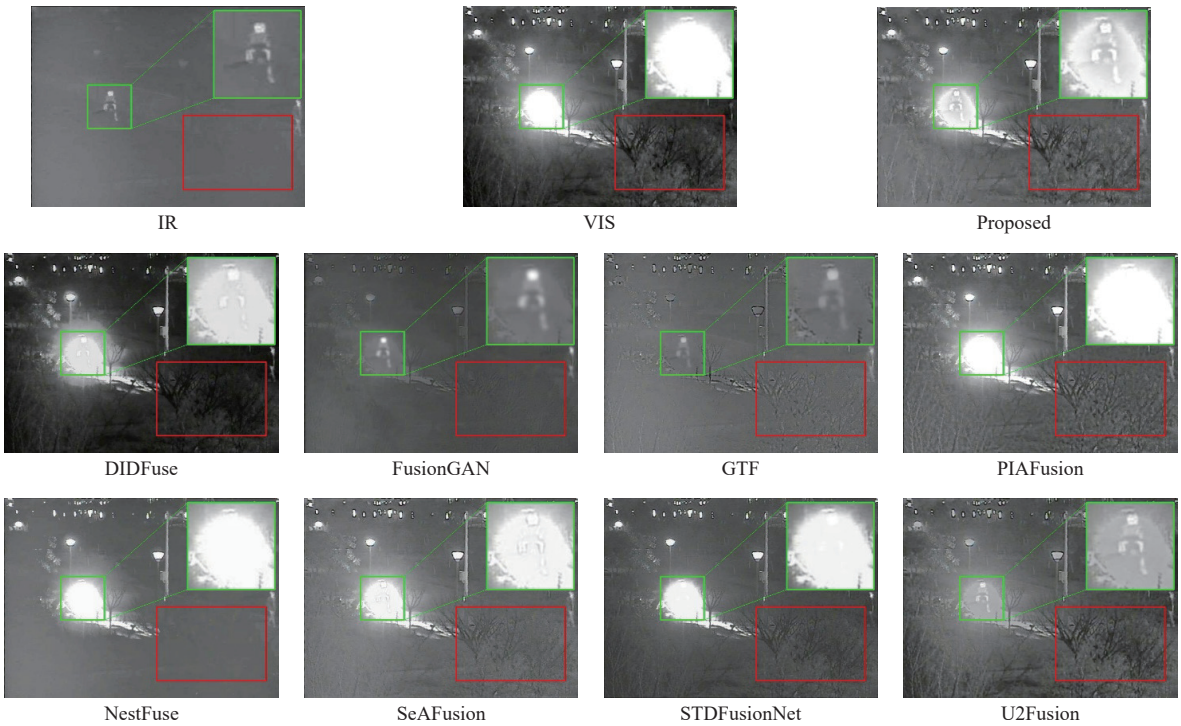


图 4　来自 VIFB 图像对的融合结果

Fig. 4　Fusion results from VIFB image pairs

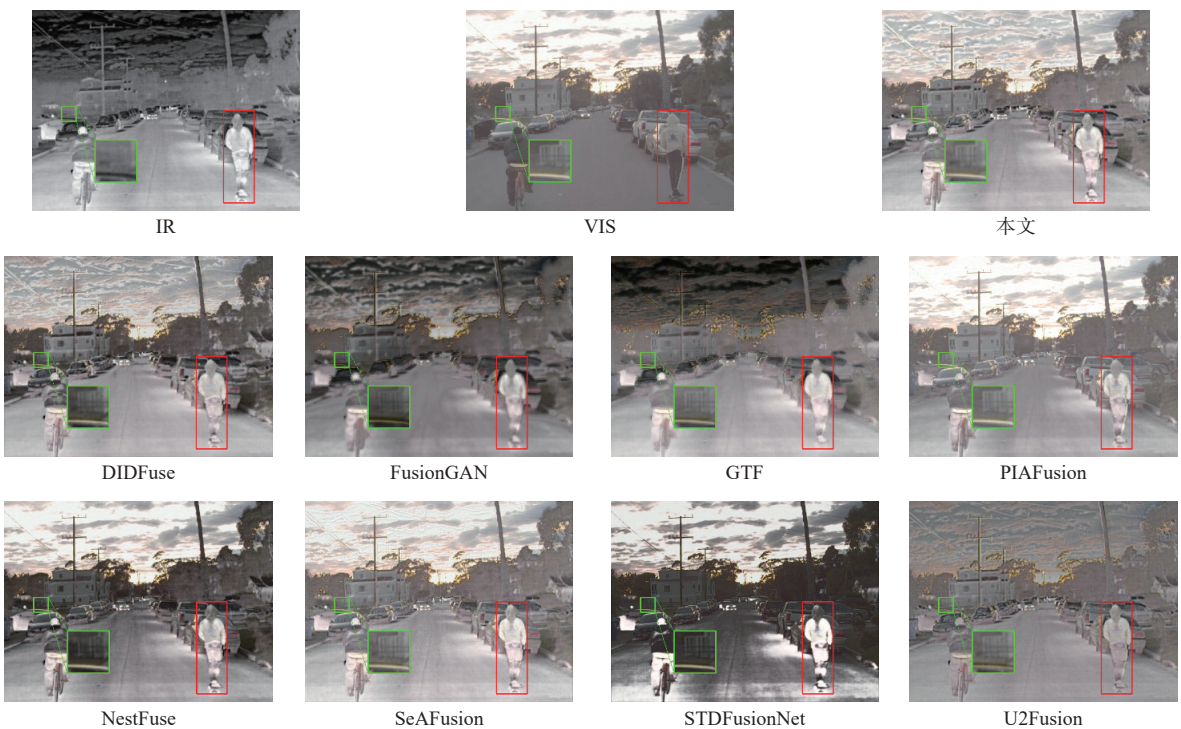


图 5　来自 RoadScene 图像对的融合结果

Fig. 5　Fusion results from RoadScene image pairs

算法的目标人物边缘不够锐利, STDFusionNet、U2Fusion 在不同程度上引入了伪影, NestFuse 等算法融合的图像从像素强度上看整体偏暗, 降低了视觉效果。而在绿框中, 与可见光图像相比大部分算法都出现了一定程度的纹理信息丢失减弱的现象。从空中云朵可以看到, FusionGAN、GTF、NestFuse、STDFusionNet、受到了一定程度的光谱污染, PIAFusion、NestFuse 和 SeAFusion 出现了纹理脱轨, 没有引入红外光纹理, 几乎所有对比的融合方法融合后的图像像素强度较低。而本文提出的模型采用强度损失、纹理损失、相似性损失 3 种损失函数, 调节图像强度损失的同时有效保持了纹

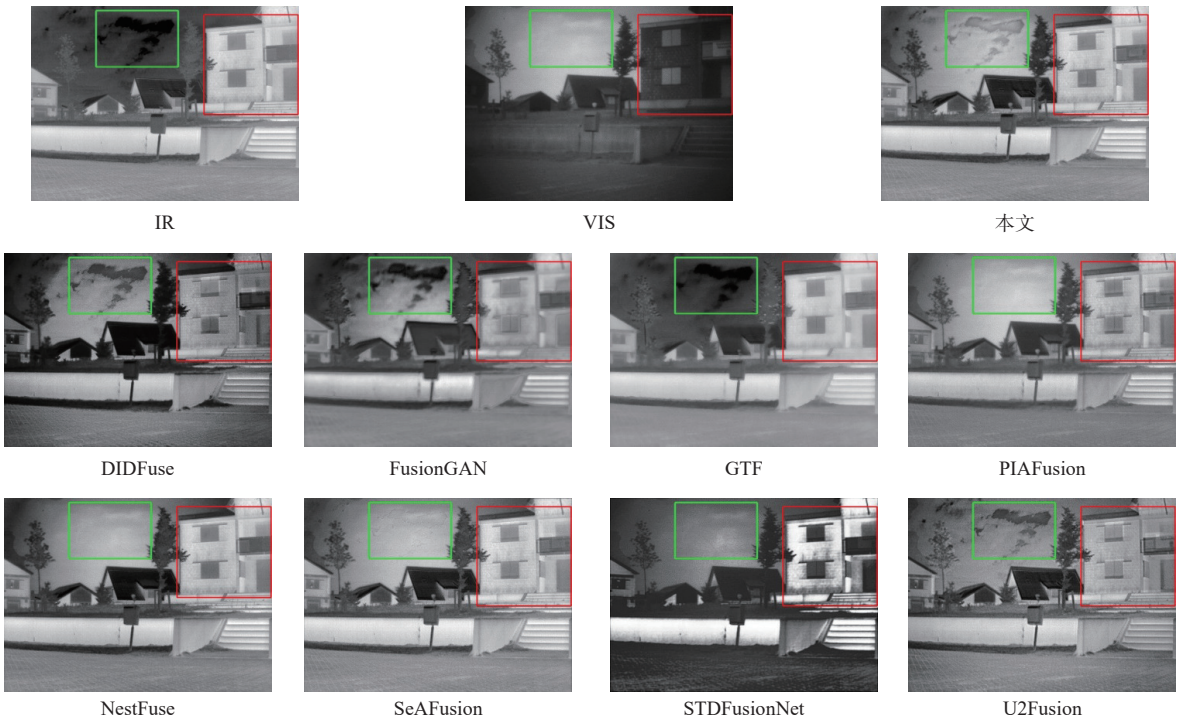


图 6 来自 TNO 图像对的融合结果

Fig. 6 Fusion results from TNO image pairs

理细节、图像对比度和相似度，因此没有出现目标强度降低、边缘不够锐利、光谱污染以及纹理丢失等问题。

图 6 显示了来自 TNO 图像对[20]的融合结果。绿框中生成的来自红外光图像的天空纹理图案中，PIAFusion、NestFuse、SeAFusion、STDFusionNet 几乎没有将纹理引入融合图像，FusionGAN 模糊了热辐射目标的轮廓使得天空纹理边界不清晰。在红框中，FusionGAN、STDFusionNet 融入了一些无用信息，再次将噪声带入到融合图像中使得显著特征内容变模糊。在地砖的整体纹理细节丰富度上，FusionGAN、GTF、PIAFusion 丢失了来自可见光的纹理细节。而本文的纹理损失函数能够有效提取更细粒度的纹理细节，多尺度的设计能使各级特征得到更加有效的提取，最大化融合源图像的各种特征。

2.2 客观指标对比

根据文献 [19] 图像融合度量可分为 4 种类型，提出了 13 个评价指标，从基于信息论的 4 个评价指标中选取交叉熵（cross entropy, CE）、峰值信噪比（peak signal-to-noise ratio, PSNR）；选取基于图像结构相似性的 2 个指标结构相似性（structural similarity index measure, SSIM）、均方根误差（root mean squared rrror, RMSE）；从基于图像特征的 5 个指标中选取标准差（standard deviation, SD）、基于梯度融合性能（gradient based fusion performance, Qabf）；选取基于人类感知度量的 2 个指标 Chen-Blum 标准（Chen-Blum metric, Qcb）、Chen-Varshney 标准（Chen-Varshney metric, Qcv）。4 种类型度量值中选出 8 种，另外选取近年文献使用的 VIF[10, 12, 21]，共 9 种指标进行评价，从 RoadScene、MSRS、TNO、VIFB 数据集中各挑选 10 对图像进行定量的客观指标对比，对比结果如表 2 所示。

由表 2 可知，不同方法因成像机理的不同导致在不同的数据集中表现出较大的差异，而本文提出的方法在各个数据集中均表现出了较为良好的性能，前 3 名的评价指标共有 29 个，排名第一的指标有 9 个。在 4 个数据集评价指标均值中，本文提出的方法在 9 个评价指标中获得了 5 个最佳值。

为更加直观比较各种算法的融合性能，负向指标取负后，将评价指标均值归一化，得到评价指标均值结果，如图 7 所示。由图可知，本文提出的方法在各项指标中均在高位区间。

2.3 运行时间对比

表 3 列出了本文所提 9 种算法在 4 个数据集共 40 个图像对的平均运行时间。从表中可以看到，各类算法有较大的差异，常规视频拍摄中，1 秒为 25 帧，平均每帧 0.04 s。与其他算法相比，本文方法每个图平均耗时小于 0.04 s，需要的时间最短，基本能够满足实时性的要求，主要原因在于采用了优化的特征提取器的同时，不增加网络架构复杂度，减少了运行时间。

表 2　实验评价结果

Table 2　Evaluation metric values of all methods

数据集	方法	评价指标								
		CE	PSNR	Qabf	Qcb	Qcv	RMSE	SSIM	SD	VIF
RoadScene	DIDFuse(2,0,0)	1.070	59.275	0.511	0.511	289	0.077	1.344	69.910	0.735
	FusionGAN(0,0,0)	1.142	58.594	0.272	0.496	573	0.091	1.318	40.735	0.365
	GTF(0,3,0)	0.523	60.121	0.344	0.498	460	0.064	1.484	58.562	0.430
	U2Fusion(3,1,1)	0.822	60.194	0.526	0.573	464	0.063	1.534	39.507	0.437
	PIAFusion(1,1,0)	0.503	59.665	0.426	0.525	318	0.071	1.516	47.939	0.392
	NestFuse(0,0,3)	0.730	59.279	0.536	0.496	285	0.077	1.303	64.756	0.598
	STDFusion(0,1,0)	1.622	58.145	0.312	0.386	880	0.100	1.215	69.405	0.589
	SeAFusion(1,1,2)	0.673	59.389	0.540	0.545	248	0.075	1.486	57.127	0.573
	Proposed(2,2,3)	**0.585**	**59.918**	**0.673**	**0.615**	**262**	**0.067**	**1.467**	**52.243**	**0.688**
MSRS	DIDFuse(0,0,0)	2.274	59.374	0.241	0.375	717	0.077	0.631	34.740	0.419
	FusionGAN(0,1,0)	1.919	60.094	0.171	0.390	3 892	0.064	1.266	20.460	0.122
	GTF(0,0,0)	0.891	59.994	0.271	0.432	4 305	0.066	1.218	19.645	0.166
	U2Fusion(0,0,0)	2.694	59.184	0.380	0.393	1 364	0.080	1.144	23.494	0.321
	PIAFusion(3,0,2)	0.638	60.003	0.696	0.583	89	0.066	1.434	52.559	0.795
	NestFuse(2,4,0)	0.697	60.188	0.723	0.567	68	0.062	1.428	50.046	0.715
	STDFusion(0,0,0)	4.365	59.361	0.497	0.350	301	0.076	1.183	37.551	0.414
	SeAFusion(1,3,3)	0.923	60.122	0.698	0.599	68	0.064	1.455	48.568	0.724
	Proposed(3,4,2)	**0.742**	**60.152**	**0.737**	**0.635**	**64**	**0.064**	**1.448**	**48.830**	**0.748**
TNO	DIDFuse(2,0,0)	1.069	59.714	0.351	0.507	315	0.071	1.282	45.935	0.643
	FusionGAN(0,0,0)	2.373	58.728	0.179	0.469	929	0.090	1.239	23.908	0.212
	GTF(2,1,1)	0.527	60.780	0.327	0.467	1 159	0.055	1.379	29.650	0.223
	U2Fusion(3,1,2)	0.713	60.790	0.424	0.567	503	0.055	1.411	34.033	0.640
	PIAFusion(1,2,0)	0.786	59.957	0.559	0.596	247	0.067	1.399	39.709	0.362
	NestFuse(0,1,2)	0.686	60.064	0.420	0.521	220	0.065	1.286	39.968	0.343
	STDFusion(0,1,0)	1.222	59.398	0.339	0.464	382	0.076	1.217	40.790	0.271
	SeAFusion(1,0,0)	0.770	60.097	0.410	0.539	218	0.065	1.354	38.775	0.478
	Proposed(1,2,4)	**0.600**	**60.238**	**0.486**	**0.611**	**222**	**0.063**	**1.320**	**38.724**	**0.479**
VIFB	DIDFuse(2,0,0)	2.087	57.961	0.501	0.523	404	0.107	1.260	53.718	0.748
	FusionGAN(0,0,0)	1.947	58.633	0.223	0.357	1 649	0.093	1.407	24.191	0.186
	GTF(0,2,0)	1.356	58.819	0.465	0.405	2 178	0.090	1.433	27.859	0.256
	U2Fusion(3,0,2)	1.099	59.304	0.572	0.556	724	0.081	1.483	34.047	0.552
	PIAFusion(1,4,2)	1.051	58.633	0.657	0.575	250	0.093	1.457	46.840	0.621
	NestFuse(0,0,0)	1.535	58.536	0.494	0.435	397	0.095	1.451	43.936	0.386
	STDFusion(0,0,2)	1.669	58.692	0.503	0.411	511	0.095	1.419	45.281	0.456
	SeAFusion(0,1,2)	1.413	58.377	0.579	0.470	328	0.099	1.458	43.936	0.512
	Proposed(3,2,1)	**1.010**	**58.658**	**0.693**	**0.581**	**278**	**0.092**	**1.450**	**45.171**	**0.657**
数据集均值	DIDFuse(1,1,0)	1.625	59.081	0.401	0.479	431	0.083	1.129	**51.076**	0.636
	FusionGAN(0,0,0)	1.845	59.012	0.211	0.428	1 761	0.084	1.308	27.323	0.221
	GTF(2,1,0)	0.824	59.929	0.352	0.450	2026	0.069	1.379	33.929	0.269
	U2Fusion(0,2,0)	1.332	59.868	0.476	0.522	764	0.070	1.393	32.770	0.487
	PIAFusion(1,3,2)	0.745	59.565	0.585	0.570	226	0.074	**1.452**	46.762	0.542
	NestFuse(0,1,0)	0.912	59.517	0.543	0.505	243	0.075	1.368	49.677	0.511
	STDFusion(0,0,1)	2.220	58.899	0.412	0.402	518	0.087	1.259	48.257	0.433
	SeAFusion(0,2,3)	0.945	59.496	0.557	0.538	216	0.076	1.438	47.102	0.572
	Proposed(5,0,2)	**0.734**	**59.741**	**0.647**	**0.611**	**206**	**0.071**	**1.421**	**46.242**	**0.643**

注: CE、Qcv和RMSE为负向指标, 其余为正向指标, 每个指标最佳3个值分别用红色、绿色和蓝色表示, 每个方法后的3个数字为第1、第2、第3值的数量。

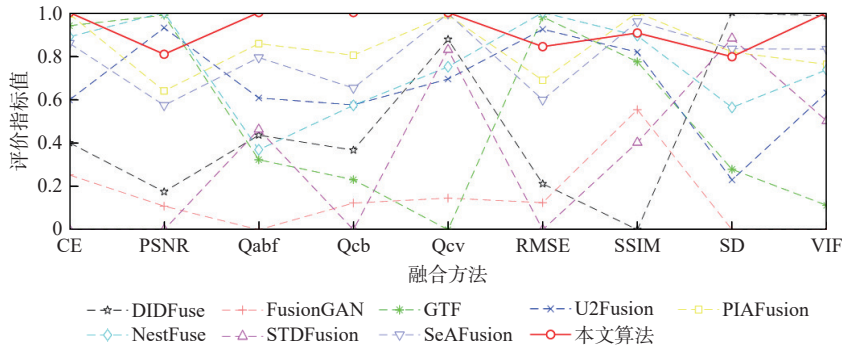


图 7 融合质量评价指标

Fig. 7 Indicators for evaluating the quality of fusion results

表 3 9 种算法在 4 个数据集共 40 个图像对的平均运行时间

Table 3 Average running time of the 9 algorithms in 40 image pairs in 4 data sets

融合方法	时间均值/s
DIDFuse	0.161
FusionGAN	2.703
GTF	8.214
U2Fusion	3.495
PIAFusion	4.531
NestFuse	0.318
STDFusion	1.415
SeAFusion	0.089
本文	0.034

注: 最佳3个值分别用红色、绿色和蓝色表示。

3 结 论

本文提出了一种基于嵌套连接的多尺度实时红外与可见光图像融合方法，利用特征提取器中的池化算子从源图像中提取出丰富的多尺度特征，然后将这些多尺度特征送入基于嵌套连接的特征融合器中生成具有多尺度深度特征的特征图像，最后通过图像重构器得到最终的输出图像。该模型能解决大多数方法存在的图像特征提取不充分、计算效率过低、依赖人工设计的融合策略以及无法提取多尺度深层次特征的问题。在公开数据集上的实验结果和分析表明，本文提出的融合网络具有优秀的融合性能，与代表性算法相比，在视觉上对于过度曝光、目标遮挡、细节模糊等复杂情况有较好的融合效果，在 3 类 9 个统计指标 CE、PSNR、Qabf、Qcb、Qcv、RMSE、SSIM、SD、VIF 上均能获得令人满意的结果。

参考文献（References）

[1] 曹金亮, 薛伟民, 蔡雨含, 等. 一种红外与可见光图像信息交换融合网络[J]. 太原科技大学学报, 2022, 43(4): 277-282.
CAO J L, XUE W M, CAI Y H, et al. Infrared and visible image information exchange-fusion algorithm[J]. Journal of Taiyuan University of Science and Technology, 2022, 43(4): 277-282 (in Chinese).

[2] 杨叶君, 刘刚, 肖刚, 等. 基于自适应特征增强和生成器路径交互的红外与可见光图像融合[J]. 激光与光电子学进展, 2022: 1-18.
YANG Y J, LIU G, XIAO G, et al. Infrared and visible image fusion based on adaptive feature enhancement and generator path interaction[J]. Laser & Optoelectronics Progress, 2022: 1-18(in Chinese).

[3] ZHANG X C, YE P, LEUNG H, et al. Object fusion tracking based on visible and infrared images: a comprehensive review[J]. Information Fusion, 2020, 63: 166-187.

[4] 王建中, 徐浩楠, 王洪枫, 等. 基于残差密集块和自编码网络的红外与可见光图像融合[J]. 北京理工大学学报, 2021, 41(10): 1077-1083.
WANG J Z, XU H N, WANG H F, et al. Infrared and visible image fusion based on residual dense block and auto-encoder network[J]. Transactions of Beijing Institute of Technology, 2021, 41(10): 1077-1083 (in Chinese).

[5] 宋加文, 朱大明, 付志涛, 等. 基于对比度和结构提取的红外与可见光图像融合[J]. 激光与光电子学进展, 2023, 60(14): 1410005.
SONG J W, ZHU D M, FU Z T, et al. Infrared and visible image fusion based on contrast and structure extraction[J]. Laser & Optoelectronics Progress, 2023, 60(14): 1410005(in Chinese).

[6] 张冬冬, 王春平, 付强. 深度学习框架下的红外与可见光图像融合算法综述[J]. 激光与红外, 2022, 52(9): 1288-1298.
ZHANG D D, WANG C P, FU Q. Overview of infrared and visible image fusion algorithms based on deep learning framework[J]. Laser & Infrared, 2022, 52(9): 1288-1298 (in Chinese).

[7] 魏琦, 赵娟. 红外和可见光图像融合算法的研究进展[J]. 计算机科学, 2023, 50(2): 11.
WEI Q, ZHAO J. Research progress of infrared and visible image fusion algorithms[J]. Computer Science, 2023, 50(2): 11(in Chinese).

[8] 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 Measurement, 2020, 69(12): 9645-9656.

[9] 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.

[10] 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.

[11] ZHAO Z X, XU S, ZHANG C X, et al. DIDFuse: deep image decomposition for infrared and visible image fusion[C]// Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. California: International Joint Conferences on Artificial Intelligence Organization, 2020: 1-5.

[12] 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.

[13] 汤璐, 卢荣胜, 史艳琼, 等. 基于 YCbCr 空间融合的高动态范围成像方法[J]. 激光与光电子学进展, 2022, 59(14): 308-319.
TANG L, LU R S, SHI Y Q, et al. High dynamic range imaging method based on YCbCr Color space fusion[J]. Laser & Optoelectronics Progress, 2022, 59(14): 308-319(in Chinese).

[14] 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.

[15] ZHOU Z W, SIDDIQUEE M M R, TAJBAKHSH N, et al. UNet++: a nested U-net architecture for medical image segmentation[J]. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018. Granada, Spain: [s.n.], 2018, 11045: 3-11.

[16] 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.

[17] 闵莉, 曹思健, 赵怀慈, 等. 改进残差密集生成对抗网络的红外与可见光图像融合[J]. 控制与决策, 2022, 37(1): 1-8.
MIN L, CAO S J, ZHAO H C, et al. Infrared and visible image fusion with improved residual dense generative adversarial network[J]. Control and Decision, 2022, 37(1): 1-8(in Chinese).

[18] 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.

[19] ZHANG X C, YE P, XIAO G. VIFB: a visible and infrared image fusion benchmark[C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2020: 468-478.

[20] 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.

[21] 张勇, 马东玺, 闫鹏程. 灰度夜视融合图像质量结构相似度评价方法[J]. 光子学报, 2014, 43(S1): 242-248.
ZHANG Y, MA D X, YAN P C. Structural similarity evaluation method to night vision fusion image quality[J]. Acta Photonica Sinica, 2014, 43(S1): 242-248(in Chinese).

Multi-scale infrared and visible image fusion based on nest connection

WU Nan，MU Chengpo*，HE Yang，LIU Taiye

(School of Mechatronical Engineering，Beijing Institute of Technology，Beijing 100080，China)

Abstract：This work suggests a multi-scale real-time image fusion model based on nest connection, which aims to address the issues of lengthy running times, unnatural fusion strategies, and the incapacity to extract multi-scale deep features in existing infrared and visible image fusion approaches. Firstly, multi-scale deep features are extracted by a feature extractor. Then, feature maps with multi-scale deep features are generated from the fusion network. Finally, the fusion image is reconstructed by an image re-constructor. In comparison to other algorithms tested on common datasets, a subjective qualitative comparison shows that the algorithm in this paper can contain sharp image edges while maintaining image intensity and it has a better fusion effect under complex conditions such as overexposure, target occlusion, and detail blurring. In objective quantitative metrics comparison, 5 best values and 2 second best values have been obtained on 9 evaluation metrics of 4 categories which are based on information theory, based on image feature, based on image structure similarity and human perception. Additionally, the other two metrics results have also showed good performance. Moreover, there has been a noticeable reduction in fusion time. The model presented in this paper has a high practicability and may successfully overcome the current issues with picture fusion techniques through experimental verification.

Keywords：image fusion; deep learning; nest connection; multi-scale; real-time

Received：2023-02-25；**Accepted**：2023-03-31；**Published Online**：2023-04-19 17:03

URL：link.cnki.net/urlid/11.2625.V.20230419.1544.001

* **Corresponding author**. E-mail：muchengpo@bit.edu.cn