

基于拆分注意力残差网络的红外和 可见光图像融合算法

钱坤^{1,2}, 李湑雨¹, 李喆², 陈美杉¹

(1.海军航空大学 岸防兵学院, 山东 烟台 264000; 2.中国人民解放军 32127 部队, 辽宁 大连 116100)

摘 要:在红外和可见光图像融合算法中,图像信息的丢失始终是制约融合图像质量提升的关键问题,为此,提出了一种基于拆分注意力残差网络的红外和可见光图像融合算法,使用带有拆分注意力模块的深层残差网络拓展感受野和提高跨通道信息融合能力,运用平滑最大值单元函数作为激活函数进一步提升网络性能;特征提取后运用零相位分量分析和归一化算法得到融合权重后完成图像融合。实验结果表明,融合后的图像细节丰富,边缘锐利;在峰值信噪比、结构相似性指数度量和基于梯度的融合性能等指标上与经典的 6 种算法相比均有不同程度提升。

关 键 词:图像融合;拆分注意力;残差网络;零相位分量分析;平滑最大值单元函数
中图分类号:TP391 **文献标志码:**A **文章编号:**1000-2758(2022)06-1404-10

近年来,红外与可见光融合技术一直是研究的热点领域。红外成像是景物热辐射成像,适宜在低光照、强闪光和浓雾烟幕等极端条件下使用,缺点是容易丢失图像细节信息,成像不符合人眼视觉偏好等。可见光成像是物体的反射成像,特点是具有丰富的色彩信息和大量的纹理细节,符合人类视觉习惯,但成像质量受制于光照和大气条件。将红外与可见光进行融合可以充分发挥二者优势,弥补单波段成像弊端,特别是在军事领域,融合图像对于分辨假目标(诱饵),提升全天候作战能力和提高精确打击能力等有着重要意义。

在传统的红外与可见光融合算法中,比较有代表性的有:①基于多尺度变换(multi-scale transformation, MST)理论的算法,包括基于交叉双边滤波器算法(cross bilateral filter, CBF)^[1]、各向异性融合算法(anisotropic diffusion fusion, ADF)^[2]等;②基于梯度信息理论的算法,如梯度转移融合算法(gradient transfer fusion, GTF)^[3]、基于梯度结构相似性的多模态图像融合(GSF)^[4]等;③基于显著性理论的算法,如潜在低秩表示算法(latent low-rank representation, LatLRR)^[5]等;④基于稀疏表示的算法,如文献[6-7]所示等;⑤上述算法的组合算法,如多尺度分解算法(multi-scale decomposition, MSD)^[8]等。

近年来,随着深度学习算法的不断发展,相关成果也应用在了红外和可见光图像融合领域,取得了较好的效果。比较有代表性的有 Prabhakar 等^[9]提出的基于卷积神经网络(convolutional neural network, CNN)的融合方法(deepfuse),其最终的融合图像由一个 3 层的 CNN 解码重建而来,能够适应多种融合任务,但网络结构比较简单,难以充分保留源图像中的信息,融合质量一般。Zhang 等^[10]提出了基于卷积神经网络的通用图像融合框架(IFCNN),利用 2 个卷积层从多个输入图像中提取显著的图像特征,根据输入图像的类型选择适当规则进行卷积特征融合,最后使用 2 个卷积层重构融合特征,并生成融合图像,其优势在于多任务通用,但对于特定任务融合效果并非最优。Li 等^[11]提出了深度学习框架算法(deep learning framework, DLF),利用牛津大学计算机视觉组(visual geometry group, VGG)提出的 VGG 网络进行图像融合,其中

收稿日期:2022-03-14

作者简介:钱坤(1986—),海军航空大学博士研究生,主要从事计算机视觉及机器学习研究。

通信作者:李湑雨(1993—),海军航空大学讲师,主要从事机器学习及图像处理研究。e-mail:2823426785@qq.com

的 VGG-19 网络在深度上胜过 CNN,但融合方法相对简单,对于图像的深度信息挖掘不够,而且随着网络的加深,存在网络退化的现象。随后 Li 等^[12]又提出了一种基于残差网络(residual network, ResNet)的红外和可见光图像融合框架,旨在改善深度网络性能,提高对图像细节信息的利用能力。文献[13]提出了一种将多尺度局部极值分解与 ResNet152 残差网络相结合的融合方法,使用多尺度局部极值分解将源图像分解为近似图像和细节图像,采用 ResNet152 网络提取显著性特征,设计了基于显著特征图和能量显著图加权的融合规则,取得较好效果,但算法复杂,时间开销较大。

针对深层网络退化与红外和可见光图像融合算法中的图像信息保留不完整等问题,进一步挖掘残差网络性能潜力和拓展应用范围,本文提出了一种基于拆分注意力残差网络(split-attention residual network, ResNeSt)^[14]的红外和可见光图像融合算法,这是一种端到端的网络模型,输入红外和可见光图像,在特征提取、权重计算和图像重建等过程后,直接输出融合后的图像。

1 基于 ResNeSt 的红外和可见光图像图像融合算法

1.1 ResNeSt 网络框架

基于 ResNeSt 的红外和可见光图像融合算法框架如图 1 所示。将红外图像和可见光图像分别送入训练过的 ResNeSt-50 网络进行特征提取,得到各自的特征图;对特征图使用零相位分量分析(zero-phase component analysis, ZAC)方法进行数据白化,使其投影在同一子空间;利用 L_1 范数获得初始权重图,然后使用双三次插值算法进行上采样并使用 softmax 函数进行权重归一化,得到和原图像大小一致的权重矩阵;运用加权平均策略,对原始红外和可见光图像进行加权平均,得到最终的融合图像。

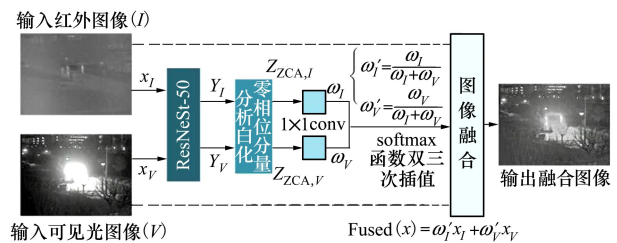


图 1 算法框架

1.2 拆分注意力残差网络

在深度学习中,神经网络的深度和宽度对于性能的影响非常明显,但网络的层级不能通过简单的堆叠进行加深,原因是随着网络的加深会发生网络退化的现象,为此 He 等^[15]提出了 ResNet,通过加入跳跃连接,保证网络加深的最坏结果也不会比原网络性能更差,这个过程不显著增加计算负担,且网络依然可以通过反向传播进行训练学习,ResNet 模块结构见图 2。通过对残差模块的反复堆叠,可以产生深层的、无退化现象的残差网络,以便从输入数据中提取更为丰富的特征。ResNet 解决了网络退化问题,但通过进一步研究发现,其性能依旧有提升的空间,ResNet 主要存在的不足为:①感受野相对有限;②缺乏跨通道的信息融合,即缺少上下文之间的交流^[16-17]。

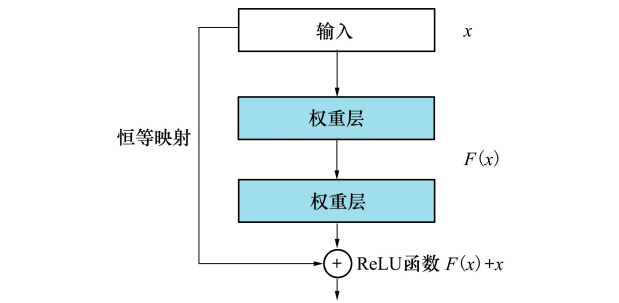


图 2 残差模块

为此,本文使用 ResNeSt 对输入的源图像进行特征提取,具体结构见图 3。ResNeSt 将特征图划分为 K 个基群(cardinal),基群内分为更细粒度 R 个切片(split),并在基群内加入拆分注意力模块(split-

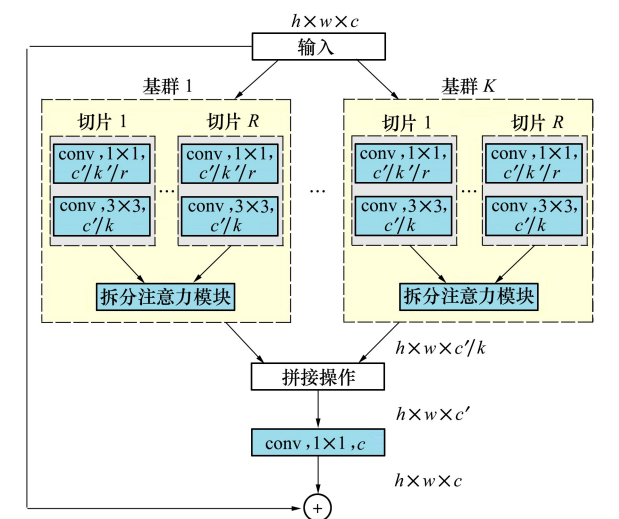


图 3 拆分注意力残差模块

attention block), 结构见图 4。每个基群的特征通过上下文信息确定的权重进行组合。ResNeSt 在分支网络中使用了不同尺寸的卷积核, 可提供不同尺度的感受野, 拆分注意力模块中的全局池化层能够进一步扩展感受野, Dense 模块中的全连接层增加了计算的非线性, 也能更好地拟合跨通道的相关性。

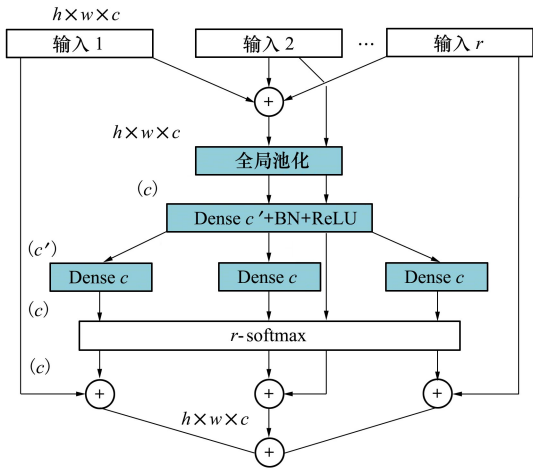


图 4 拆分注意力模块

ResNeSt 网络结构通过拆分共有 KR 个切片, 输入 x 经过切片的处理, 进行一系列的特征提取后得到 $\{F_1, F_2, \dots, F_{KR}\}$, 最后输出特征 $U_i = F_i(x), i \in \{1, 2, \dots, KR\}$ 。而后将同一基群中的特征输入拆分注意力模块, 特征求和以获得联合特征表示 $\hat{U}^k$, 见公式(1)。

$$\hat{U}^k = \sum_{j=R(k-1)+1}^{RK} U^j, \hat{U}^k \in \mathbf{R}^{H \times W \times C/K}, k \in (1, 2, \dots, K) \tag{1}$$

式中: k 表示第 k 个基群, j 表示第 j 个切片。

之后, 通过跨空间维度的全局平均池化, 获得每个通道上的统计信息, 作为全局上下文信息, 记作 $s^k, s^k \in \mathbf{R}^{C/K}$, 第 c 个通道上的信息可以表示为公式(2)。

$$s_c^k = \frac{1}{H \times W} \sum_i^H \sum_j^W \hat{U}_c^k(i, j) \tag{2}$$

而后, 再将第 c 个通道上同一基群中所有切片信息进行求和, 表示为 V_c^k , 见公式(3)。

$$V_c^k = \sum_i^R \alpha_i^k(c) U_{R(k-1)+i} \tag{3}$$

式中, $\alpha_i^k(c)$ 表示赋值权重, 具体见公式(4)。

$$\alpha_i^k(c) = \begin{cases} \frac{\exp[\zeta_i^c(s^k)]}{\sum_{j=0}^R \exp[\zeta_j^c(s^k)]}, R > 1 \\ \frac{1}{1 + \exp[-\zeta_i^c(s^k)]}, R = 1 \end{cases} \tag{4}$$

如果每个基群的切片数 R 大于 1, 使用 softmax 函数作为每个切片的权重, 如果 R 等于 1, 则使用 sigmoid 函数作为切片的权重, ζ_i^c 表示所用 softmax 或 sigmoid 函数。

最后将所有基群输出的特征进行拼接, 得到最后的特征 $V, V = \text{Concat}\{V^1, V^2, \dots, V^K\}$, 与 ResNet 中含有一个恒等映射相同, 在 ResNeSt 的最终输出里还包含了直接输入 x , 所以最终输出可以表示为 $Y = V + x$ 。

本文所用的拆分注意力残差网络 ResNeSt-50 在结构上与 ResNet-50 保持一致, 共有 50 层卷积, 分为 16 组残差结构, 见图 5。用拆分注意力残差模块替换原有残差模块, 得到 ResNeSt-50 网络, 网络通过 ImageNet 2012 数据集^[18]进行预训练。

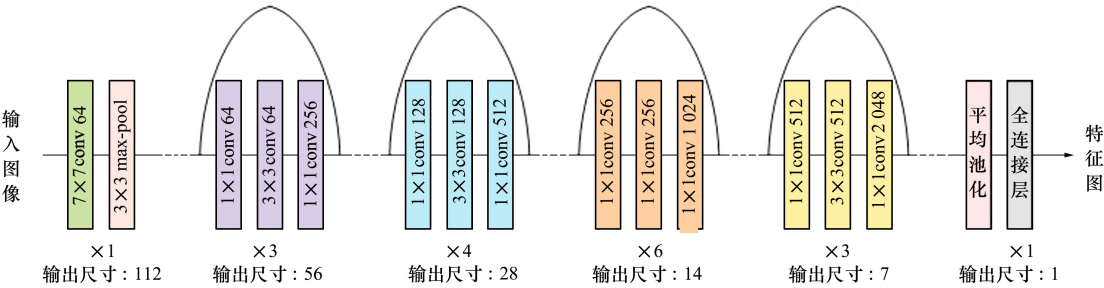


图 5 ResNet-50 网络结构

1.3 平滑最大值单元函数

激活函数在神经网络有效性和训练过程中起到

非常重要的作用, 优化激活函数可以显著提升网络的综合性能。在 ResNet-50 网络的残差模块中, 前

向神经网络与跳跃连接叠加后使用线性整流函数 (rectified linear unit, ReLU) 作为激活函数, ReLU 函数见公式(5)。ReLU 函数优势在于简单易用, 计算效率高, 网络收敛快, 但弊端在于当输入趋近于零或者为负时, 函数梯度为零, 网络无法反向传播, 不能继续“学习”。

$$f_{\text{ReLU}}(x) = \max(0, x) \quad (5)$$

在优化 ReLU 函数过程中, 提出了多种激活函数, 如带泄露线性整流函数 (leaky ReLU)^[19]、指数线性单元函数 (exponential linear unit, ELU)^[20] 等, 文献[21]提出了一种平滑最大值单元函数 (smooth maximum unit, SMU), 能够有效避免 ReLU 函数的缺陷, 在分类、目标检测和语义分割等应用领域被验证效果优于 ReLU。SMU 函数见公式(6)。

$$f_{\text{SMU}}(x, \alpha; \mu) = \frac{(1 + \alpha)x + (1 - \alpha)\text{xerf}[\mu(1 - \alpha)x]}{2} \quad (6)$$

其中, $\text{erf}(x)$ 为高斯误差函数, 见公式(7)

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (7)$$

式中, α, μ 为超参, 本文中选取 $\alpha = 0.25, \mu = 0.5$ 。

1.4 零相位分量分析和归一化

经过 ResNeSt 网络模型输出的是多维特征向量, 运用 ZCA 方法可以将高维特征向量通过向量矩阵转换后进行去相关性操作^[22], 同时, 保留主要信息, 忽略掉偏差大的信息。假设在 ResNeSt 网络模型提取的是 m 条 n 维的向量数据, 将数据组成 n 行 m 列的矩阵 $\mathbf{Y}$, ZCA 白化计算过程如下:

① 计算矩阵 $\mathbf{Y}$ 的协方差矩阵 Σ , 如公式(8)

$$\Sigma = \frac{1}{m} \sum_{i=1}^m [\mathbf{y}^{(i)}][\mathbf{y}^{(i)}]^T \quad (8)$$

② 对 Σ 进行奇异值分解, 得到向量 $\mathbf{U}$, 计算得到 $\mathbf{U}^T \mathbf{y}$, 即为数据经过坐标轴旋转之后得到的矩阵 $\mathbf{Z}$, 如公式(9)

$$\mathbf{Z} = \mathbf{U}^T \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^T \mathbf{y} \\ \mathbf{u}_2^T \mathbf{y} \\ \vdots \\ \mathbf{u}_n^T \mathbf{y} \end{pmatrix} \quad (9)$$

③ 对 $\mathbf{Z}$ 进行 PCA 白化, 得到 $\mathbf{Z}_{\text{PCA}, i}$

$$\mathbf{Z}_{\text{PCA}, i} = \frac{\mathbf{Z}_i}{\sqrt{\lambda_i}} \quad (10)$$

式中, λ_i 是 $\mathbf{Z}$ 协方差矩阵对角元素的值。

④ 将 $\mathbf{Z}_{\text{PCA}, i}$ 左乘 $\mathbf{U}$, 得到经过白化后的特征图 $\mathbf{Z}_{\text{PCA}}$ 。

将红外和可见光图像的特征图 $\mathbf{Z}_{\text{PCA}}$ 通过一个大小为 1×1 的卷积核, 得到初始权重 ω_I (红外图像) 和 ω_V (可见光图像), 这样一方面可以实现数据降维, 减少参数量, 实现对不同尺度特征进行尺寸的归一化, 另一方面也能起到跨通道特征融合的作用。

1.5 图像融合

特征图归一化后运用 softmax 函数, 得到最终的权重图 ω'_I (红外图像) 和 ω'_V (可见光图像), 如公式(11)所示

$$\begin{cases} \omega'_I = \frac{\omega_I}{\omega_I + \omega_V} \\ \omega'_V = \frac{\omega_V}{\omega_I + \omega_V} \end{cases} \quad (11)$$

归一化的特征图通过双三次插值算法缩放放到与输入图像一致大小, 并按照各自权重通过加权平均算法得到最终的融合图像, 如公式(12)所示。

$$\text{Fused}(x) = \omega'_I x_I + \omega'_V x_V \quad (12)$$

2 实验结果与分析

2.1 数据集及实验环境

在图像融合研究中, 需要使用经过配准后的红外和可见光图像作为输入数据, 但公开并配准的双光数据集选择非常有限, 本文采用 TNO^[23] 中的 40 对红外和可见光图像作为主观视觉实验的输入数据, 同时, 为检验算法对于红外和彩色可见光图像的融合能力, 还选取了部分红外和彩色可见光图像对进行融合实验。在多指标客观评价中, 选取文献[24]的红外和可见光图像融合基准 (visible and infrared image fusion benchmark, VIFB) 中的 21 对图像进行测试, 其中 TNO 数据集示例见图 6, VIFB 数据集示例见图 7。

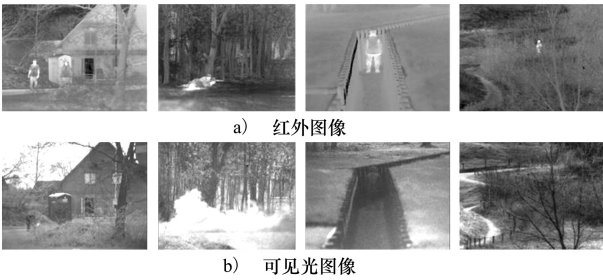


图 6 TNO 数据集示例

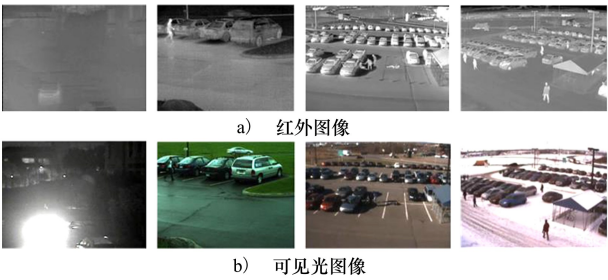


图 7 VIFB 测试集示例

算法所用的 ResNeSt-50 网络已在 ImageNet 2012 数据集上进行了训练,但为进一步优化网络性能,使其在图像融合计算中更具针对性,还对 ResNeSt-50 进行了迁移学习,训练数据选自 TNO 数据集,按照 3 : 1 的比例划分训练集和测试集,由于

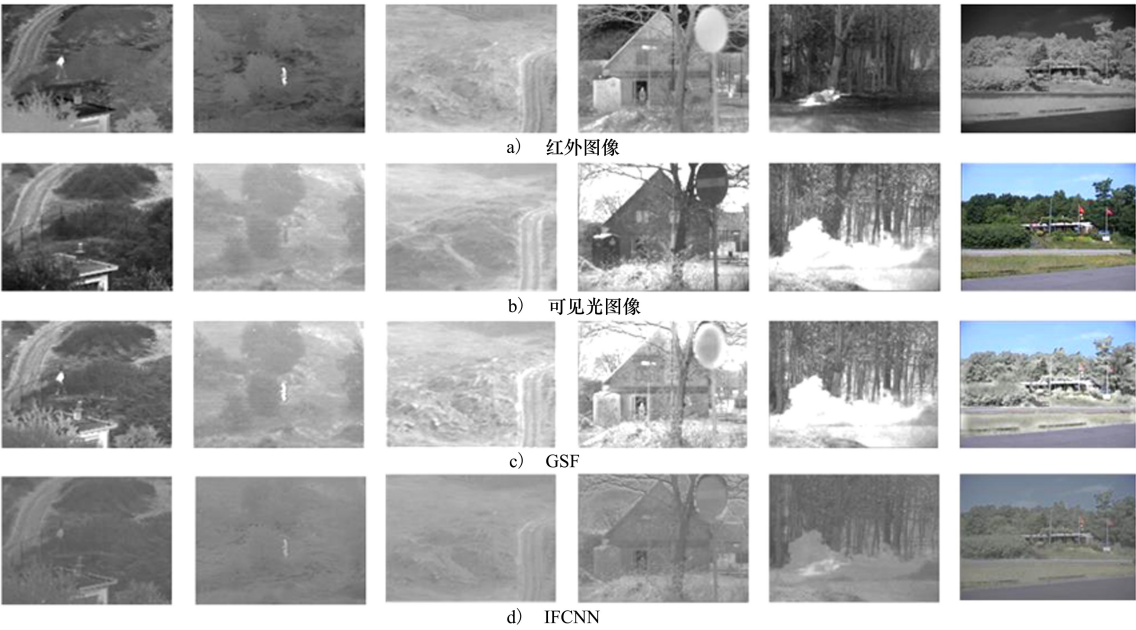
表 1 实验环境

软(硬)件	版本(型号)
操作系统	Windows10
CPU	i7-9700 3.00 GHz
GPU	NVIDIA 2060 SUPER
GPU 显存	8 GB
内存	16G
加速环境	CUDA v11.0、CuDNN v7.6
显卡驱动版本	27.21.14.6172
MATLAB	R2020a
Pytorch 版本	1.7
Python 版本	3.8.5

训练数据数量较少,需要对训练集数据进行数据增广处理^[25],训练集中图像尺寸有 280×280,360×270,595×328 等多种,为避免裁剪过小导致训练出的模型对于大尺寸物体识别能力不足,统一按照 180×180 大小,步长 20 进行裁剪,经人工筛选后最终得到图像 1 120 对。迁移学习中使用 Pytorch 深度学习框架,使用 Adam 优化器,学习速率为 2×10^{-4} ,每批训练图像为 8 幅,训练 100 轮。实验环境见表 1。

2.2 主观视觉评价

将训练后的网络在 TNO 数据集的测试图像上进行实验,选取其中的 6 幅图像进行主观视觉对比,对比对象分别是 GSF 算法^[4]、IFCNN 算法^[10]、LatLRR 算法^[5]、VGG 算法^[22]、DLF 算法^[11] 和 ResNet 算法^[12]。输入图像和各算法融合后图像见图 8。



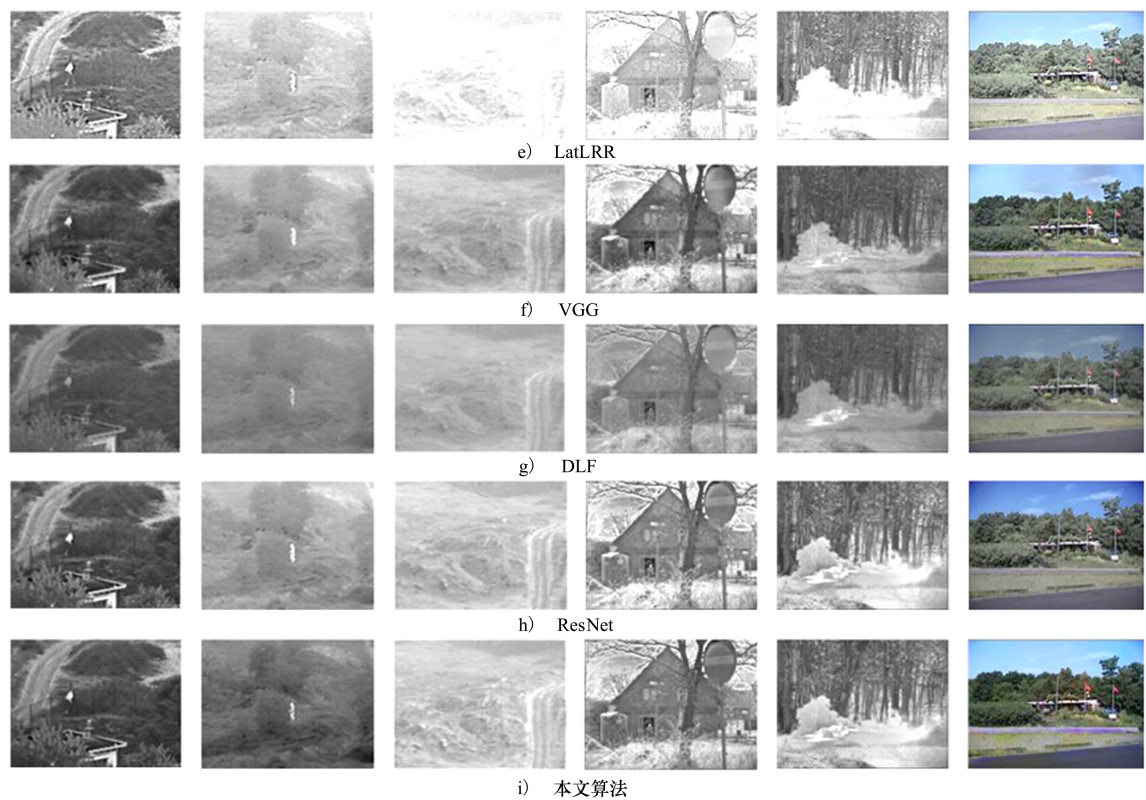


图 8 红外与可见光图像融合主观评价对比

如图 8c) 所示,GSF 算法在第 4 列中的交通标志边缘处融合失真,出现高亮光晕,标志牌中央的横杠几乎不可见,第 5 列中对于红外图像信息利用不佳,烟幕中的射手未能有效融合;第 6 列中对于色彩的还原不准确,整体偏灰白。如图 8d) 所示,IFCNN 算法在整体对比度上相对偏低,导致纹理细节还原质量一般,在第 1 列灌木丛和第 3 列地形起伏处尤其明显,第 5 列色彩还原不够准确,明显偏灰。如图 8e) 所示,LatLRR 算法处理后的图像整体过曝,第 3 列上表现尤其突出,第 5 列中烟幕遮挡下的射手未能有效融合显示,丢失了红外图像的热信息;在第 6 列上伪影明显,观感不自然。如图 8f) 所示,VGG 算法在第 5 列中,对于烟幕的融合效果不佳,画面大面积过曝。如图 8g)~8h) 所示,DLF 算法和 ResNet 算法整体效果较好,图像细节保留比较完整,从第 4~5 列中的树木效果看,ResNet 算法优于 DLF 算法,其纹理更清晰,在彩色图像融合上 2 幅图像均有不同程度失真,DLF 算法偏灰,ResNet 算法颜色较深。在图 8i) 所示,本文算法在图像细节的提取和保留上表现最好,纹理边缘清晰锐利,在第 3 列中地形起伏处表现最为明显;在对烟幕的处理上,既保留了烟

幕的轮廓,也清晰显示了被烟幕遮挡的人;同时,本文算法对于色彩的还原最为准确,相较其他算法表现优异。

2.3 多指标客观评价

客观评价需要计算相关指标,旨在定量模拟人眼视觉对图像质量的感知。相关指标种类繁多,本文选取比较有代表性的 4 种:①基于信息理论的峰值信噪比 (peak signal-to-noise ratio, PSNR)^[26],其值越大说明融合图片和原始图片之间差别就越小,细节保留越完整;②基于结构相似性理论的结构相似性指数度量 (structural similarity index measure, SSIM)^[27],SSIM 数值在 [0,1] 之间,数值越大代表和原来 2 幅图像越接近;③基于图像特征的空间频率 (spatial frequency, SF)^[28],空间频率数值越大越好,越大图像质量越高,观感更清晰;④基于梯度的融合性能 (gradient-based fusion performance, $Q^{AB/F}$)^[29], $Q^{AB/F}$ 值越大图像质量越高,保留的原始图像的信息也就越多。

在 4 种常用指标中,只有 SF 的计算不依赖原始输入图像,即只需计算融合图像的 SF 后进行比较即可,其他 3 种指标都需要分别计算融合图像与原始

输入的红外图像和可见光图像的指标数值,再将 2 个数值以适当方式求和得到最终融合图像的指标数

值,具体实验结果如图 9 所示。

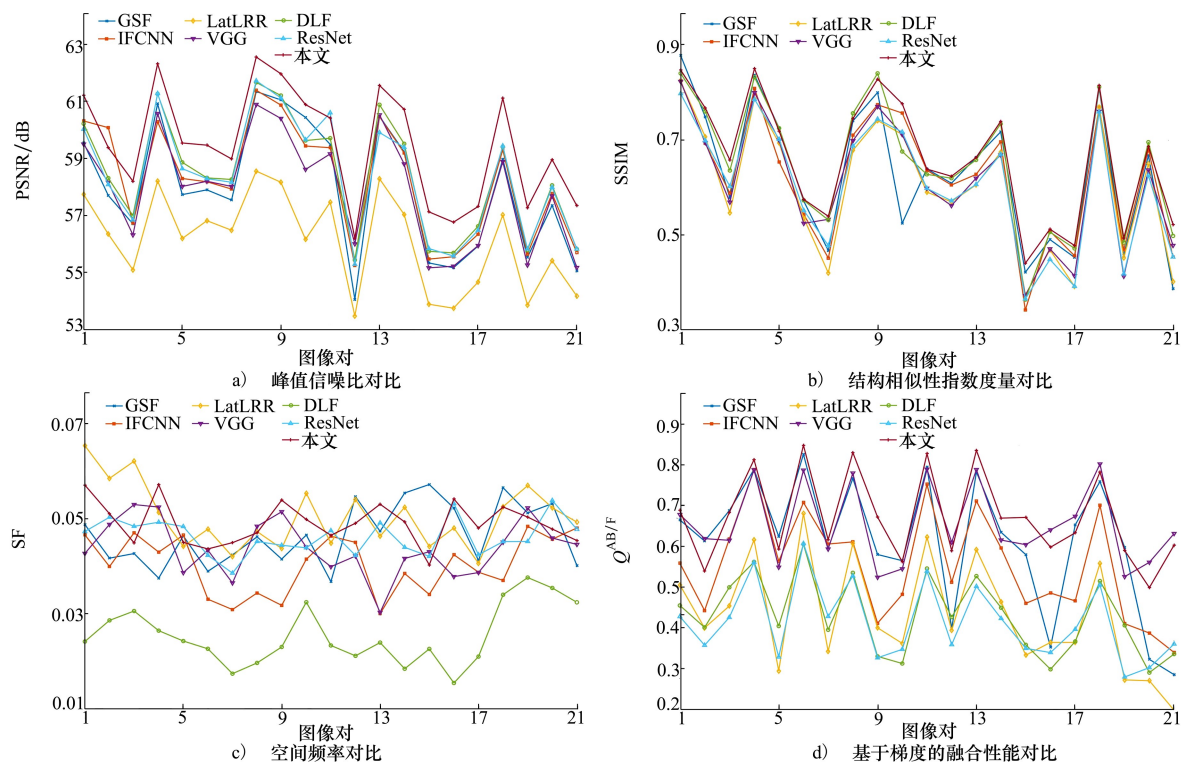


图 9 7 种算法的 4 项客观指标对比

实验结果表明,在 PSNR、SSIM 和 $Q^{AB/F}$ 3 项评价指标中,本文提出的基于 ResNeSt 的融合算法表现最优,对比其他 6 种算法分别至少提高了 1.78%, 2.00% 和 3.10%, 在 SF 指标中排名第二, 仅落后 LatLRR 算法 2.73%, 但对比传统 ResNet 算法, 也有 6.52% 的提升。同时, 为衡量各算法的执行效率, 还将融合图像时间开销进行了比较, 本文算法继承了传统 ResNet 算法执行效率高的优势, 在算法复杂度提高的情况下, 平均每对图像的融合时间相对 ResNet 算法小幅增加了 4.5%, 全部 21 组融合图像客观评价结果平均值如表 2 所示。

2.4 消融实验

为验证算法的合理性, 设计了 3 组消融实验, 实验 1 是使用 ResNeSt 网络进行特征提取, 内部使用 ReLU 函数作为激活函数; 实验 2 使用传统的 ResNet 网络进行特征提取, 在其内部使用 SMU 函数作为激活函数; 实验 3 使用 ResNeSt 网络进行特征提取, 并在内部使用 SMU 函数作为激活函数, 即本文完整算法。消融实验也分为主观视觉评价和多指标客观评价两部分, 选取了测试集中最具代表性的一组图像及其融合后图像进行具体说明, 消融实验图像融合结果见图 10。

图 10a) 为输入的红外图像, 图 10b) 为输入的可见光图像, 图 10c) 为实验 1 所得融合图像, 图像保留了烟幕中的战士, 能够呈现树干的纹理细节; 图 10d) 为实验 2 所得融合图像, 也较好地完成了融合任务, 细节同样比较丰富, 战士的轮廓边缘也更锐利; 图 10e) 为实验 3 所得融合图像, 效果最为理想, 树干的纹理清晰, 树丛细节保留更加丰富完整。表 3 为消融实验 4 项客观评价指标对比, 4 项指标中实验 3 均优于实验 1 和实验 2, 表明算法的多处改进

表 2 21 组融合图像 5 项客观评价结果平均定量值

算法	PSNR/dB	SSIM	SF	$Q^{AB/F}$	时间/s
GSF	57.971	0.628	0.047	0.613	11.276
IFCNN	58.261	0.622	0.040	0.553	33.638
LatLRR	56.133	0.600	0.050	0.433	198.515
VGG	57.926	0.611	0.044	0.653	28.679
DLF	58.443	0.650	0.025	0.429	9.977
ResNet	58.376	0.604	0.046	0.414	3.754
本文	59.484	0.663	0.049	0.673	3.923

均能对最终结果产生正向优化。

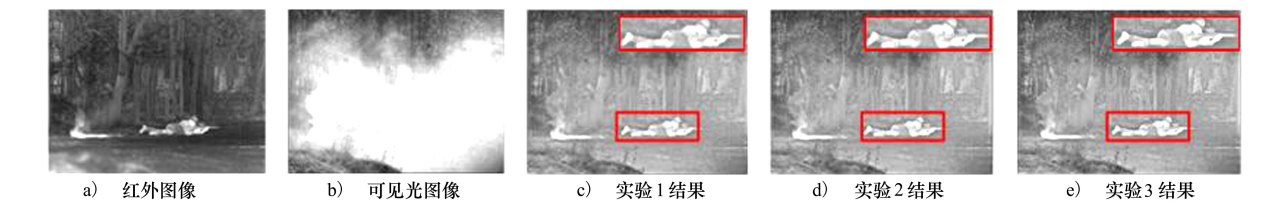


图 10 消融实验示例

表 3 消融实验 4 项客观评价指标对比				
组别	PSNR	SSIM	SF	$Q^{AB/F}$
实验 1	57.891	0.594	0.043	0.660
实验 2	57.599	0.609	0.043	0.651
实验 3	58.476	0.617	0.045	0.664

3 结 论

本文充分利用深层网络对于图像特征提取的优势,提出了基于拆分注意力残差网络的红外和可见

光图像融合算法,算法利用带有拆分注意力模块的深层残差网络分别提取红外和可见光图像多尺度特征,通过后续的权重计算和图像重建,得到融合后的图像。与经典的 6 种融合算法进行主客观对比,本算法融合图像细节丰富,既体现了红外图像中的热信息,又保留了可见光图像的纹理细节,在多项定量评价中表现突出。后续工作中,将继续优化网络结构及超参寻优,进一步提升融合图像质量,同时,针对配准的双光数据较少的实际,还将着手建立更为丰富的数据集,以方便开展后续研究工作。

参考文献:

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

[2] BAVIRISETTI D P, DHULI R. Fusion of Infrared and visible sensor images based on anisotropic diffusion and Karhunen-Loeve transform[J]. IEEE Sensors Journal, 2015, 16(1): 203-209

[3] 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: 100-109

[4] FU Z Z, ZHAO Y F, XU Y W, et al. Gradient structural similarity based gradient filtering for multi-modal image fusion[J]. Information Fusion, 2020, 53: 251-268

[5] LI H, WU X J. Infrared and visible image fusion using latent low-rank representation[EB/OL]. (2018-4-24) [2022-11-11]. <https://arxiv.org/abs/1804.08992v4>.

[6] 裴佩佩, 杨艳春, 党建武, 等. 基于滚动引导滤波器和卷积稀疏表示的红外与可见光图像融合方法[J]. 激光与光电子学进展, 2022, 59(12): 56-63

PEI Peipei, YANG Yanchun, DANG Jianwu, et al. Infrared and visible image fusion method based on rolling guidance filter and convolution sparse representation[J]. Laser & Optoelectronics Progress, 2022, 59(12): 56-63 (in Chinese)

[7] 杨勇, 李露奕, 黄淑英, 等. 自适应字典学习的卷积稀疏表示遥感图像融合[J]. 信号处理, 2020, 36(1): 125-138

YANG Yong, LI Luyi, HUANG Shuying, et al. Remote sensing image fusion with convolutional sparse presentation based on adaptive dictionary learning[J]. Journal of Signal Processing, 2020, 36(1): 125-138 (in Chinese)

[8] ZHOU Z, WANG B, LI S, et al. Perceptual fusion of infrared and visible images through a hybrid multi-scale decomposition with gaussian and bilateral filters[J]. Information Fusion, 2016, 30: 15-26

[9] PRABHAKAR K R, SRIKAR V S, BABU R V. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure image pairs[C]//Proceeding of the 2017 IEEE International Conference on Computer Vision, 2017

[10] ZHANG Y, LIU Y, SUN P, et al. IFCNN: a general image fusion framework based on convolutional neural network[J]. Infor-

mation Fusion, 2020, 54: 99-118

- [11] LI H, WU X J, KITTTLER J. Infrared and visible image fusion using a deep learning framework[C]//Proceeding of the 24th International Conference on Pattern Recognition, 2018
- [12] LI H, WU X J, DURRANI T S. Infrared and visible image fusion with resnet and zero-phase component analysis[J]. Infrared Physics & Technology, 2019, 102: 103039
- [13] 陈广秋, 王帅, 黄丹丹, 等. 基于多尺度局部极值分解与 ResNet152 的红外与可见光图像融合[J]. 光电子·激光, 2022, 33(3): 283-295
CHEN Guangqiu, WANG Shuai, HUANG Dandan, et al. Infrared and visible image fusion based on multiscale local extrema decomposition and resnet152[J]. Journal of Optoelectronics · Laser, 2022, 33(3): 283-295 (in Chinese)
- [14] ZHANG H, WU C, ZHANG Z, et al. ResNeSt: split-attention networks[C]//Proceeding of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022
- [15] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognitio[C]//Proceeding of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016
- [16] JIE H, LI S, GANG S, et al. Squeeze-and-excitation networks[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023
- [17] LI X, WANG W, HU X, et al. Selective kernel networks[C]//Proceeding of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020
- [18] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database[C]//Proceeding of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009: 248-255
- [19] MAAS A, HANNUN A, NG A. Rectifier nonlinearities improve neural network acoustic models[C]//Proceeding of the 30th International Conference on Machine Learning, 2013
- [20] CLEVERT D, UNTERTHINER T, HOCHREITER S. Fast and accurate deep network learning by exponential linear units (ELUs)[EB/OL]. (2015-11-23)[2022-11-11]. <https://arxiv.org/abs/1511.07289>.
- [21] BISWAS K, KUMAR S, BANERJEE S, et al. SMU: smooth activation function for deep networks using smoothing maximum technique[EB/OL]. (2021-11-23)[2022-11-11]. <https://arxiv.org/abs/2111.04682>
- [22] 马旗, 朱斌, 张宏伟. 基于 VGG 网络的双波段图像融合方法[J]. 激光与红外, 2019, 49(11): 7
MA Qi, ZHU Bin, ZHANG Hongwei. Dual-band image fusion method based on VGGNet[J]. Laser & Infrared, 2019, 49(11): 7
- [23] TOET A. The TNO multiband image data collection[J]. Data in Brief, 2017, 15: 249-251
- [24] ZHANG X, YE P, XIAO G. VIFB: a visible and infrared image fusion benchmark[C]//Proceeding of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020
- [25] 徐东东. 基于无监督深度学习的红外与可见光图像融合方法研究[D]. 长春: 中国科学院大学, 2020: 72-74
XU Dongdong. Research on infrared and visible image fusion based on unsupervised deep learning[D]. Changchun: Chinese Academy of Sciences, 2020: 72-74 (in Chinese)
- [26] JAGALINGAM P, HEGDE A V. A review of quality metrics for fused image[J]. Aquatic Procedia, 2015, 4: 133-142
- [27] WANG Z, BOVIK A C, SHEIKH H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Trans on Image Processing, 2004, 13(4): 600-612
- [28] ESKICIOGLU A M, FISHER P S. Image quality measures and their performance[J]. IEEE Trans on Communications, 1995, 43(12): 2959-2965
- [29] XYDEAS C S, PV V. Objective image fusion performance measure[J]. Military Technical Courier, 2000, 56(4): 181-193

Infrared and visible image fusion algorithm based on split-attention residual networks

QIAN Kun^{1,2}, LI Tianyu¹, LI Zhe², CHEN Meishan¹

(1.Coastal Defense Academy, Naval Aeronautical University, Yantai 264000, China;
2.No. 32127 Unit of PLA, Dalian 116100, China)

Abstract: In the infrared and visible image fusion algorithm, the loss of image information is always the key problem to restrict the improvement of fusion image quality. Therefore, an infrared and visible image fusion algorithm based on split-attention residual network is proposed. The residual network with split-attention block is used to expand the receptive field and improve the ability of cross-channel information fusion, and the smooth maximum unit function is used as the activation function to further improve the network performance. Then, the extracted feature map is whitened by using the zero-phase component analysis method to project it in the same subspace, and the initial weight map is obtained by L_1 -norm. Then the bicubic interpolation algorithm is used for upsampling and the softmax function is used for weight normalization to obtain the weight matrix consistent with the size of the original image. Finally, the weighted average strategy is used to weighted average the original infrared and visible images to obtain the final fused image. In order to verify the performance of the algorithm, the subjective and objective evaluation is compared with the six classical fusion algorithms in experiment. In the subjective evaluation, the fusion image of the present algorithm is in detail, which not only reflects the thermal information in the infrared image, but also retains the texture details of the visible image, with sharp edges, high color restoration and natural appearance. Among the three evaluation indexes of peak signal-to-noise ratio, structural similarity index measure and gradient-based fusion performance, the present algorithm improves at least 1.78%, 2.00% and 3.10% by comparing with the other six algorithms.

Keywords: image fusion; split-attention; residual network; zero-phase component analysis; smooth maximum unit

引用格式:钱坤, 李湉雨, 李喆, 等. 基于拆分注意力残差网络的红外和可见光图像融合算法[J]. 西北工业大学学报, 2022, 40(6): 1404-1413

QIAN Kun, LI Tianyu, LI Zhe, et al. Infrared and visible image fusion algorithm based on split-attention residual networks [J]. Journal of Northwestern Polytechnical University, 2022, 40(6): 1404-1413 (in Chinese)