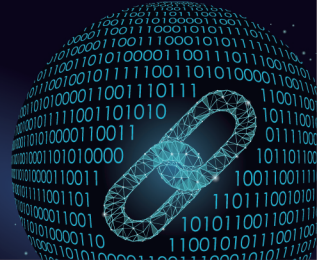




中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



空洞卷积并行注意力机制和纹理对比度增强的红外与可见光图像融合

邸敬, 任莉, 刘冀钊, 郭文庆, 廉敬

引用本文:

邸敬,任莉,刘冀钊,郭文庆,廉敬. 空洞卷积并行注意力机制和纹理对比度增强的红外与可见光图像融合[J]. 控制与决策, 2024, 39(11): 3586–3596.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2023.1353>

您可能感兴趣的其他文章

Articles you may be interested in

基于混合模型驱动的红外与可见光图像融合

Infrared and visible image fusion based on hybrid model driving

控制与决策. 2021, 36(9): 2143–2151 <https://doi.org/10.13195/j.kzyjc.2019.1749>

基于FRC框架的红外与可见光图像融合方法

Infrared and visible image fusion based on FRC algorithm

控制与决策. 2021, 36(11): 2690–2698 <https://doi.org/10.13195/j.kzyjc.2020.0669>

自适应直觉模糊相异直方图裁剪的图像增强算法

Adaptive intuitionistic fuzzy dissimilar histogram clipping image enhancement algorithm

控制与决策. 2021, 36(12): 2919–2928 <https://doi.org/10.13195/j.kzyjc.2020.0845>

基于自注意力生成对抗网络的图像超分辨率重建

Image super-resolution reconstruction based on self-attention GAN

控制与决策. 2021, 36(6): 1324–1332 <https://doi.org/10.13195/j.kzyjc.2019.1290>

基于双边分解与 L_1 暗通道的战场图像增强算法

A battlefield image enhancement algorithm based on bilateral decomposition and L_1 dark channel prior

控制与决策. 2021, 36(5): 1165–1172 <https://doi.org/10.13195/j.kzyjc.2019.1213>

空洞卷积并行注意力机制和纹理对比度增强的 红外与可见光图像融合

邸敬^{1†}, 任莉¹, 刘冀钊², 郭文庆¹, 廉敬¹

(1. 兰州交通大学 电子与信息工程学院, 兰州 730070; 2. 兰州大学 信息科学与工程学院, 兰州 730000)

摘要: 在机器视觉中, 视频监控的红外与可见光图像融合可以帮助机器更有效地识别目标和环境. 针对现有视频监控的红外和可见光图像算法中存在人物细节提取不充分、目标物体轮廓模糊等问题, 提出空洞卷积并行注意力机制和纹理对比度增强的红外与可见光图像融合. 首先, 融合网络采用多尺度密集连接和空洞卷积并行注意力机制构建局部-全局并行网络提取图像中的梯度信息和强度信息; 然后, 利用 Schar 滤波器和深度可分离卷积构建纹理对比度增强网络, 增强融合特征的对比度和纹理细节; 最后, 利用一种信息交换流网络设计分解网络, 由于分解图像的质量直接取决于融合结果, 分解过程可以使融合图像包含更多的场景信息. 通过与其他 8 种具有代表性的图像融合方法对比, 所提出方法的 7 项客观评价指标有 5%~62% 的提升, 表明所提出方法不仅能够充分提取源图像信息, 获得纹理细节更清晰、对比度更好的融合结果, 而且可以有效解决因源图像分辨率差异大而阻碍的多光谱遥感分析、军事侦察等实际应用.

关键词: 红外与可见光融合; 多尺度密集连接; 注意力机制; 纹理对比度增强网络; 深度可分离卷积; 信息交换流网络

中图分类号: TP391.41

文献标志码: A

DOI: 10.13195/j.kzyjc.2023.1353

引用格式: 邸敬, 任莉, 刘冀钊, 等. 空洞卷积并行注意力机制和纹理对比度增强的红外与可见光图像融合[J]. 控制与决策, 2024, 39(11): 3586-3596.

A dilated convolution parallel attention mechanism and texture contrast enhancement for infrared and visible image fusion

DI Jing^{1†}, REN Li¹, LIU Ji-zhao², GUO Wen-qing¹, LIAN Jing¹

(1. School of Electronic & Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; 2. School of Information Science & Engineering, Lanzhou University, Lanzhou 730000, China)

Abstract: In the realm of machine vision, the fusion of infrared and visible images for video surveillance enhances the ability of machines to recognize targets and environments with greater effectiveness. Aiming at the problems of insufficient character detail extraction and blurred target object contours in existing infrared and visible light image algorithms for video surveillance, a parallel attention mechanism with dilated convolution and texture contrast enhancement for infrared and visible image fusion are proposed. Firstly, the fusion network uses multi-scale dense connection and dilated convolution parallel attention mechanism to extract gradient and intensity information from the images. Then, the texture contrast-enhanced network is constructed using Schar filters and depthwise separable convolutions to enhance the contrast and texture details of fused features. Finally, the decomposition network is designed with an information exchange flow network. Since the quality of the decomposed images depends directly on the fusion result, the decomposition process can enable the fused image to contain more scene information. Compared with other eight representative image fusion methods, the seven objective evaluation indexes of this paper's method have an improvement of 5%~62%, which indicates that the proposed method not only can fully extract source image information and obtain fusion results with clearer texture details and better contrast, but also effectively solves the problem of multispectral remote sensing analysis, military reconnaissance, and other practical applications hindered by the large difference in resolution between source images.

Keywords: infrared and visible image fusion; multi-scale dense connection; attention mechanism; texture contrast enhancement network; depthwise separable convolution; information exchange flow network

收稿日期: 2023-09-22; 录用日期: 2024-01-16.

基金项目: 甘肃省科技计划项目(22JR5RA360); 国家自然科学基金项目(62061023); 甘肃省杰出青年基金项目(21JR7RA345).

责任编辑: 张国山.

[†]通讯作者. E-mail: 46891771@qq.com.

0 引言

图像融合是图像处理的一个分支,其目的在于通过使用适当的特征提取方法和融合策略,从源图像中生成包含显著特征和互补信息的单个图像。其中,红外与可见光图像融合是图像融合领域的一个重要课题,它可以有效提取每个原始图像之间的互补冗余信息,并将它们组合成高质量、信息丰富的图像,在遥感、目标检测、军事侦察等^[1-3]领域具有重要的应用价值。

红外和可见光图像融合方法大致分为两类:传统方法和深度学习方法^[4-7]。传统方法通常在空间域或频率域设计活动水平测量和融合规则,利用相关的数学变换实现融合,如多尺度变换^[8]和稀疏表示^[9]方法等。然而,这些方法都需要手动设置融合规则,且越来越趋于复杂。Veshki等^[10]提出一种耦合字典学习的多模态图像融合算法(coupled dictionary learning, CDL),通过保留融合图像中的不相关分量来增加融合图像纹理细节,用最大绝对值规则增强对比度分辨率。此方法保留了功能图像的色彩信息,高度满足人眼视觉特性,但用最大绝对值规则时忽略了结构图像的轮廓信息,导致融合图像中轮廓信息丢失,边缘模糊。继而Veshki等^[11]提出一种基于卷积稀疏编码(convolutional sparse coding, CSC)的红外和可见光图像融合算法,通过最大方差规则对耦合滤波器进行融合,并使用最大绝对值规则对稀疏编码进行融合,此方法在保留细节和局部强度方面有较好的提升,但鲁棒性欠佳。深度学习由于其强大的特征提取能力,已成为当前图像融合领域的主流方法。Liu等^[12]首次将卷积神经网络应用于图像融合,相比传统算法取得了很大进步,但需要进行有监督的训练。Prabhakar等^[13]提出了一种无监督端到端神经网络融合框架,其编码部分使用卷积神经网络提取特征,但是仅利用最后一层的特征进行融合,导致没有用到所有层信息,融合效果较差。随后,Zhang等^[14]提出了基于密集连接的图像融合方法SDNet,该方法使用两条密集连接构建的双分支网络进行特征提取,但未考虑单一尺度卷积核感受野受限,导致融合后图像清晰度不高。在此基础上,陈永等^[15]引入多尺度卷积,解决了单尺度卷积核感受野单一、特征提取不足等问题,但融合图像对比度不高,导致小目标轮廓不清晰。Wang等^[16]提出密集Res2net和注意力机制的红外与可见光图像融合算法,采用密集Res2net提取多尺度深度特征,再通过注意力机制细化特征,关注源图像中典型目标和细节,但该方法关注特征目标时忽略了背景信息,导致

背景纹理和细节信息被模糊化。

针对以上问题,本文提出一种空洞卷积并行注意力机制和纹理对比度增强的红外与可见光图像融合网络模型。采用多尺度密集连接和空洞卷积并行注意力机制(dilated convolution parallel attention mechanism, DCPAM)构建局部-全局并行网络提取图像中的梯度信息和强度信息,减少中间层信息丢失,并相应地设计代表梯度信息和强度信息的损失项函数。利用Scharr滤波器和深度可分离卷积构建纹理对比度增强融合网络以整合互补信息,增强融合特征的对比度和纹理细节,使融合结果更符合人眼感知视觉。此外,不仅考虑从源图像到融合图像的融合过程,而且考虑从融合图像到源图像的分解过程,并提出信息交换流网络。该网络能充分提取来自融合图像的红外和可见光图像信息,并利用融合结果作为分解质量的参考,从而强制融合结果包含更多的场景细节。

1 图像融合方法

1.1 网络总体框架

红外图像具有很强的抗干扰能力,可见光图像具有较好的纹理细节和空间分辨率。为了提高红外与可见光图像的融合质量,提出一种无监督端到端红外和可见光图像融合模型,网络总体框架如图1所示。框架主要由融合网络、增强网络和分解网络3部分组成。首先,融合网络利用卷积提取图像中的浅层特征信息;其次,将浅层特征信息输入到由多尺度密集连接和空洞卷积并行注意力机制构建的局部-全局并行网络模块中进一步提取深层特征信息和显著信息,并利用Scharr滤波器和深度可分离卷积构建纹理对比度增强融合网络;接着,将提取的深层特征、对比度增强特征和纹理特征信息进行融合;最后,分解网络利用卷积构建信息交换流网络从融合图像中提取特征,尽可能分解得到与源图像一致的图像,迫使融合结果包含更多的场景信息。

1.2 融合网络

1.2.1 多尺度密集连接

现有深度学习方法使用单一尺度卷积提取特征,导致中间层大量信息丢失,融合效果较差。为了能够充分提取红外和可见光图像的特征信息,首先选择 5×5 和 3×3 两种大小的卷积核对输入图像进行不同维度特征信息的提取,然后将其使用DenseNet密集连接,保证网络中每层信息最大传输的前提下,利用concat操作将所有层连接起来,减少信息丢失。网络结构如图2所示。

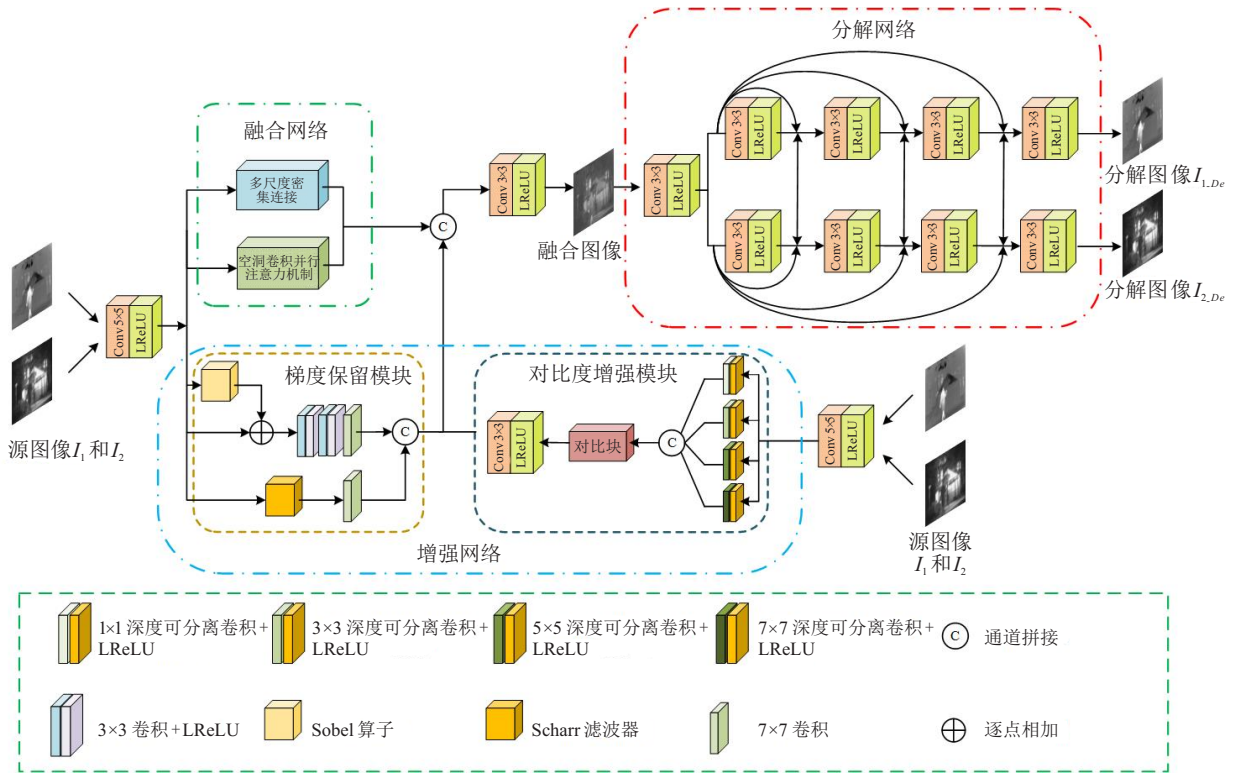


图1 网络总体框架

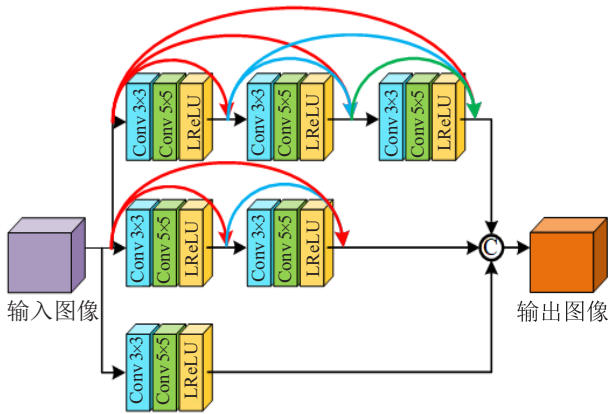


图2 多尺度密集连接网络结构

1.2.2 空洞卷积并行注意力机制

多尺度密集连接可以充分提取红外和可见光图像的局部信息,但无法获得全局关联信息,而注意力机制能够捕获全局关联信息,因此本文提出一种空洞卷积并行注意力机制(dilated convolution parallel attention mechanism, DCPAM)进行全局特征信息关联,有效融合上下文信息.卷积注意力机制中的空间模型使用7×7的卷积核来扩大感受野以提高空间注意力能力,导致模块参数增加.本文提出的DCPAM注意力机制利用空洞卷积减少模块参数数量,在不影响空间注意力聚合能力的情况下简化模型.具体空间注意力计算过程如下:

$$M_s(F) = \sigma(f_{dilat}^{3 \times 3}([\text{AvgPool}(F); \text{MaxPool}(F)])) =$$

$$\sigma(f_{dilat}^{3 \times 3}([\text{AvgPool}(F)_{avg}^S; (F)_{max}^S])). \tag{1}$$

其中: F 表示输入特征图; $M_s(F)$ 表示基于空间注意力 F 的输出权重; σ 表示Sigmoid函数; $f_{dilat}^{3 \times 3}$ 表示卷积核大小为3的空洞卷积,空洞率为2.

卷积注意力机制不仅有助于图像融合聚焦重要部分和保留更多的细节信息,而且是一种轻量级网络,可以很容易地嵌入网络框架中的任何位置.然而,卷积注意力机制通常需要两个阶段的特征图输出,前一阶段输出的特征图在一定程度上已经对原始输入特征进行了美化,导致后一阶段注意力所学习到的特征发生了变化.在图像融合任务中,这种变化会使注意力模块效果变差,从而影响融合质量.因此,在本文提出的DCPAM中将两阶段的特征图输出变为一阶段的输出,可以同时学习输入特征图,在减少运行速度的同时不会降低模型性能,其过程公式为

$$F'' = M_C(F) \otimes M_S(F) \otimes F, \tag{2}$$

$$M_C(F) = \sigma(\text{MLP}(\text{AvgPool}(F)) + \text{MLP}(\text{MaxPool}(F))) = \sigma(W_1(W_0(F_{avg}^C)) + W_1(W_0(F_{max}^C))). \tag{3}$$

其中: F'' 表示最终的输出特征图, F 表示输入特征图, $M_C(F)$ 表示特征图 F 通过信道注意力的输出权重, $M_S(F)$ 表示特征图 F 通过空间注意力的输出权重, F_{avg}^C 和 F_{max}^C 分别表示全局最大池化和平均池化,

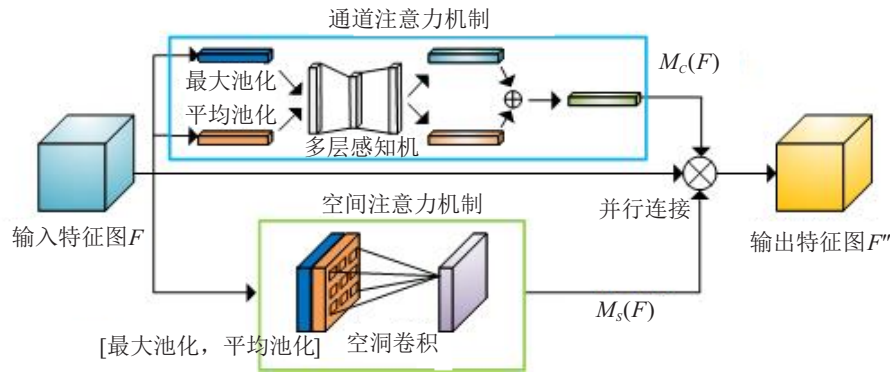


图3 DCPAM模块

⊗表示对应元素的乘法运算,MLP表示多层感知机, W_0 和 W_1 表示经过MLP的权重值.DCPAM整体结构框架如图3所示.

1.3 增强网络

纹理对比度增强网络由两个模块组成:梯度保留模块(**gradient retention module, GRM**)和对比度增强模块(**contrast enhancement module, CEM**). GRM主要由两部分组成,一部分采用Sobel算子,以保留强大的特征纹理,并经过 1×1 卷积消除通道尺寸差异;另一部分将原始输入特征和经过Scharr滤波器的特征图进行add操作后,经两个 3×3 卷积层和一个 1×1 卷积层进一步提取图像的弱纹理特征,最后使用concat操作将其连接,以捕获具有详细、细致的深度特征,实现纹理增强.GRM网络结构如图4所示.

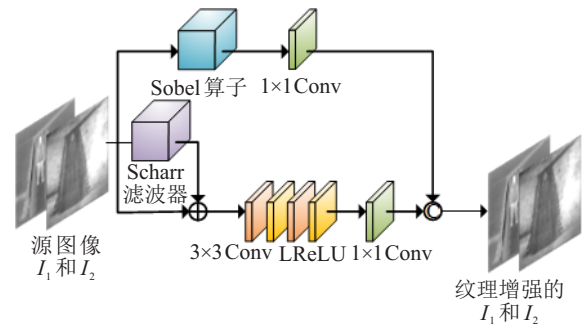


图4 GRM网络结构

构建轻量级网络;然后将获得的特征以通道尺寸连接,并发送到对比块,在对比块中计算每个像素位置的标准差和分别以 (i, j) 为中心窗口、半径为 r 的标准偏差,并计算偏差;最后利用最大池化层和一个具有LReLU的全连接层输出表示特征对比权值的激活向量生成增强特征.

1.4 分解网络

分解网络的目的是对融合后的图像进行分解,尽可能产生与源图像相一致的结果,从而迫使网络得到的融合图像保留更多的场景细节信息,具体结构如

CEM网络结构如图5所示,使用卷积核大小分别为 1×1 、 3×3 、 5×5 、 7×7 的深度可分离卷积提取信息.深度卷积操作首先将输入的特征进行多尺度独立空间卷积,再通过逐点卷积操作将其映射到新的空间通道,在保证进行深层特征提取的同时减少网络参数,

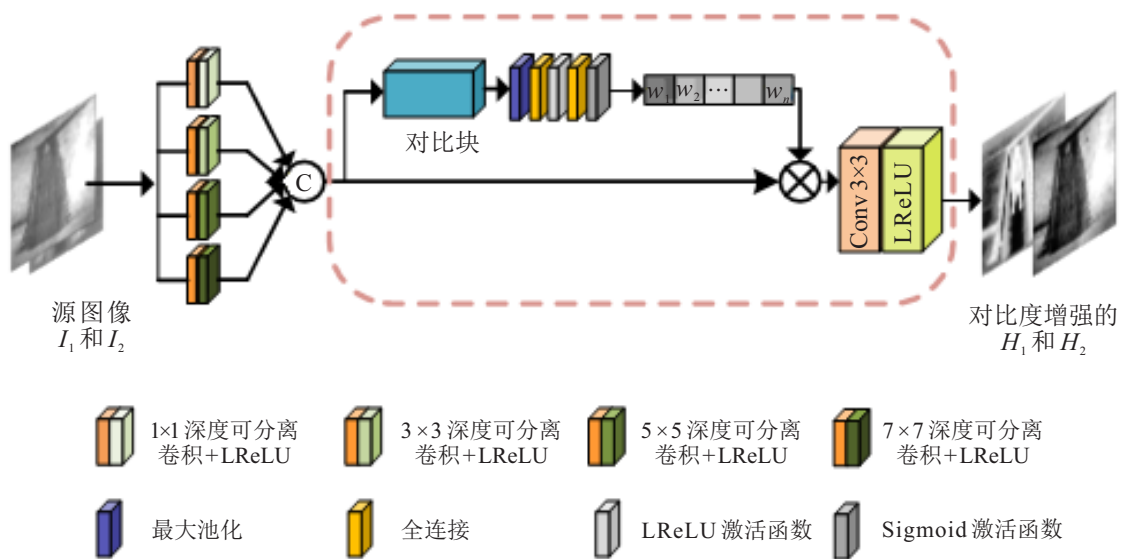


图5 CEM网络结构

- 

1x1 深度可分离卷积+LReLU
- 

3x3 深度可分离卷积+LReLU
- 

5x5 深度可分离卷积+LReLU
- 

7x7 深度可分离卷积+LReLU
- 

最大池化
- 

全连接
- 

LReLU 激活函数
- 

Sigmoid 激活函数

图6所示.在分解网络中,考虑到分解是从高维度到低维度的过程,并且要更好地捕捉红外与可见光之间特征信息的复杂关系,提出一种交换流网络,专注于对输入特征进行交叉操作,有助于学习输入特征之间的高度交互性,可以更好地捕捉特征间的复杂关系.首先利用公共卷积提取融合图像的浅层特征,其次构建双分支网络进行深层特征提取,且使用3个交换流(图中虚线)用于在每个特征提取阶段之后进行

特征信息的交换和融合,充分提取来自融合图像的信息.以A1和B1为例,交换流的具体操作是A1卷积后的特征图与B1卷积后的特征图进行链接,同时B1卷积后的特征图与A1卷积后的特征图进行连接.在这一基础上,创新地将融合图像直接插入到每一层中,实现跳跃连接.这样的设计使得整个网络能够捕捉不同红外与可见光图像特征之间的关联,而不仅仅限于线性关系,进一步促使融合图像达到更佳的效果.

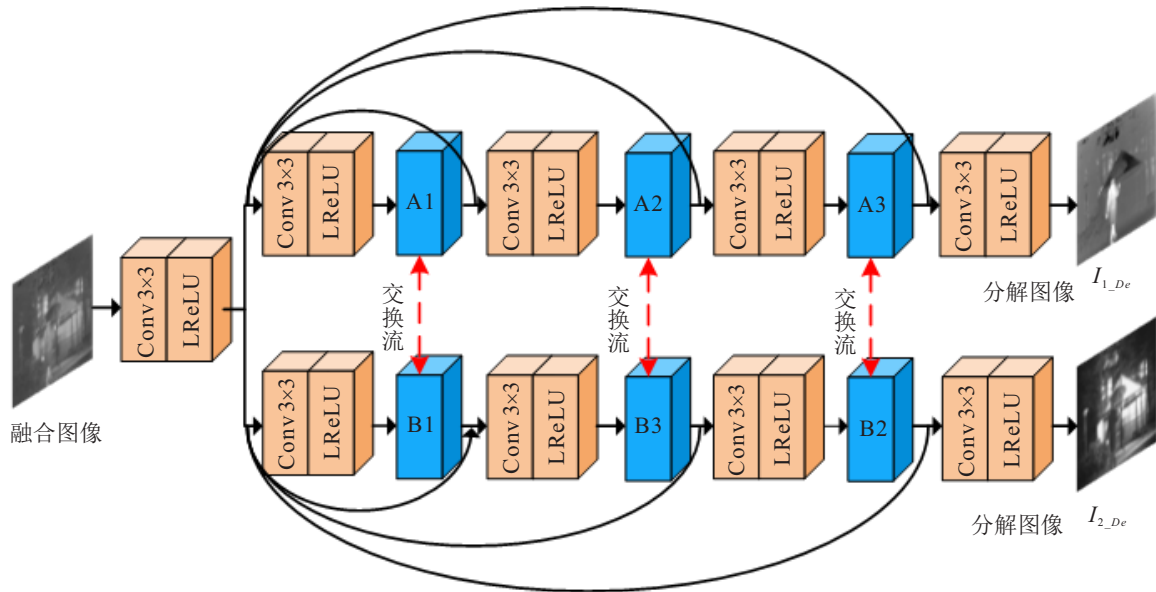


图6 分解网络结构

1.5 损失函数

红外和可见光图像的基本组成元素是像素,像素的强度可以表示整体亮度分布,像素之间的差异构成梯度,可以表示图像中的纹理细节.因此,将红外与可见光图像融合建模为对图像中梯度信息和强度信息的提取和重构问题.融合网络和增强网络用于提取、增强和融合这两种信息,使得融合图像对比度高、纹理细节清晰.分解网络用于对强度和梯度信息进行分解和重构.相应地,损失函数由3部分组成:强度损失、梯度损失和分解损失,定义如下:

$$L = L_{\text{grad}} + L_f + L_{\text{dc}}. \quad (4)$$

其中: L 为总损失函数, L_{grad} 为梯度损失, L_f 为强度损失, L_{dc} 为分解损失.

1.5.1 梯度损失

梯度损失 L_{grad} 迫使融合图像包含丰富的纹理细节特征.在梯度损失中引入自适应权重块,首先利用高斯低通滤波,尽可能地减少噪声对权重块的影响,然后基于梯度丰富度评估源图像对应像素的优化目标,使融合图像中的纹理自适应地向源图像中的纹理靠近,有

$$L_{\text{grad}} = \frac{1}{HW} \sum_i \sum_j S_{1,i,j} \cdot (\nabla I_{\text{fused},i,j} - \nabla I_{1,i,j})^2 + S_{2,i,j} \cdot (\nabla I_{\text{fused},i,j} - \nabla I_{2,i,j})^2, \quad (5)$$

$$S_{1,i,j} = \text{sign}(|\nabla(L(I_{1,i,j}))|) - \min(|\nabla(L(I_{1,i,j}))|, |\nabla(L(I_{2,i,j}))|), \quad (6)$$

$$S_{2,i,j} = 1 - S_{1,i,j}. \quad (7)$$

其中: i 和 j 为像素在第 i 行和第 j 列的权重, H 和 W 为图像的高度和宽度, $I_{1,i,j}$ 和 $I_{2,i,j}$ 为源图像第 i 行和第 j 列的像素值, $I_{\text{fused},i,j}$ 为融合图像, $|\cdot|$ 为绝对值函数, $\nabla(\cdot)$ 为拉普拉斯算子, $\nabla I_{1,i,j}$ 为第1幅源图像的拉普拉斯算子, $\nabla I_{2,i,j}$ 为第2幅源图像的拉普拉斯算子, $L(\cdot)$ 为高斯低通滤波器函数, $\min(\cdot)$ 为最小函数, $\text{sign}(\cdot)$ 为符号函数, $S_{i,j}$ 为基于源图像梯度水平生成的权重块.

1.5.2 强度损失

强度损失采用改进Frobenius范数,其不同于欧氏距离,是矩阵中每一个位置元素的平方和取平均数

再开方,作用是衡量融合图像像素与源图像像素的距离,有效调整融合图像与两个源图像之间的权重,迫使网络选择更有效的信息,公式如下:

$$L_{\text{int}} = \sqrt{\frac{1}{HW} \sum_i \sum_j |I_{\text{fused},i,j} - I_{1,i,j}|^2 + \alpha |I_{\text{fused},i,j} - I_{2,i,j}|^2}. \quad (8)$$

其中: I_1 和 I_2 为第1幅和第2幅源图像, I_{fused} 为融合图像, H 和 W 为图像的高度和宽度, α 为强度损失项的权重比例.

1.5.3 分解损失

分解损失 L_{dc} 要求融合后的单幅图像分解结果尽可能与源图像相近,其被定义为

$$L_{\text{dc}} = \frac{1}{HW} \sum_i \sum_j (I_{1\text{-de},i,j} - I_{1,i,j})^2 + (I_{2\text{-de},i,j} - I_{2,i,j})^2. \quad (9)$$

其中: I_1 和 I_2 为第1幅和第2幅源图像, $I_{1\text{-de}}$ 和 $I_{2\text{-de}}$ 为分解后的第1幅和第2幅图像结果, H 和 W 为图像的高度和宽度. 由于分解后图像与输入图像之间的差异程度决定融合结果的效果,分解损失函数将尽可能使融合结果获得更多图像信息,从而使融合质量达到最好.

2 实验结果与分析

2.1 实验设置与环境

本文算法选用公开数据集 TNO、RoadScene、MSRS 进行红外与可见光图像融合任务. 在训练时,由于红外和可见光图像数据集有限,通过裁剪分解的方法对数据进行扩张,从 TNO 数据集中挑选最大像素 576×768 和最小像素 360×270 的红外和可见光图像进行剪裁,得到 35 484 个大小为 12×120 的图像. 在测试阶段选择 10 对图像作为测试数据,与训练数据不同,测试是在源图像的原始大小上执行.

实验在 Windows10 操作系统的 Intel Core i5-1035G1 CPU 下进行,主要采用 tensorflow、imageio 库在 Pycharm 编译器中训练和测试网络性能. 实验设置 $\text{epoch} = 15$ 、 $\text{batch_size} = 32$ 、 $\text{learning_rate} = 1e-4$. 在优化算法方面,选择适应性强、收敛性好的自适应优化算法 Adam. 此外,总损失中包含的梯度损失、强度损失、分解损失的比例经过重复实验后设置为 80: 1: 1.

对于包含彩色图像融合任务的 MSRS 数据集,采用 YCbCr 模型进行颜色空间转化. 首先,将彩色图像从 RGB 转换到 YCbCr 颜色空间,由于 Y 通道可以反映结构细节和亮度变化,本文致力于对彩色图像的 Y

通道灰度图像进行融合;然后,采用不同的融合方法将可见光图像的 Y 通道与红外图像进行融合;最后,直接将彩色图像的 Cb 和 Cr 通道与融合后的 Y 通道级联,并将这些分量转移到 RGB 空间,得到最终的融合结果.

2.2 融合图像分析

为了验证算法的有效性,将本文算法与其他 8 种算法进行比较,包括 CDL^[10]、CCFL^[11]、MFBEMD^[17]、RTVD^[18]、SMVIF^[19]、SDNet^[14]、BF^[20]、PMGI^[21],从主观视觉和客观评价指标两方面进行分析.

2.2.1 TNO 数据集主客观评价

从构建的 TNO 测试数据集中选取 6 种经典场景作为红外与可见光融合的源图像,其中包括 Nato camp、helicopter、Marne_04、Movie_04、Movie_18、sandpath. 图 7 展示了 6 组源图像融合结果. 通过对比可以看出,RTVD 算法融合结果虽然能够突出红外目标人物,但未充分提取源可见光信息,导致融合结果严重缺失场景信息. BF 算法则是充分提取了可见光图像信息,但丢失红外图像目标特征. CDL 和 CCFL 算法将红外与可见光图像信息全部融合,但噪声现象严重,部分融合场景出现失真. SMVIF 算法相对于 CDL 算法和 CCFL 算法的融合结果噪声现象得到改善,但整体对比度低,融合特征细节信息缺失. PMGI 算法很大程度上解决了上述问题,但一些重要信息仍无法准确体现,如第 5 组图像中,窗户边缘不清晰. MFBEMD 算法和 SDNet 算法整体对比度较好,目标突出,但是目标边缘不够清晰,一些细节信息表征不明显,如 MFBEMD 算法中第 1 组图像左下角的草丛缺失纹理细节信息,SDNet 算法中第 6 组左下角河边的草堆同样缺失纹理细节. 本文算法在保留特征信息的同时,图像亮度和边缘梯度最好,目标突出,背景纹理信息清晰,场景信息丰富,视觉效果佳.

为了使融合结果更具说服力,选取信息熵 (entropy, EN)、平均梯度 (average gradient, AG)、改进空间频率 (spatial frequency, SF)、标准差 (standard deviation, SD)、基于视觉信息保真度 (visual information fidelity for fusion, VIFF)、互信息 (mutual information, MI) 和基于多尺度结构相似性 (multi-scale structural similarity, MSSSIM) 七种客观评价指标进行评价. 同时,从融合结果中选取一组最具代表性的图像进行评价,如表 1 所示.

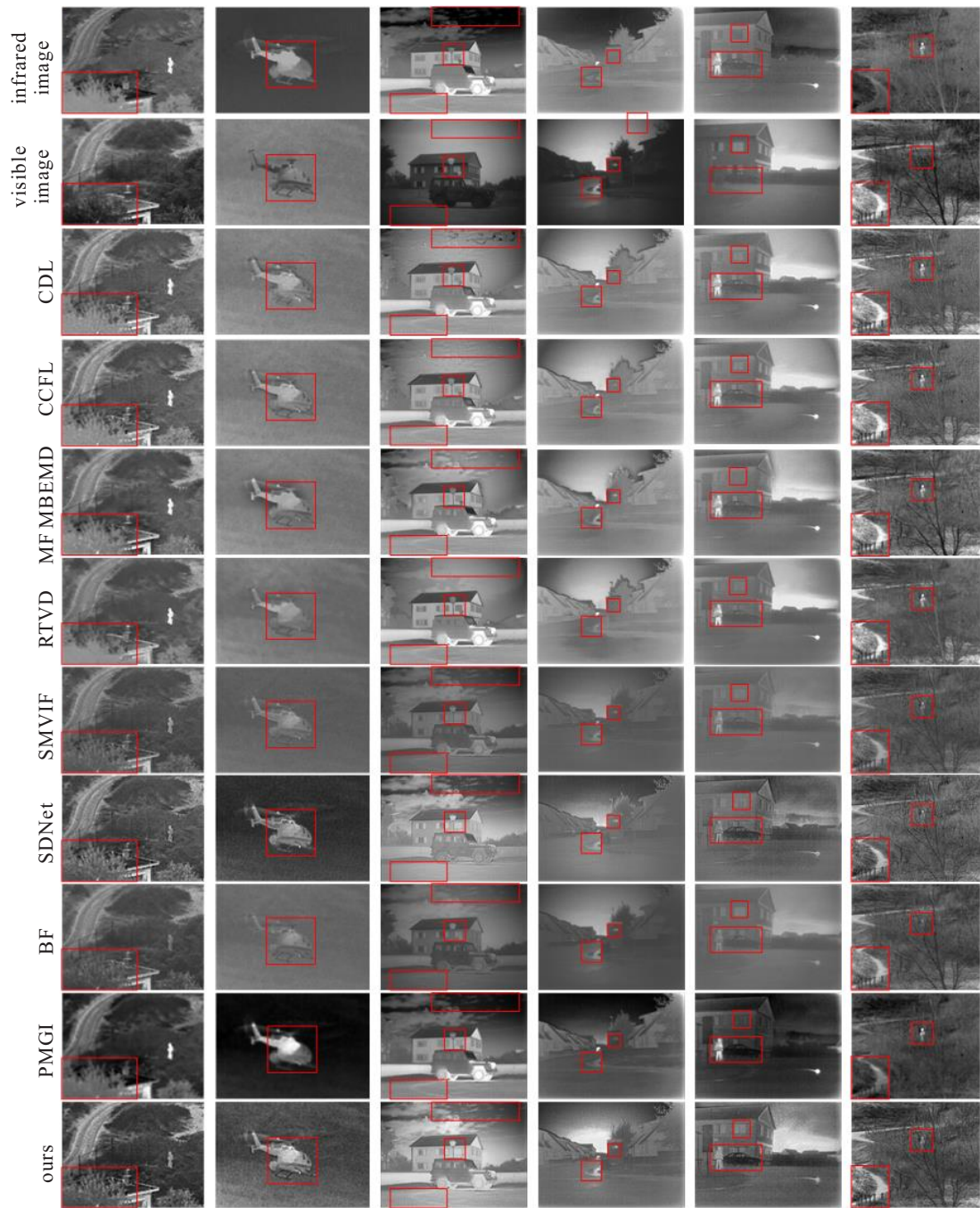


图7 TNO数据集6组场景的融合结果

表1 代表性图像的客观评价结果(第1列图像)

Indicators	CDL	CCFL	MFM BEMD	RTVD	SMVIF	SDNet	BF	PMGI	ours
EN	6.7934	6.8127	6.8323	6.6069	6.3308	6.7848	6.4218	6.8585	6.9714
AG	5.4709	4.8605	4.5634	4.0701	5.2782	6.1397	3.6227	2.6098	6.2186
SF	16.1523	15.2942	15.3061	12.7258	15.6479	18.5612	11.3134	8.6448	18.7816
SD	5.5624	5.5693	5.5664	5.3593	4.8745	5.5237	5.1595	5.9259	6.1900
VIFF	0.4261	0.4496	0.3331	0.2498	0.3497	0.4653	0.2690	0.3202	0.7061
MI	13.5869	13.6254	13.6652	13.2140	12.6616	13.5696	12.8435	13.7170	13.9428
MSSSIM	0.8856	0.8916	0.8833	0.7744	0.8947	0.9079	0.8676	0.7892	0.9358

表2为6组图像客观评价指标的平均值。可以看出:本文方法的EN值高,表明融合后图像包含信息丰富;AG值高,表明融合后图像的细节表征和纹理特征明显;SF值高,说明融合后图像清晰度高;SD值高,表明图像灰度值离散度高,图像携带的细节信息丰

富;VIFF和MSSSIM值高,表明图像与人眼视觉特性一致;MI值高,表明融合后图像获取了大量源图像信息。从主观视觉分析和客观评价指标上都可以看出所提出方法在特征提取、纹理细节保存、融合质量等方面均优于其他融合算法。

表2 TNO数据集6组图像的客观评价指标平均值

Indicators	CDL	CCFL	MFBMBEMD	RTVD	SMVIF	SDNet	BF	PMGI	ours
EN	6.6228	6.6061	6.6187	6.5121	5.8899	6.5903	6.0056	6.6125	6.9049
AG	4.8630	3.4308	4.1184	3.6115	4.5372	5.8549	2.9596	1.7732	6.0983
SF	12.5029	10.1055	11.3858	9.9835	11.5560	15.3832	7.7413	6.7259	16.2124
SD	5.3177	5.2977	5.3927	5.3262	4.1291	5.2015	4.4970	5.7060	5.9810
VIFF	0.3345	0.2960	0.2758	0.2032	0.2520	0.3969	0.1545	0.3042	0.7468
MI	13.2476	13.2141	13.2555	13.0262	11.7798	13.1806	12.0112	13.1750	13.8099
MSSSIM	0.8925	0.9017	0.9034	0.8113	0.9025	0.8951	0.8668	0.8230	0.9222

2.2.2 RoadScene数据集主客观评价

为了进一步验证本文方法的优越性,选取RoadScene中的30对源图像进行测试,并从中选取“FLIR_00497”一组典型图像进行主客观对比,如图8所示。从图8可以看出:PMGI方法路面细节信息丢失,指示牌信息模糊,整体融合效果最差;RTVD方法、SMVIF方法和BF方法指示牌信息清晰,路面信息有

所保留,但是部分可见光信息提取不充分,导致融合结果纹理细节信息丢失;CDL方法、CCFL方法和MFBMBEMD融合结果整体较好,但图像对比度略低,一些纹理细节信息丢失;SDNet高效融合了红外信息与可见光信息,但是路面细节信息不够突出;本文方法有效提取了纹理细节信息,目标显著,实现了红外与可见光图像的良好平衡。

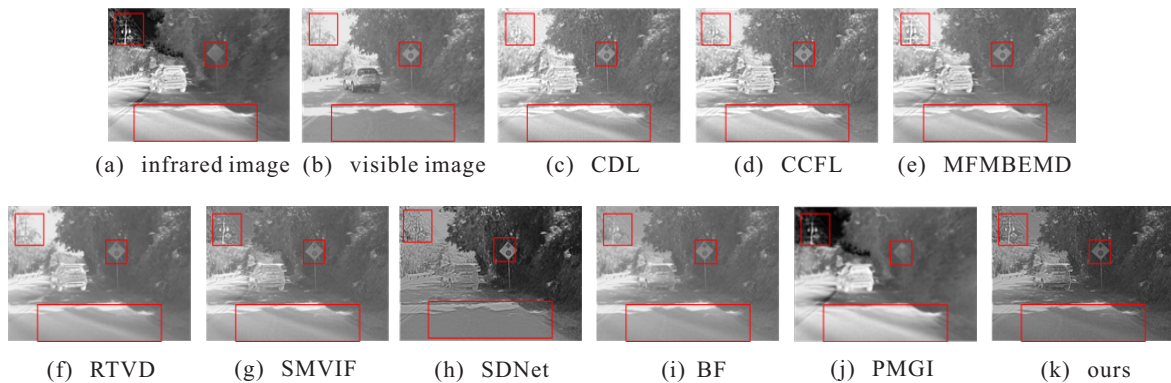


图8 “LIR_00497”场景融合结果

LIR_00497场景融合结果如表3所示。由表3可知,本文方法EN、AG、SF、VIFF、MSSSIM指标高,表明融合图像信息丰富、纹理细节、对比度较好与主观视觉一致。CDL方法的SD指标高,表示细节纹理信息较好;RTVD方法的MI指标高,也侧面印证了其

未能充分提取可见光图像信息;而本文方法5种评价指标均最佳,表现本文充分提取了红外和可见光图像信息,并在红外与可见光特征中实现了良好的平衡,在保留细节的同时突出了红外目标。

表3 FLIR_00497场景客观评价指标

Indicators	CDL	CCFL	MFBMBEMD	RTVD	SMVIF	SDNet	BF	PMGI	ours
EN	7.1850	7.1561	7.1434	6.9331	6.9771	7.0584	6.8438	7.2784	7.3709
AG	6.0560	5.3246	5.5241	2.7439	5.9952	7.3269	3.9004	2.1651	7.6549
SF	18.2404	17.6876	18.2400	11.5091	18.8586	22.8114	12.2613	7.8543	24.0680
VIFF	0.4240	0.4952	0.4273	0.3269	0.4888	0.4484	0.3511	0.2834	0.4995
MI	3.5319	3.5017	3.8899	3.9024	3.1610	2.9504	3.5149	3.6255	2.8531
MSSSIM	0.8709	0.9031	0.9109	0.8153	0.9291	0.8900	0.8663	0.7128	0.9459

2.2.3 MSRS数据集主客观评价

继续从MSRS测试数据集中挑选一组白天场景和夜间场景的源图像进行融合,如图9和图10所示。可以看出:PMGI方法融合结果对比度低,清晰度不高,整体融合效果差;BF方法、CCFL方法和SMVIF方法在一定程度上提升了清晰度,但整体对比度低,

导致一些细节信息无法体现;MFBMBEMD、SDNet和RTVD方法融合效果相比之前有了明显改善,但可见光信息丢失,例如大楼上的字母没有体现在融合图像中;CDL方法融合的图像整体融合效果较好,但是细节信息存在丢失;而本文融合方法不存在失真,整体对比度亮,图像清晰,目标显著,融合效果最佳。

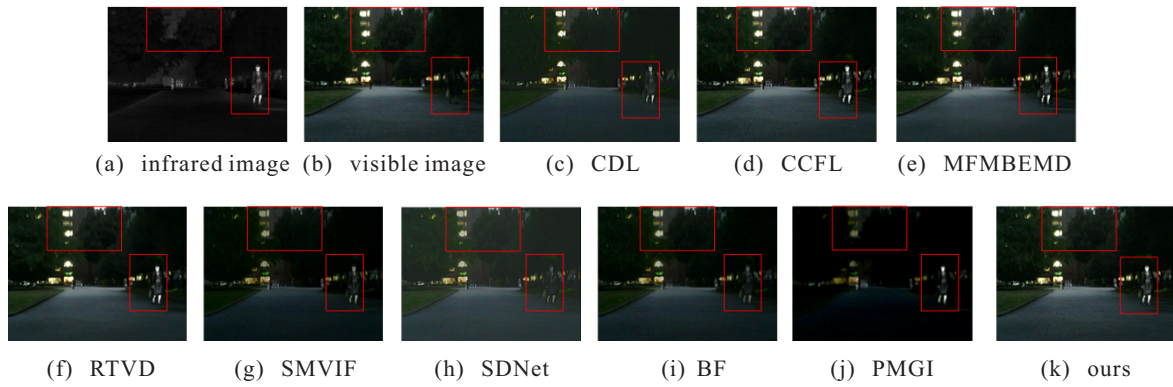


图9 MSRS数据集夜间场景融合结果

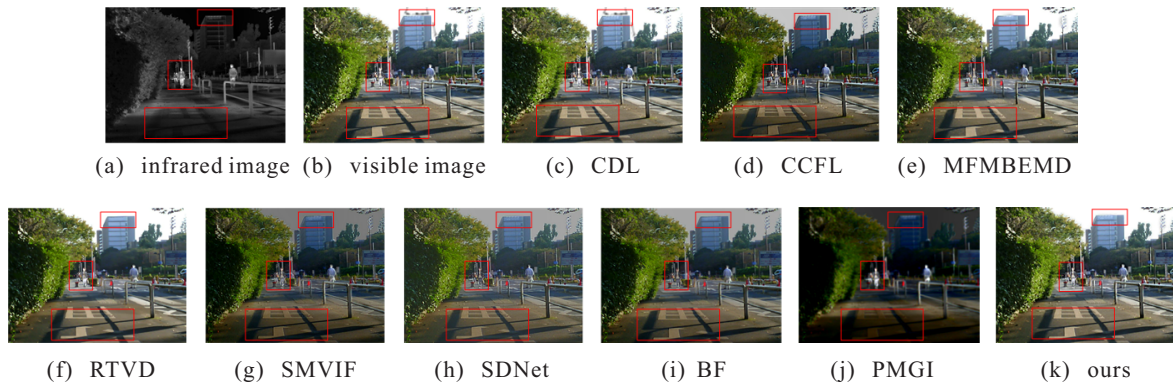


图10 MSRS数据集白天场景融合结果

表4 MSRS夜间场景客观评价指标

Indicators	CDL	CCFL	MFMBEMD	RTVD	SMVIF	SDNet	BF	PMGI	ours
EN	5.7175	6.4095	6.4003	6.3916	5.1616	5.8396	5.7380	2.5474	6.4217
AG	2.1890	2.4176	2.6148	1.9920	2.2300	2.4964	1.8866	1.0741	3.0119
SF	10.4922	10.6580	10.8453	10.3395	8.5336	9.6964	7.5502	8.5919	11.1707
SD	4.7355	5.6424	5.6228	5.6396	4.3582	4.4815	4.7356	3.8180	5.6452
VIFF	0.5073	0.8534	0.8053	0.7763	0.4792	0.4632	0.5354	0.2377	0.8564
MI	1.9326	3.4506	3.1932	3.4823	2.1728	2.6310	2.9784	1.2846	3.2309
MSSSIM	0.9209	0.9724	0.9628	0.9336	0.9436	0.9409	0.9530	0.8408	0.9714

MSRS 夜间场景客观评价指标如表4所示。由表4所知,本文方法的EN、AG、SF、SD、VIFF指标较高,而MI指标和MSSSIM指标分别略低于RTVD方法和CCFL方法。MI指标主要关注基于均值法的亮度信息,如果融合后的图像最终含有大量噪声,则会导致亮度信息增加;RTVD方法侧重于红外信息,忽略了可见光信息。综合各种指标可以得出,本文方法既能保留图像特征信息,又能获得最佳的亮度、清晰度和细节信息。

2.3 消融实验

2.3.1 网络框架

为了验证所提出的多尺度密集连接、空洞卷积并行注意力机制、纹理对比度增强网络、分解网络的有效性,设计一系列消融实验将所提出方法降级为不同网络结构进行测试。第1组实验将多尺度密集连接换成单一尺度卷积,其余网络结构保持不变。第

2组实验将空洞卷积并行注意力机制去掉,其他网络结构保持不变。第3组实验将纹理对比度增强网络去掉,其他网络结构保持不变。第4组实验将分解网络去掉,其他网络结构不变。第5组实验则为整体网络模型。在测试数据集中选择一张Kaptein_1654融合结果作为主观评价,如图11所示。同时选取10组图像的平均值作为客观评价指标进行对比,如表5所示。

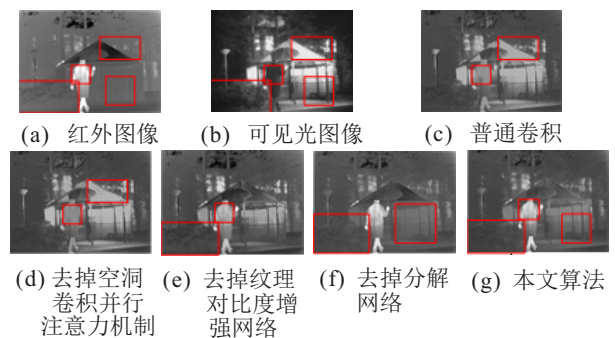


图11 各模型消融实验结果

表 5 消融实验客观评价结果

moudle	AG	VIFF
图 11(c)	4.859 3	0.422 9
图 11(d)	5.202 6	0.502 1
图 11(e)	5.811 5	0.586 8
图 11(f)	5.026 8	0.372 5
图 11(g)	6.361 9	0.623 6

由图 11 和表 5 可见,完整的框架无论从主观视觉还是客观评价指标都优于其他 4 个降级的模型框架,这也充分表明了本文网络中采用的多尺度密集连接、空洞卷积并行注意力机制、对比度增强网络和分解网络的有效性.

2.3.2 强度损失

强度损失的引入可以促使融合图像保留重要的信息,如对比度和亮度.同时,该方法有助于在融合后的图像中保持更自然的整体场景风格,进行消融实验验证其有效性,结果如图 12 所示.

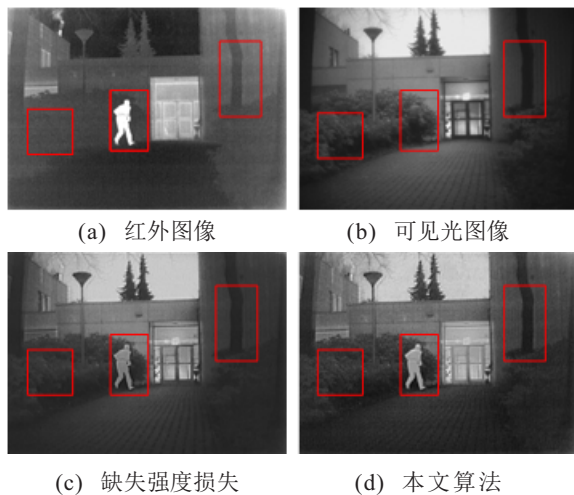


图 12 强度损失消融实验结果

由图 12 可见,缺乏强度损失导致融合后整体亮度偏低,图像表达的视觉感受不佳,一些细节信息无法充分体现,这也得到了表 6 中数值的验证.

表 6 强度损失消融实验客观评价结果

moudle	EN	AG	VIFF	SCD
missing intensity loss	6.640 0	3.330 4	0.278 7	1.696 1
ours	7.036 4	4.541 6	0.454 5	1.837 3

2.3.3 梯度损失

进行消融实验以验证梯度损失可以增强融合图像的纹理细节,结果如图 13 所示.缺乏梯度损失会导致融合后的图像出现纹理丢失和锐度降低,而使用梯度损失则可以保留原有的清晰度并增强纹理细节,这也在表 7 中得到了验证.

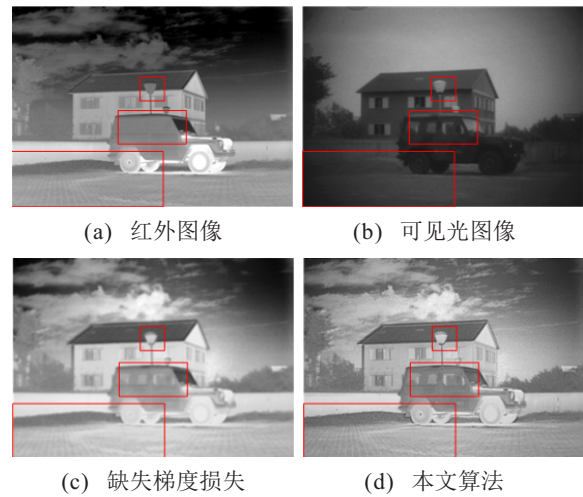


图 13 梯度损失消融实验结果

表 7 梯度损失消融实验客观评价结果

moudle	SF	AG	MSSSIM
missing gradient loss	9.981 4	2.672 7	0.909 5
ours	12.955 2	4.786 3	0.948 8

表 7 的数据表明梯度损失的加入可以进一步提升融合结果的性能,有力证明了梯度损失在融合任务中的重要性.

3 结 论

为了提高人眼或机器视觉特性的融合效果,综合利用红外和可见光图像中的显著目标和纹理细节信息,本文提出了一种空洞卷积并行注意力机制和纹理对比度增强的红外与可见光图像融合.该算法利用多尺度密集连接、空洞卷积并行注意力机制、Scharr 滤波器和深度可分离卷积等优势,在提高融合图像的亮度、对比度和纹理细节的同时突出显著目标.实验表明,本文算法的客观评价指标 EN、AG、SF、SD、MI、MSSSIM 分别提高了 7.61%、62.61%、57.31%、17.80%、7.68%、5.63%,有效融合了红外与可见光图像,达到单一图像无法提供全面信息的优势.相比于现有算法,所提出算法在特征提取、细节保留、纹理对比度提高方面具有一定的优势.然而,本文图像融合框架适用于对齐图像,这对实时、非对齐图像有限制.在未来的工作中,不仅将探索如何有效地融合未对齐的图像用于实时任务,还将集成到更先进的图像处理技术中,同时设计统一的融合框架来处理其他复杂的图像融合任务.

参考文献(References)

[1] Tang L F, Xiang X Y, Zhang H, et al. DIVFusion: Darkness-free infrared and visible image fusion[J]. Information Fusion, 2023, 91: 477-493.

- [2] Yang B, Jiang Z H, Pan D, et al. Detail-aware near infrared and visible fusion with multi-order hyper-Laplacian priors[J]. *Information Fusion*, 2023, 99: 101851.
- [3] Zhou T, Li Q, Lu H L, et al. GAN review: Models and medical image fusion applications[J]. *Information Fusion*, 2023, 91: 134-148.
- [4] 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*, 2023, 38(3): 721-728.
(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*, 2023, 38(3): 721-728.)
- [5] Zhong R J, Fu Y, Song Y S, et al. A fusion approach to infrared and visible images with Gabor filter and sigmoid function[J]. *Infrared Physics & Technology*, 2023, 131: 104696.
- [6] 闵莉, 田林林, 赵怀慈, 等. 基于多尺度和注意力机制的红外与可见光图像融合[J]. *控制与决策*, 2024, 39(1): 227-235.
(Min L, Tian L L, Zhao H C, et al. Infrared and visible image fusion based on multi-scale and attention mechanism[J]. *Control and Decision*, 2024, 39(1): 227-235.)
- [7] Wang J, Jiang M, Kong J. MDAN: Multilevel dual-branch attention network for infrared and visible image fusion[J]. *Optics and Lasers in Engineering*, 2024, 176: 108042.
- [8] Panigrahy C, Seal A, Mahato N K. Parameter adaptive unit-linking dual-channel PCNN based infrared and visible image fusion[J]. *Neurocomputing*, 2022, 514: 21-38.
- [9] 冯鑫. 基于内生长机制和卷积稀疏表示的红外与可见光图像融合[J]. *控制与决策*, 2022, 37(1): 167-174.
(Feng X. Infrared and visible light image fusion based on internal generative mechanism and convolution sparse representation[J]. *Control and Decision*, 2022, 37(1): 167-174.)
- [10] Veshki F G, Ouzir N, Vorobyov S A, et al. Multimodal image fusion via coupled feature learning[J]. *Signal Processing*, 2022, 200: 108637.
- [11] Veshki F G, Vorobyov S A. Coupled feature learning via structured convolutional sparse coding for multimodal image fusion[C]. *IEEE International Conference on Acoustics, Speech and Signal Processing*. Singapore, 2022: 2500-2504.
- [12] Liu Y, Chen X, Cheng J, et al. Infrared and visible image fusion with convolutional neural networks[J]. *International Journal of Wavelets, Multiresolution and Information Processing*, 2018, 16(3): 1850018.
- [13] 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: IEEE, 2017: 4714-4722.
- [14] 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.
- [15] 陈永, 张娇娇, 王镇. 多尺度密集连接注意力的红外与可见光图像融合[J]. *光学精密工程*, 2022, 30(18): 2253-2266.
(Chen Y, Zhang J J, Wang Z. Infrared and visible image fusion based on multi-scale dense attention connection network[J]. *Optics and Precision Engineering*, 2022, 30(18): 2253-2266.)
- [16] Wang Z S, Wu Y Y, Wang J Y, et al. Res2Fusion: Infrared and visible image fusion based on dense Res2net and double nonlocal attention models[J]. *IEEE Transactions on Instrumentation Measurement*, 2022, 71: 3139654.
- [17] Xie Q, Hu J P, Wang X C, et al. Novel and fast EMD-based image fusion via morphological filter[J]. *The Visual Computer*, 2023, 39(9): 4249-4265.
- [18] Chen J, Li X J, Wu K L. Infrared and visible image fusion based on relative total variation decomposition[J]. *Infrared Physics & Technology*, 2022, 123: 104112.
- [19] Chen J, Wu K L, Cheng Z, et al. A saliency-based multiscale approach for infrared and visible image fusion[J]. *Signal Processing*, 2021, 182: 107936.
- [20] Zhao Z X, Xu S, Zhang C X, et al. Bayesian fusion for infrared and visible images[J]. *Signal Processing*, 2020, 177: 107734.
- [21] Zhang H, Xu H, Xiao Y, et al. Rethinking the image fusion: A fast unified image fusion network based on proportional maintenance of gradient and intensity[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12797-12804.

作者简介

邸敬(1979—), 女, 副教授, 硕士, 从事图像检测识别、信号处理技术、宽带无线通信等研究, E-mail: 46891771@qq.com;

任莉(1997—), 女, 硕士生, 从事图像融合的研究, E-mail: 1427594961@qq.com;

刘冀钊(1989—), 男, 副教授, 博士, 从事混合电路的研究, E-mail: liujz@lzu.edu.cn;

郭文庆(1998—), 男, 硕士生, 从事图像融合的研究, E-mail: 344385945@qq.com;

廉敬(1983—), 男, 副教授, 博士, 从事人工智能、模式识别等研究, E-mail: lian322sec@163.com.