

# 模态自适应的红外与可见光图像融合

曲海成, 王宇萍, 高健康, 赵思琪

(辽宁工程技术大学 软件学院, 辽宁 葫芦岛 125105)

**摘要:** 为解决低照度和烟雾等恶劣环境条件下融合图像目标对比度低、噪声较大的问题, 提出一种模态自适应的红外与可见光图像融合方法 (mode adaptive fusion, MAFusion)。首先, 在生成器中将红外图像与可见光图像输入自适应加权模块, 通过双流交互学习二者差异, 得到两种模态对图像融合任务的不同贡献比重; 然后, 根据各模态特征的当前特性自主获得各模态特征的相应权重, 进行加权融合得到融合特征; 最后, 为了提高模型的学习效率, 补充融合图像的多尺度特征, 在图像融合过程中加入残差块与跳跃残差组合模块, 提升网络性能。在 TNO 和 KAIST 数据集上进行融合质量测评, 结果表明: 主观评价上, 提出的方法视觉效果良好; 客观评价上, 信息熵、互信息和基于噪声的评价性能指标均优于对比方法。

**关键词:** 图像融合; 模态自适应; 对抗生成网络; 残差网络

**中图分类号:** TP391      **文献标志码:** A      **文章编号:** 1001-8891(2022)03-0268-09

## Mode Adaptive Infrared and Visible Image Fusion

QU Haicheng, WANG Yuping, GAO Jiankang, ZHAO Siqi

(School of Software, Liaoning Technical University, Huludao 125105, China)

**Abstract:** To solve the problems of low contrast and high noise of fused images in low illumination and smoky environments, a mode-adaptive infrared and visible image fusion method (MAFusion) is proposed. Firstly, the infrared and visible images are input into the adaptive weighting module in the generator, and the difference between them is learned through two streams interactive learning. The different contribution proportion of the two modes to the image fusion task in different environments is obtained. Then, according to the characteristics of each modal feature, the corresponding weights of each modal feature are obtained independently, and the fusion feature is obtained by weighted fusion. Finally, to improve the learning efficiency of the model and supplement the multi-scale features of the fused image, a residual block and jump connection combination module are added to the image fusion process to improve the network performance. The fusion quality was evaluated using the TNO and KAIST datasets. The results show that the visual effect of the proposed method is good in subjective evaluation, and the performance indexes of information entropy, mutual information, and noise-based evaluation are better than those of the comparison method.

**Key words:** image fusion, mode adaptive, GAN, ResNet

## 0 引言

图像融合属于信息融合的一部分, 其目的是针对同一场景, 将来自不同传感器的图像联合起来, 生成一幅信息丰富的融合图像, 以便于后续处理。CT、磁共振成像、可见光图像与红外图像属于不同成像类型的源图像, 在这些源图像组合中, 本文重

点研究红外与可见光图像融合, 原因如下: 第一, 红外图像通过捕获物体热辐射信息成像, 可见光图像通过捕获反射光成像, 二者成像原理不同, 因此可以从不同的角度提供互补信息。第二, 红外图像和可见光图像几乎可以呈现所有物体的固有特性。可见光图像分辨率高, 色彩和纹理信息丰富, 但容易受到恶劣天气的影响; 红外图像中热辐射信息丰

收稿日期: 2021-07-18; 修订日期: 2021-09-23.

作者简介: 曲海成 (1981-), 男, 博士, 副教授, 主要研究方向: 图像与智能信息处理。E-mail: quhaicheng@lntu.edu.cn.

基金项目: 辽宁省教育厅一般项目 (LJ2019JL010); 辽宁工程技术大学学科创新团队资助项目 (LNTU20TD-23)。

富,目标突出,不受天气影响,但分辨率低,细节不清晰,利用二者的优势进行图像融合可以得到一幅信息丰富的图像。此外,这些图像可以通过相对简单的设备获得。因此,红外与可见光图像融合技术可以比其它融合方案应用范围更广,例如可以应用到目标识别<sup>[1]</sup>、目标检测<sup>[2]</sup>、图像增强<sup>[3]</sup>、监控<sup>[4]</sup>和遥感<sup>[5]</sup>等领域。

图像融合包括像素级融合、特征级融合和决策级融合方法。像素级融合方法只对图像信息作特征提取并直接使用,代表性的方法包括变换域的各向异性扩散融合(anisotropic diffusion fusion, ADF)<sup>[6]</sup>、交叉双边滤波(cross bilateral filter, CBF)<sup>[7]</sup>和最小加权二乘法(weighted least square, WLS)<sup>[8]</sup>和稀疏域的自适应稀疏表示(adaptive sparse representation, ASR)<sup>[9]</sup>、拉普拉斯金字塔(laplacian pyramid, LP)<sup>[10]</sup>等方法。在变换域方法中,比较热门的是基于多尺度变换<sup>[11]</sup>的方法,但其忽略了融合过程中的空间一致性,容易在融合图像中引入光晕。稀疏域的方法<sup>[12]</sup>需要构建一个过完备字典,需要多次迭代,非常耗时。近年来,针对特征级图像融合,国内外学者提出了许多基于深度学习的融合方法。2017年,Prabhakar等人<sup>[13]</sup>提出了基于卷积神经网络的图像融合方法,该方法适用于多曝光图像融合,网络结构简单,仅使用编码网络中最后一层计算的结果,会造成中间层有用特征信息丢失的问题。2018年,Ma等人<sup>[14]</sup>提出了一种以生成对抗网络为模型的图像融合方法(FusionGAN),该方法通过生成器与判别器对抗性训练产生融合图像,无需预训练,是端到端的无监督学习模型,但其生成器与判别器网络结构相对简单,可以进行优化改进。2018年,Li等人<sup>[15]</sup>提出了一种基于深度网络的红外与可见光图像融合方法,该方法使用VGG网络为特征提取工具,在计算时参数较多,耗费大量计算资源。2019年,Li等人<sup>[16]</sup>提出了以密集连接网络为模型的图像融合方法(Dense Fuse),训练时只考虑编码层与解码层,不考虑融合层,测试时需要根据不同情况来设计融合规则。2020年,董安勇<sup>[17]</sup>等人改进了以卷积神经网络为模型的图像融合方法,有效地解决了低频子带融合模式给定的问题,同时又克服了手动设置参数的缺陷。2020年,Xu等人<sup>[18]</sup>提出了一种端到端的统一无监督学习(U2Fusion)图像融合方法,可以解决多模态、多聚焦和多曝光图像融合问题,但这种方法训练起来非常复杂,且占用大量内存资源。

上述方法主要针对清晰明亮的图像进行融合。可见光图像在照明条件良好的情况下纹理清晰,红外图像在光照较差和存在烟雾的条件下也能较好地凸

显出目标,二者在不同条件下各有所长,考虑到图像融合应该适应环境改变为源图像分配权重,从而更加合理地利用多模态信息,于是,本文提出一种模态自适应的红外与可见光图像融合方法(MAFusion)。主要工作如下:①采用双路特征提取网络针对性的提取成像特点不同的两类图像特征,以减少特征丢失;②建立自适应加权模块,通过双流交互学习红外与可见光图像差异,得到当前环境条件下两种模态在融合图像中所占的权重;③在图像融合过程中加入残差块与跳跃残差组合模块,补充融合图像的多尺度特征,提升网络性能;④在TNO和KAIST数据集上将MAFusion与其他方法进行对比实验,实验结果证明本文方法效果较好。

## 1 相关理论与分析

### 1.1 对抗生成网络模型

生成对抗网络是由Goodfellow等人<sup>[19]</sup>提出的,由生成器(generative, G)和判别器(discriminative, D)组成。生成器用来生成需要的图片,判别器类似图像分类器,判断图片是由生成器生成的融合图像还是真实的可见光图像。训练后得到一个比较智能的自动生成器和一个判断能力较强的分类器。GAN模型的优化函数公式(1)所示:

$$\min_G \max_D V(D, G) = E_{x \sim p_{\text{data}}(x)} [\lg D(x)] + E_{z \sim p_z(z)} [\lg \{1 - D[G(z)]\}] \quad (1)$$

式中: $E(*)$ 表示分布函数的期望值; $x$ 表示输入的真实值; $z$ 表示输入生成器的噪声大小; $p_{\text{data}}(x)$ 表示真实样本的分布; $p_z(z)$ 是定义在低维的噪声分布;希望 $D(x)$ 越大越好, $D[G(z)]$ 越小越好,函数的目的是将最大损失最小化,即可得到损失函数的最优解。

### 1.2 残差网络模型

为解决模型增加网络深度后网络性能变差的问题,何凯明等人<sup>[20]</sup>创新性地提出了深度残差网络ResNet。残差网络的主要思想是将输入层通过跳跃连接与卷积层后的输出相加,有效地解决了梯度消失问题。残差块的结构如图1所示,其中 $F(x)$ 为残差学习函数,该函数的学习相对容易,有助于提取深层特征信息,激活函数为Relu。

### 1.3 问题分析

在深度学习图像融合方法中,大多数方法采用相同的编解码网络进行特征提取,由于红外与可见光图像来自不同的信号,采用相同的编解码网络会造成突出特征提取不当的问题,因此,本方法将采用双流特征提取网络更具针对性地提取两类图像的特征。

传统的模态堆叠策略在训练过程中对两类源图像在通道维按 1:1 的权重直接 Concat 级联, 没有考虑两类图像在不同环境条件的优缺点, 容易引入冗余特征, 降低融合图像质量, 因此需要一种融合策略能够自适应环境改变提取红外与可见光两个模态的重要特征, 因此设计自适应加权模块为两个模态分配合理的权重。

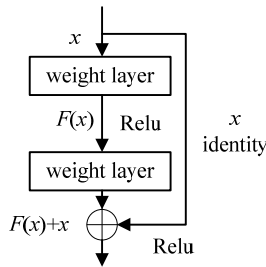


图1 残差块结构图

Fig.1 Residual block structure

为了提高模型性能, 使网络训练更加容易, 同时解决反向传播时的梯度消失问题, 本文方法引入了残差块, 将 Conv1 的输出直接连接到 Conv3 后。利用残差块的恒等映射和跨层连接使前向传播与反向传播更加容易, 同时有助于补充多尺度特征。在生成器网络中构建跳跃残差块, 这种大幅度跨层连接方式使前后层特征联系更加密切, 可以获取更多有利特征信息。

## 2 模态自适应的红外与可见光图像融合方法

本方法主体网络为 GAN, 包括生成器与判别器。生成器是整个模型的主要网络, 影响最终融合图像的质量, 主要用来生成可以骗过判别器的融合图像。判别器的作用不被生成器生成的图像骗过, 经过训练达到很强的区分可见光图像与生成器生成的融合图像的能力。生成器与判别器均参与训练过程, 测试时则直接利用训练好的生成器模型生成融合图像。

训练过程如下:

1) 将红外与可见光图像随机采样级联输入生成器, 得到初步融合图像 (标记为假数据), 记为  $D(F)$ ;

2) 将可见光图像作为真实数据, 随机采样记为  $x$ ;  
3) 将前两步中某一步产生的数据 (标记为假数据的融合图像和标记为真实数据的可见光图像) 输入判别器, 判别网络的输出值为该输入属于真实数据 (可见光图像) 的概率, 真为 1, 假为 0。

4) 然后根据得到的概率值计算损失函数;

5) 根据判别模型和生成模型的损失函数, 利用反向传播算法, 更新模型的参数。先更新判别模型的参数, 然后通过再采样得到的红外与可见光图像更新生成器的参数。

### 2.1 网络框架

生成器包含了双路输入与特征提取模块、自适应加权模块、残差块、跳跃残差块、特征重构模块和输出模块 6 个部分, 其网络结构图如图 2 所示。

生成器首先对红外图像和可见光图像进行特征提取, 然后通过 Concat 操作将二者特征图级联为 256 通道特征图, 再输入权值学习融合网络, 完成权值分配。

然后, 重新分配权值的融合特征输入一个普通卷积 Conv1 后, 进入残差块与跳跃残差模块, 以实现提取深层特征的同时保留浅层特征; 最后, 通过输出模块进行特征重构得到融合图像。其具体网络结构参数见表 1。

表中字母及缩写定义如下: VIS 代表可见光图像, IR 代表红外图像,  $ir\_weight$  为红外图像权重,  $vis\_weight$  为可见光图像权重,  $k$  表示卷积核大小,  $s$  表示移动步长,  $n_1$  表示输入通道数,  $n_2$  表示输出通道数。

判别器希望训练出来的模型具有很强的判别输入图像是融合图像或可见光图像的能力, 它以可见光图像为真实参考值, 希望不被生成器生成的融合图像骗过, 输出的判别值在 (0, 1) 之间, 该值越趋近于 1 表明判别能力越强, 即越趋向于真实参考值 (可见光图像)。其网络结构如图 3 所示。

判别器由 4 个卷积层、一个特征线性转换模块和一个线性层组成, 卷积层以  $3 \times 3$  大小的卷积核进行特征提取, 填充方式均为 VALID, 激活函数为 LeakyReLU, 卷积层 2、3 和 4 使用 BatchNorm 来提高模型收敛速度。判别器的网络结构及参数信息见表 2。

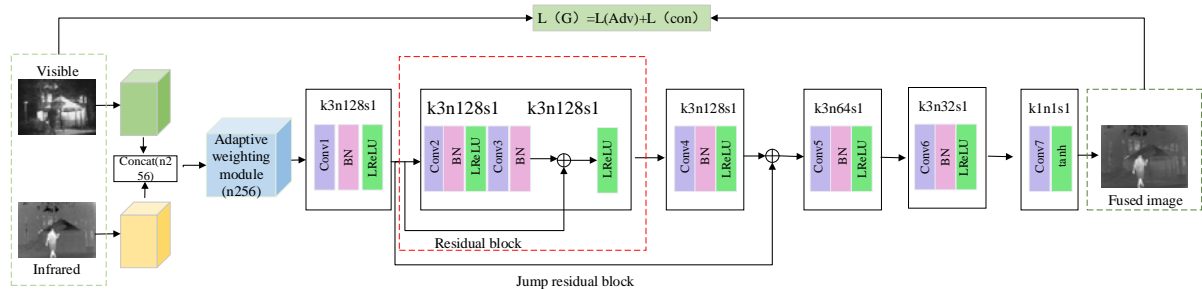


图2 生成器网络结构图

Fig.2 Structure diagram of generator network

表1 生成器网络整体结构

Table 1 Overall structure of generator network

	layer	$k$	$s$	$n_1$	Input	$n_2$	Output	fill	function
Feature extraction	10_ir	$5 \times 5$	1	1	VIS	128	Conv0_ir	VALID	LReLU
	10_vis	$5 \times 5$	1	1	IR	128	Conv0_vis	VALID	LReLU
Adaptive weighting module	L11_w_ir	$3 \times 3$	1	256	Concat(Conv0_ir, Conv0_vis)	128	ir_weight	SAME	LReLU
	L22_w_vis	$3 \times 3$	1	256	Concat(Conv0_ir, Conv0_vis)	128	vis_weight	SAME	LReLU
Routine convolution	Layer1	$5 \times 5$	1	256	Concat[multiply(Conv0_ir, ir_weight), multiply(Conv0_vis, vis_weight)]	128	Net1	VALID	LReLU
Residual block	Layer2	$3 \times 3$	1	128	Net1	128	Net2	SAME	LReLU
	Layer3	$3 \times 3$	1	128	Net2	128	Net3	SAME	-
	Add1	-	-	-	Net1/Net3	-	Net1+Net3	-	LReLU
Transfer layer	Layer4	$3 \times 3$	1	128	LReLU(Net1+Net3)	128	Net4	SAME	LReLU
Jump residual	Add2	-	-	-	Net1/net4	-	Net1+net4	-	-
Output layer	Layer5	$3 \times 3$	1	128	Net4	64	Net5	VALID	LReLU
	Layer6	$3 \times 3$	1	64	Net5	32	Net6	VALID	LReLU
	Layer7	$1 \times 1$	1	32	Net6	1	Fused	VALID	tanh

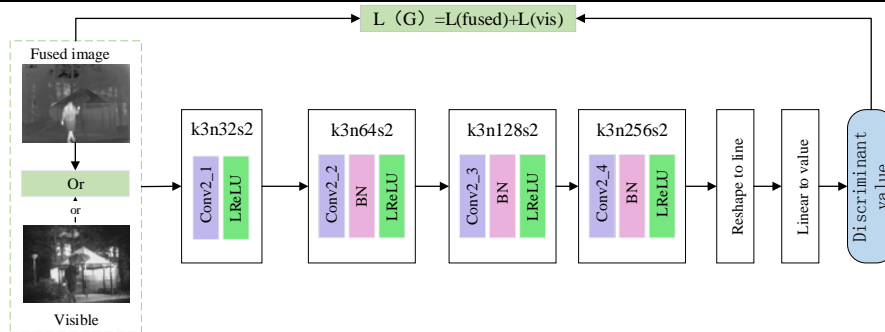


图3 判别器网络结构图

Fig.3 Structure diagram of generator network

表2 判别器网络整体结构

Table 2 Overall structure of discriminator network

Layer	$k$	$s$	$n_1$	Input	$n_2$	Output	Padding	Activation function
Conv_1	$5 \times 5$	2	1	VIS/Fused	32	Net1	VALID	LReLU
Conv_2	$3 \times 3$	2	32	Net1	64	Net2	VALID	LReLU
Conv_3	$3 \times 3$	2	64	Net2	128	Net3	VALID	LReLU
Conv_4	$3 \times 3$	2	128	Net3	256	Net4	VALID	LReLU
Line_5	-	-	-	Net4	-	Net5(6*6*256)	-	-
Output	-	-	-	Net5	1	Discriminant value (1)	-	-

表中字母缩写定义如下: VIS 代表可见光图像, Fused 代表融合图像,  $k$  表示卷积核大小,  $s$  表示移动步长,  $n_1$  表示输入通道数,  $n_2$  表示输出通道数。

## 2.2 自适应加权模块

为了在融合图像中充分利用红外图像目标区域

的热辐射信息, 保留可见光图像有利的纹理信息, 抑制噪声等干扰信息引入融合图像, 提出自适应加权模块来根据环境变化为两个模态分配合理的权重值, 将源图像在当前环境条件下的优势充分体现出来。在生成器中, 将该模块嵌入红外图像与可见光图像输入之

后, 最终得到一个包含两模态重要特征的融合特征图。其网络结构图如图4所示。

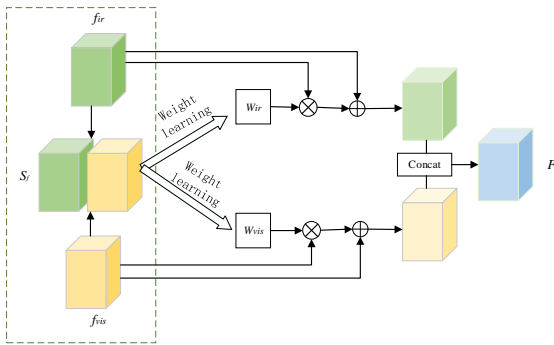


图4 自适应加权模块结构图

Fig.4 Adaptive weighted fusion module structure diagram

该模块以前一阶段得到的红外与可见光各自特征图 Concat 作为输入, 首先通过一个  $3 \times 3$  卷积分离两类特征, 然后通过 sigmoid 函数计算当前环境下的红外图像与可见光图像所占的比例, 红外权重矩阵  $W_{ir}$  和可见光权重矩阵  $W_{vis}$  的计算公式如式(2)所示:

$$\begin{aligned} W_{ir} &= \sigma(C_1 * S_f + b_1) \\ W_{vis} &= \sigma(C_2 * S_f + b_2) \\ \sigma(x) &\triangleq \frac{1}{1 + e^{-x}} \end{aligned} \quad (2)$$

式中:  $C_1$  和  $C_2$  表示卷积权重;  $b_1$  和  $b_2$  为卷积偏置;  $W_{ir}$  和  $W_{vis}$  通过自主学习得来。最后将获得的权重与原始特征图像在通道维相乘, 再与源图像相加以避免特征丢失, 从而得到两个模态的优势特征, 干扰信息被很好地抑制。

### 2.3 残差块与跳跃残差块

为解决由于网络过深导致反向传播时梯度消失的问题, 本文方法引入了残差块, 学习到的残差信息有助于补充多尺度特征。同时, 为了增加特征图前后层的关联性, 降低网络训练复杂性, 在残差块的输入与其后卷积层的输出通过跳跃连接相连, 构建了一个跳跃残差模块。残差块与跳跃残差块的结构图如图5所示。

### 2.4 损失函数与优化方法

MAFusion 的损失函数包括两部分: 生成器损失和判别器损失。生成器损失计算融合图像与源图像之间的差别; 判别器损失以可见光图像作为标签, 训练辨别可见光图像与融合图像的能力。

#### 1) 生成器损失函数

生成器损失  $L_G$  包含对抗损失  $L_{Adv}$  及内容损失  $L_{con}$ , 其定义如式(3)所示:

$$L_G = L_{Adv} + L_{con} \quad (3)$$

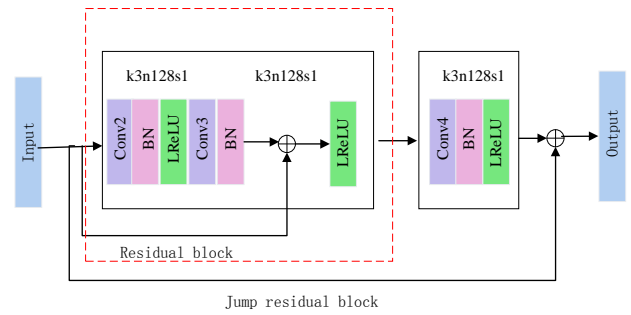


图5 残差块及跳跃连接组合模块

Fig.5 Combined module of residual block and jump connection

内容损失  $L_{con}$  是生成器的主要损失函数, 约束了融合图像与源图像间的相关性及重点提取红外图像与可见光图像的有利信息。内容损失越小, 融合图像中源图像的特征越多, 其定义如式(4)所示:

$$L_{con} = \frac{1}{HW} \left( \|F - I\|_F^2 + \xi \|\nabla F - \nabla I\|_F^2 \right) \quad (4)$$

对抗损失  $L_{Adv}$  是生成器与判别器交互的损失函数, 其定义如式(5)所示:

$$L_{Adv} = \frac{1}{N} \sum_{n=1}^N [D(F) - a]^2 \quad (5)$$

式中:  $D(F)$  为判别器对融合图像的判别结果, 在(0,1)之间;  $a$  为软标签(取值不确定的标签)。判别器希望融合图像向可见光图像靠近, 其值越大越好, 又要保证损失最小, 根据文献[14]将软标签  $a$  的取值范围设置为(0.7, 1.2)。

#### 2) 判别器损失函数

判别器损失  $L_D$  的目的是约束判别器网络反向传播参数更新, 使判别器在训练后具有分辨可见光图像和融合图像的能力, 其定义如式(6)所示:

$$L_D = \frac{1}{N} \sum_{n=1}^N [D(V) - b]^2 + \frac{1}{N} \sum_{n=1}^N [D(F) - c]^2 \quad (6)$$

式中:  $D(V)$  表示判别器对可见光图像的判别结果;  $D(F)$  表示判别器对融合图像的判别结果,  $b, c$  为软标签(取值不确定的标签), 根据文献[14]将  $b$  的范围设置为(0.7, 1.2), 将  $c$  的范围设置为(0, 0.3)。

## 3 实验及结果分析

实验在 Windows10 操作系统和 NVIDIA GTX 1080Ti GPU 上完成, 框架为 TensorFlow 1.12.0, 训练与测试平台为 PyCharm。本文方法的训练集是在 TNO<sup>[21]</sup>数据集中选择了 39 对不同场景的红外和可见光图像, 以步长 15 将 39 对图像分成 30236 对  $120 \times 120$  的红外图像与可见光图像, 最后的融合图像为生成器



输出图像的拼接。判别器训练次数设置为2,学习率设为0.00001,将batch\_size设置为32,训练迭代次数设置为20次。本节采用TNO数据集的18对白天图像和KAIST<sup>[22]</sup>数据集的14对夜间图像评估MAFusion的性能。选择ADF<sup>[6]</sup>、CBF<sup>[7]</sup>、WLS<sup>[8]</sup>、FusionGAN<sup>[14]</sup>、Densefuse<sup>[16]</sup>和U2Fusion<sup>[18]</sup>6种方法与MAFusion作对比实验,对融合图像进行主观评价和客观评价。对比实验图片由MATLAB及PyCharm程序生成,客观评价指标计算通过MATLAB软件编程实现。

### 3.1 主观评价

MAFusion与对比方法在KAIST数据集中夜晚路灯下的“道路与行人”实验结果如图6所示。

从图6的实验结果可以看出:图6(c)中,道路右边的车轮廓模糊,整幅图像清晰度较低;图6(d)中,整体噪声大,路灯中心有明显的黑块;图6(e)中,道路右边的人边缘不清晰,整体纹理细节信息少;图6(f)中,人和车的轮廓模糊;图6(g)中,整体背景偏暗,视觉效果较好;图6(h)中,图像整体较为平滑,路灯周围有光晕;图6(i)中,整体背景较亮且清晰,目标突出,抑制路灯周围的光晕,能清楚地识别楼层、路灯、行人和车辆等目标。

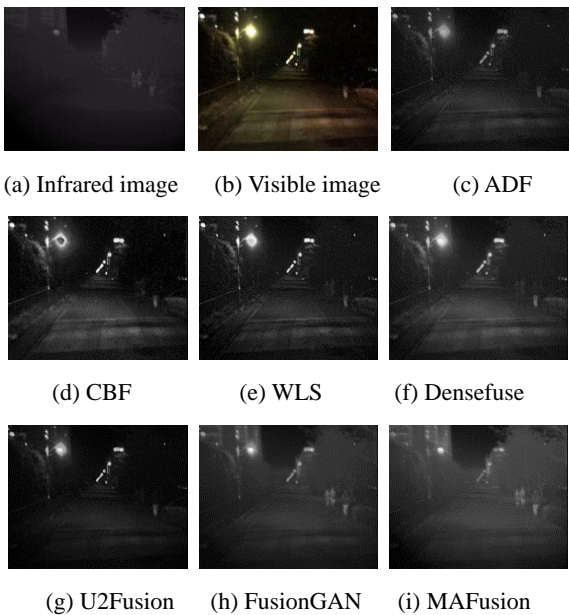


图6 KAIST数据集中“道路与行人”融合结果

Fig.6 Fusion results of "road and people" in KAIST dataset

MAFusion与对比方法在TNO数据集中的“烟雾中的士兵”实验结果如图7所示。

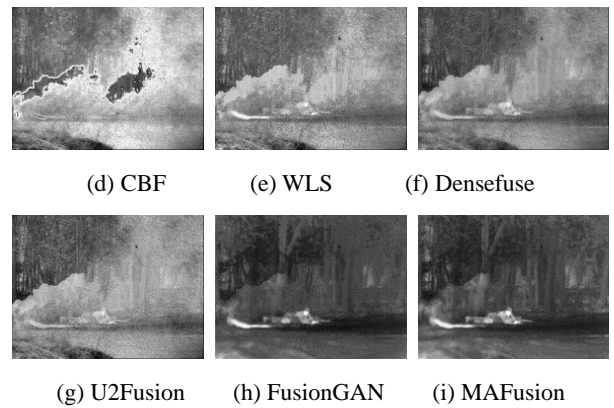
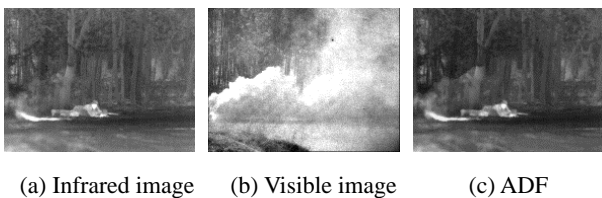


图7 TNO数据集中“烟雾中的士兵”融合结果

Fig.7 Fusion results of "a soldier in the smog" in KAIST dataset

从图7的实验结果可以看出:图7(c)、(h)中,能看到士兵与树木,但轮廓较模糊;图7(d)中,图像整体上被烟雾笼罩,无法识别士兵,图像整体噪声较大,且有不连续的黑块;图7(e)中,士兵轮廓不清晰,烟雾的干扰信息较多,图像不平滑;图7(f)中,人与汽车的轮廓都比较模糊,图像较平滑,仍有烟雾的干扰信息;图7(g)中,图像较清晰,但没有抑制烟雾的干扰,士兵轮廓不清晰;图7(i)中,士兵与树木轮廓清晰可见,抑制了烟雾的干扰,重点目标突出,图像整体平滑,有利信息得到了良好地体现。

MAFusion与对比方法在TNO数据集中的“士兵与车辆”实验结果如图8所示。

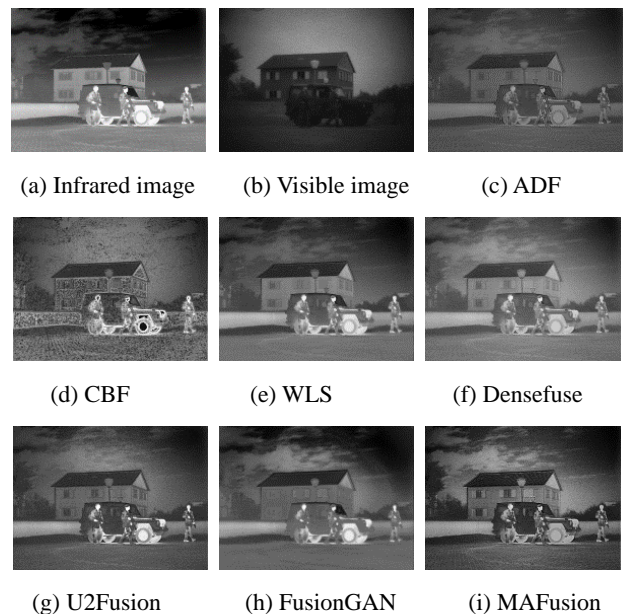


图8 TNO数据集中“士兵与车辆”融合结果

Fig.8 Fusion results of "soldiers and vehicles" in TNO dataset

从图8的实验结果可以看出:图8(c)中,背景较暗,图像整体不够平滑;图8(d)中,整体呈现颗粒状,有不连续的黑块,视觉效果差,整体噪声较大;图8(e),图像整体较模糊,窗户与左侧暗处的树木轮廓不清

晰；图 8(f)中，人的轮廓信息不清楚，部分可见光图像细节信息丢失；图 8(g)中，整体偏暗，人、车、树木、房屋均较清晰，但天空中的云轮廓较模糊；图 8(h)中，图像较平滑，但房屋的窗户、车窗与左侧树木的轮廓不清晰；图 8(i)中，整体较清晰，图像平滑，可以清晰看到房屋窗户的轮廓，车辆、房屋、士兵、地面、路灯和树木都清晰可见，细节信息丰富，热辐射信息也得到了较好的体现。

### 3.2 客观评价

为了进一步评价融合图像的融合效果，使用客观评价方法进行评估。选择信息熵 EN、互信息 MI 和基于噪声评估的融合性能  $N_{abf}$  为测评指标，从图像信息丰富度、信息互补性和噪声等不同角度对融合图像进行评价。以下为测评指标的计算公式。

1) 信息熵<sup>[23]</sup>EN。其值大小能够体现融合图像中平均信息量的多少，其定义如式(7)所示：

$$EN = \sum_{i=0}^L p_i \log_2 p_i \quad (7)$$

信息熵越大越好，表明图像所含信息越丰富。

2) 互信息<sup>[24]</sup>MI。该指标表示融合图像包含源图像的信息量大小，其值越大越好，计算方法如式(8)所示：

$$\begin{aligned} MI_{A,F} &= \sum_{i=1}^{L-1} \sum_{j=1}^{L-1} h_{A,F}(i,j) \cdot \log_2 \frac{h_{A,F}(i,j)}{h_A(i) + h_F(j)} \\ MI_{B,F} &= \sum_{i=1}^{L-1} \sum_{j=1}^{L-1} h_{B,F}(i,j) \cdot \log_2 \frac{h_{B,F}(i,j)}{h_B(i) + h_F(j)} \\ MI_{AB,F} &= \frac{MI_{A,F} + MI_{B,F}}{2} \end{aligned} \quad (8)$$

式中： $A$ 、 $B$  为源图像； $F$  为融合图像。

3) 基于噪声评价的融合性能<sup>[25]</sup> $N_{abf}$ 。该指标反映了图像中包含的不必要或冗余的干扰信息的多少，其值越小，图像噪声越少。

首先，在照明条件差且行人较小的 KAIST 数据集上选取 14 对红外与可见光图像对进行实验，图片“道路与行人”的融合效果如图 6 所示，CBF 方法得到的融合图像视觉效果差，所以直接由其他 5 种方法与 MAFusion 分别生成的融合图像进行测评，“道路与行人”的测评结果见表 4。EN 和 MI 指标的值越大越好， $N_{abf}$  的值越小越好。

由表 4 可见，MAFusion 在“道路与行人”的融合图像在 EN 和 MI 两个方面都是最优的， $N_{abf}$  值也接近于最优值。KAIST 数据集中图像偏暗，MAFusion 中模态自适应权值学习功能很好地感知了光照条件，

为两个模态分配合适的权值，在目标区域保留红外图像的热辐射信息，背景区域保留可见光图像的纹理信息，使融合图像测评指标值进一步优化。

表 4 KAIST 数据集“道路与行人”融合图像客观评价  
Table 4 Objective evaluation of "road and pedestrian" fusion image in KAIST dataset

Methods	EN	MI	$N_{abf}$
ADF	5.7176	11.4352	0.0782
WLS	5.8693	11.7387	0.2129
densefuse	6.0565	12.1129	<b>0.0121</b>
U2Fusion	5.3543	10.7086	0.0809
FusionGAN	5.8022	11.6044	0.0389
MAFusion	<b>6.3678</b>	<b>12.5356</b>	0.0447

Note: Bold font is the best value

以 KAIST 数据集中 14 对红外与可见光图像为源图像，MAFusion 与 6 种对比方法的测评结果见表 5。表中数值均为每种方法生成的 14 张融合图像在 EN、MI 和  $N_{abf}$  指标上测评结果的平均值，EN 和 MI 指标的值越大越好， $N_{abf}$  的值越小越好。

表 5 KAIST 数据集 14 组融合图像客观评价指标均值  
Table 5 Mean value of objective evaluation of 14 groups fusion images in KAIST dataset

Methods	EN	MI	$N_{abf}$
ADF	6.2336	12.4673	0.0930
WLS	6.4248	12.8496	0.2465
CBF	6.5413	13.0827	0.1551
Densefuse	6.6039	13.2078	<b>0.0223</b>
U2Fusion	5.9722	11.9444	0.1318
FusionGAN	6.2516	12.5033	0.0921
MAFusion	<b>6.7086</b>	<b>13.4042</b>	0.0562

Note: Bold font is the best value

由表 5 可见，MAFusion 在 KAIST 数据集上的融合图像在 EN 和 MI 两个方面都是最优的， $N_{abf}$  也与对比方法最优值相近，算法性能良好。

然后，在 TNO 数据集上选取 18 对红外与可见光图像进行实验，由 MAFusion 与 6 种对比方法分别生成融合图像进行测评，图片“烟雾中的士兵”的融合结果如图 7 所示，其中 CBF 与 WLS 方法得到的融合图像视觉效果差，DenseFuse 与 U2Fusion 有大量烟雾干扰信息，无法识别图中目标，所以直接由 ADF、FusionGAN 与 MAFusion 这 3 种方法分别生成的融合图像进行测评，“烟雾中的士兵”的测评结果见表 6。EN 和 MI 指标的值越大越好， $N_{abf}$  的值越小越好。

由表 6 可见，MAFusion 在“烟雾中的士兵”的

融合图像在 EN 和 MI 两个指标中取得最大值, 相比其他方法有明显地提高, 说明得到的融合图像信息更丰富, 源图像转移到融合图像的信息量较大。N<sub>abf</sub> 值最低, 说明融合图像与对比方法相比噪声最小, 抑制干扰信息引入融合图像。

表 6 TNO 数据集“烟雾中的士兵”融合图像客观评价

Table 6 Objective evaluation of "a soldier in the smog" fusion image in TNO dataset

Methods	EN	MI	N <sub>abf</sub>
ADF	6.4886	13.1772	0.104
FusionGAN	6.3959	12.7919	0.026
MAFusion	<b>6.6097</b>	<b>13.2193</b>	<b>0.018</b>

Note: Bold font is the best value

TNO 数据集中图片“士兵与车辆”的融合结果如图 8 所示, 其中由 CBF 方法生成的融合图像含有大量不连续的黑块, 图像视觉效果差, 所以直接由其他 5 种方法与 MAFusion 分别生成的融合图像进行测评, “士兵与车辆”的测评结果见表 7。EN 和 MI 指标的值越大越好, N<sub>abf</sub> 的值越小越好。

表 7 TNO 数据集“士兵与车辆”融合图像客观评价

Table 7 Objective evaluation of "Soldiers and vehicles" fusion image in TNO dataset

Methods	EN	MI	N <sub>abf</sub>
ADF	6.4620	12.9241	0.1582
WLS	6.9916	13.9833	0.2652
Densefuse	7.0227	14.0453	0.0931
U2Fusion	7.1399	14.2797	0.4285
FusionGAN	7.0007	14.0015	0.1093
MAFusion	<b>7.2017</b>	<b>14.4033</b>	<b>0.0801</b>

Note: Bold font is the best value

由表 7 可见, MAFusion 在“士兵与车辆”的融合图像在 EN 和 MI 两个指标中取得最大值, N<sub>abf</sub> 取得最小值, 相比其他方法的性能有明显提高, 说明 MAFusion 得到的融合图像信息更丰富, 噪声较小, 融合质量较好。

以 TNO 数据集中的 18 对红外与可见光图像为源

表 9 KAIST 数据集消融实验客观评价

Table 9 Objective evaluation of ablation experiment in KAIST dataset

Method	FusionGAN	Direct stacking	Two way feature extraction and adaptive weighting	Ordinary convolution	Residual block	Jump residual block	EN	MI	N <sub>abf</sub>
①	√	√	-	√	-	-	6.3787	12.7574	0.0921
②	√	√	-	-	√	√	6.5284	13.0567	0.0607
③	√	-	√	-	√	√	<b>6.7086</b>	<b>13.4042</b>	<b>0.0562</b>

Note: Bold font is the best value

图像, 由 MAFusion 与 6 种对比方法分别生成融合图像进行测评, 测评结果见表 8。表中数值均为每种方法生成的 18 张融合图像在 EN、MI 和 N<sub>abf</sub> 指标上测评结果的平均值, EN 和 MI 指标的值越大越好, N<sub>abf</sub> 的值越小越好。

表 8 TNO 数据集 18 组融合图像客观评价指标均值

Table 8 Mean value of objective evaluation of 18 groups fusion images in TNO dataset

Methods	EN	MI	N <sub>abf</sub>
ADF	6.6001	13.2003	0.0896
WLS	6.9157	13.8314	0.2404
CBF	6.9846	13.9697	0.3921
Densefuse	6.9482	13.8964	0.0854
U2Fusion	7.0457	14.0915	0.3313
FusionGAN	6.9058	13.8116	0.0892
MAFusion	<b>7.0653</b>	<b>14.1306</b>	<b>0.0608</b>

Note: Bold font is the best value

由表 8 可以看出, MAFusion 在 EN 和 MI 两个指标中取得最大值, 较对比方法有提高, 说明得到的融合图像有利信息更多, 源图像转移到融合图像的信息量较大, 噪声较小。

为验证 MAFusion 中双路特征提取与自适应加权模块、残差块及跳跃残差块的作用, 使用 KAIST 数据集中的 14 对典型红外与可见光图像做消融实验, 采用客观评价对所有方法生成的融合图像进行测评, 测评结果见表 9。表中数值均为每种方法生成的 14 张融合图像在 EN、MI 和 N<sub>abf</sub> 指标上测评结果的平均值, EN 和 MI 指标的值越大越好, N<sub>abf</sub> 的值越小越好。

在表 9 中, 方法①为 FusionGAN 方法, 其采用直接堆叠策略将红外与可见光图像 Concat 后输入特征提取网络, 特征提取阶段也使用普通卷积完成; 方法②在 FusionGAN 的基础上, 在特征提取阶段采用残差块与跳跃残差块, 得到的图像经测评后 3 个指标值均优于方法①, 体现了残差块和跳跃残差块恒等映射的优势, 加强浅层特征与深层特征的联系, 获得的融合图像能较好地兼顾不同种类源图像信息。方法③



在方法②的基础上,将红外与可见光图像送入双路特征提取网络并进行权值自适应学习,得到的图像测评后的指标值均优于方法②。方法③即MAFusion生成的融合图像如图6、图7和图8所示,图像中的信息更丰富,视觉效果良好。

## 4 结论

针对恶劣环境条件下红外与可见光图像融合效果较差且容易引入噪声的问题,本文提出了一种模态自适应的红外与可见光图像融合方法。该方法使用双路特征提取及权值学习融合网络,自适应提取两个模态的重要信息,并分配权重,强化对重要信息的学习。同时,在生成器特征提取阶段加入残差块与跳跃残差块,使得深层网络特征提取能力增强,反向传播也更加容易,增强模型学习能力。实验结果表明,提出的方法在图像的视觉观察、目标清晰度、图像信息丰富性和抑制噪声等方面具有良好的性能。

### 参考文献:

- [1] 段辉军, 王志刚, 王彦. 基于改进YOLO网络的双通道显著性目标识别算法[J]. *激光与红外*, 2020, **50**(11): 1370-1378.  
DUAN H J, WANG Z, WANG Y. Two-channel saliency object recognition algorithm based on improved YOLO network[J]. *Laser & Infrared*, 2020, **50**(11): 1370-1378.
- [2] 李舒涵, 宏科, 武治宇. 基于红外与可见光图像融合的交通标志检测[J]. *现代电子技术*, 2020, **43**(3): 45-49.  
LI S H, XU H K, WU Z Y. Traffic sign detection based on infrared and visible image fusion[J]. *Modern Electronics Technique*, 2020, **43**(3): 45-49.
- [3] Reinhard E, Adhikhmin M, Gooch B, et al. Color transfer between images[J]. *IEEE Comput. Graph. Appl.*, 2001, **21**(5): 34-41.
- [4] Kumar P, Mittal A, Kumar P. Fusion of thermal infrared and visible spectrum video for robust surveillance[C]// *Proceedings of the Indian Conference on Computer Vision, Graphics and Image Processing*, 2006: 528-539.
- [5] 常新亚, 丁一帆, 郭梦瑶. 应用整体结构信息分层匹配的红外与可见光遥感图像融合方法[J]. *航天器工程*, 2020, **29**(1): 100-104.  
CHANG X Y, DING Y F, GUO M Y. Infrared and visible image fusion method using hierarchical matching of overall structural information[J]. *Spacecraft Engineering*, 2020, **29**(1): 100-104.
- [6] Bavirisetti D P, Dhuli R. Fusion of infrared and visible sensor images based on anisotropic diffusion and karhunen-loeve transform[J]. *IEEE Sensors*, 2016, **16**(1): 203-209.
- [7] Kumar B S. Image fusion based on pixel significance using cross bilateral filter[J]. *Signal Image Video Process*, 2015, **9**(5): 1193-1204.
- [8] MA J, ZHOU Z, WANG B. Infrared and visible image fusion based on visual saliency map and weighted least square optimization[J]. *Infrared Physics & Technology*, 2017, **82**: 8-17.
- [9] LIU Y, WANG Z. Simultaneous image fusion and denoising with adaptive sparse representation[J]. *IET Image Process*, 2014, **9**(5): 347-357.
- [10] Burt P, Adelson E. The Laplacian pyramid as a compact image code[J]. *IEEE Trans. Commun.*, 1983, **31**(4): 532-540.
- [11] 马雪亮, 柳慧超. 基于多尺度分析的红外与可见光图像融合研究[J]. *电子测试*, 2020, **24**(4): 57-58.  
MA Xueliang, LIU Huichao. Research on infrared and visible image fusion based on multiscale analysis[J]. *Electronic Test*, 2020, **24**(4): 57-58.
- [12] LI S, YIN H, FANG L. Group-sparse representation with dictionary learning for medical image denoising and fusion[J]. *IEEE Trans Biomed Eng.*, 2012, **59**(12): 3450-3459.
- [13] Prabhakar K R, Srikar V S, Babu R V. DeepFuse: A deep unsupervised approach for exposure fusion with extreme exposure image[C]//*Proc of the 2017 IEEE International Conference on Computer Vision*, 2017: 4724-4732.
- [14] 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.
- [15] LI H, WU X J, Kittler J. Infrared and visible image fusion using a deep learning framework[C]//*The 24th International Conference on Pattern Recognition (ICPR)*, 2018: 2705-2710.
- [16] LI H, WU X. DenseFuse: A fusion approach to infrared and visible images[J]. *IEEE Transactions on Image Processing*, 2019, **28**(5): 2614-2623.
- [17] 董安勇, 杜庆治, 苏斌, 等. 基于卷积神经网络的红外与可见光图像融合[J]. *红外技术*, 2020, **42**(7): 660-669.  
DONG Anyong, DU Qingzhi, SU Bin, et al. Infrared and visible image fusion based on convolutional neural network[J]. *Infrared Technology*, 2020, **42**(7): 660-669.
- [18] XU Han, MA Jiayi, JIANG Junjun, et al. U2Fusion: A unified unsupervised image fusion network[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, **44**: 502-518.
- [19] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//*Proceedings of the 27th International Conference on Neural Information Processing Systems*, 2014: 2672-2680.
- [20] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016:770-778.
- [21] Figshare. TNO Image Fusion Dataset[OL]. [2018-09-15]. [https://figshare.com/articles/TNO Image Fusion Dataset/1008029](https://figshare.com/articles/TNO_Image_Fusion_Dataset/1008029).
- [22] Hwang S, Park J, Kim N, et al. Multispectral pedestrian detection: benchmark dataset and baseline[C]//*2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015: 1037-1045.
- [23] Roberts J W, Aardt J V, Ahmed F. Assessment of image fusion procedures using entropy, image quality, and multispectral classification[J]. *Appl. Remote Sens.*, 2008, **2**(1): 023522-023522-28.
- [24] QU G, ZHANG D, YAN P. Information measure for performance of image fusion[J]. *Electron Lett.*, 2002, **38**(7): 313-315.
- [25] Kumar B K S. Multifocus and multispectral image fusion based on pixel significance using discrete cosine harmonic wavelet transform[J]. *Signal, Image and Video Processing*, 2013, **7**(6): 1125-1143.