



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



## 基于结构重参数化的红外与可见光图像融合

陈昭宇, 范洪博, 马美燕, 赵艺博

引用本文:

陈昭宇, 范洪博, 马美燕, 赵艺博. 基于结构重参数化的红外与可见光图像融合[J]. 控制与决策, 2024, 39(7): 2275–2283.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.2003>

## 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>

#### 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>

#### 基于双分支特征融合的场景文本检测方法

A scene text detection based on dual-path feature fusion

控制与决策. 2021, 36(9): 2179–2186 <https://doi.org/10.13195/j.kzyjc.2020.0002>

#### 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>

# 基于结构重参数化的红外与可见光图像融合

陈昭宇<sup>1</sup>, 范洪博<sup>2†</sup>, 马美燕<sup>1</sup>, 赵艺博<sup>1</sup>

(1. 昆明理工大学 信息工程与自动化学院, 昆明 650500; 2. 昆明理工大学 现代农业工程学院, 昆明 650500)

**摘要:** 红外与可见光图像融合的目的是通过将不同模态的互补信息融合来增强源图像中的细节场景信息, 然而, 现有的深度学习方法存在融合性能与计算资源消耗不平衡的问题, 并且会忽略红外图像存在噪声的问题. 对此, 提出一种基于结构重参数化的红外与可见光图像融合算法. 首先, 通过带权重共享的双分支残差连接型网络分别对两种源图像进行特征提取, 分别得到的特征级联后图像重建; 然后, 用结构相似性损失与双边滤波去噪的内容损失联合指导网络的训练; 最后, 在训练完成后进行结构重参数化将训练网络优化成直连型网络. 在多个公共数据集上与 7 种领先的深度学习融合算法进行了定性与定量的实验对比, 所提出的融合算法在更低的资源耗费下能够实现多个评价指标的提升, 融合结果具有更丰富的场景信息、更强的对比度以及更符合人眼的视觉效果.

**关键词:** 图像融合; 结构重参数化; 红外与可见光图像; 双边滤波; 深度学习

中图分类号: TP391

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.2003

引用格式: 陈昭宇, 范洪博, 马美燕, 等. 基于结构重参数化的红外与可见光图像融合[J]. 控制与决策, 2024, 39(7): 2275-2283.

## Infrared and visible image fusion based on structural re-parameterization

CHEN Zhao-yu<sup>1</sup>, FAN Hong-bo<sup>2†</sup>, MA Mei-yan<sup>1</sup>, ZHAO Yi-bo<sup>1</sup>

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. Faculty of Modern Agricultural Engineering, Kunming University of Science and Technology, Kunming 650500, China)

**Abstract:** The purpose of infrared and visible image fusion is to enhance the detailed scene information in the source image by fusing the complementary information of different modalities. However, the existing deep learning methods have the problem of unbalanced fusion performance and computing resource consumption, and ignore the problem of noise in infrared images. Aiming at these two problems, this paper proposes an infrared and visible image fusion algorithm based on structural reparameterization. Firstly, the algorithm performs feature extraction on the two source images through a two-branch residual connection network with weight sharing, and the obtained features are cascaded to reconstruct the images. Then, the structural similarity loss and the content loss with bilateral filtering denoising are used to jointly guide the training of the network. Finally, after the training is completed, the structure reparameterization is performed to optimize the training network into a direct connection network. Qualitative and quantitative experiments are compared with seven leading deep learning fusion algorithms on multiple public data sets. The proposed fusion algorithm achieves the improvement of multiple evaluation indicators with lower resource consumption. The fusion results have richer scene information, stronger contrast and more in line with the visual effect of the human eye.

**Keywords:** Image fusion; structural reparameterization; infrared and visible images; bilateral filtering; deep learning

## 0 引言

图像融合作为一种图像增强技术,其作用是融合不同的单一类型传感器或单一拍摄设置下的图像,以获得具有更全面的场景表征信息图像. 其中,红外与可见光图像融合是典型代表之一<sup>[1]</sup>. 红外图像是通过红外传感器热辐射成像,因此它可以突出显著目标,受光照变化或伪装影响小,能够很好地区分显著目标

与背景,但它具有噪声大、细节模糊以及视觉效果差的特点. 而可见光图像则包含丰富的纹理细节及结构信息,图像分辨率较高,但其易受外界环境影响,例如天气、光线等都会影响可见光图像的质量,导致信息丢失严重<sup>[2]</sup>. 因此,红外图像与可见光图像的互补特性,使得图像融合应运而生. 将红外图像与可见光图像融合可以整合各自的独特信息,去除冗余特征,

收稿日期: 2022-11-18; 录用日期: 2023-04-04.

责任编辑: 张文安.

<sup>†</sup>通讯作者. E-mail: qingchen233@foxmail.com.

生成具有丰富纹理和突出目标的融合图像,以便于语义分割、目标检测以及目标跟踪等后续任务的进行<sup>[3]</sup>.

在深度学习出现之前,传统图像融合方法主要有基于多尺度分解的方法<sup>[4-5]</sup>、基于稀疏表示的方法<sup>[6-7]</sup>、基于子空间的方法<sup>[8]</sup>、基于优化的方法<sup>[9]</sup>和混合方法<sup>[10]</sup>. 这些方法通常使用各种数学变换,通过设计空间域或变换域中的活动测量水平、融合规则来实现融合. 因此,虽然传统算法具有较强的可解释性,但由于图像任务变得越来越复杂,人工设计的融合规则会限制融合的性能. 而基于数据驱动的深度学习方法解决了传统方法的这些弊端.

基于深度学习的融合方法主要分为3大类: 基于自编码器(auto-encoder, AE)的方法、基于卷积神经网络(convolutional neural network, CNN)的方法和基于生成对抗网络(generative adversarial network, GAN)的方法. 文献[11]在基于AE的框架基础上加入了密集连接,以提高网络提取特征的能力,使网络易于训练. 文献[12]则在AE中引入注意力机制,使网络能够关注图像中的显著目标和纹理细节. 但上述基于AE的方法都需要手工设计融合规则,导致其性能被限制而无法达到进一步的高度. 基于CNN的方法是一种端到端的、无需手工设计融合规则的方法. 文献[13]通过在CNN中加入跨模差分感知融合模块(CMDAF)来提升网络特征提取的能力. 文献[14]提出了基于双注意力特征融合的图像融合方法. 文献[15]的FusionGAN将GAN引入图像融合任务中,并将融合任务定义为生成器与鉴别器之间的对抗博弈,但它削弱了纹理细节. 文献[16]又设计了DDcGAN,其根据红外图像与可见光图像的模态差异设计了双鉴别器,但该方法导致部分结果中存在伪影.

现有融合方法常常会堆叠(如密集连接等)大量的网络模块,耗费较多的存储和计算量来换取性能指标的提升,而使用简单的卷积神经网络通常也难以训练出好的效果. 其次,由于红外图像的热辐射成像原理,导致红外图像容易存在大量噪声<sup>[17]</sup>,若忽略红外图像存在噪声的问题,则会导致融合图像引入源图像的噪声,可视性较差.

针对这些问题,本文提出一种基于结构重参数化的红外与可见光图像融合的方法. 首先,主网络采用结构重参数化特征提取模块进行特征提取和图像重建;然后,在红外分支和可见光分支的结构重参数化特征提取块之间用权重共享的方式,将不同模态特征的信息域拉近;最后,用结构相似性损失与双边滤波去噪的内容损失共同指导网络的训练,使得快速生成

的融合图像既能保留红外图像的显著性信息,又能保留可见光图像的纹理细节.

## 1 相关工作

### 1.1 结构重参数化

大多数的现有方法通过加入复杂网络模块实现红外与可见光图像融合,这在一定程度上能提高融合结果的指标,但往往会增大网络的计算量. 既能有效提取图像特征、又能高计算效率的网络鲜为少见. 文献[18]提出了一种基于结构重参数化的VGG式极简单架构,既可以充分提取图像特征,又能降低网络计算量,在目标检测等领域拥有极佳的表现. 受该方法的启发,本文引入这种结构重参数化的思想,用于解决红外与可见光图像融合任务中融合性能与计算资源消耗不平衡的问题.

结构重参数化是指构造一系列结构在训练阶段使用,待训练完成后,将其参数等效转换为另一组用于推理的参数,从而将这一系列结构等效转换为另一系列结构的过程. 结构重参数化模块在训练阶段使用多分支的卷积进行特征提取与图像重建,而在测试阶段之前,将多分支卷积等效转换为单路卷积,从而拥有更快、更少资源耗费的推理能力. 结构重参数化具体操作如图1所示,与 $3 \times 3$ 卷积层并行的 $1 \times 1$ 卷积分支和恒等映射分支,分别等价变换成 $3 \times 3$ 卷积,再根据卷积的线性可加性<sup>[19]</sup>对结构重参数化模块做等效转换,从而得到图中最后阶段的 $3 \times 3$ 卷积. 该卷积块等效融合了多分支的参数,并以更快的推理速度达到原先的训练效果. 卷积的线性可加性公式为

$$\text{Conv}(x, \omega_1) + \text{Conv}(x, \omega_2) + \text{Conv}(x, \omega_3) = \text{Conv}(x, (\omega_1 + \omega_2 + \omega_3)). \quad (1)$$

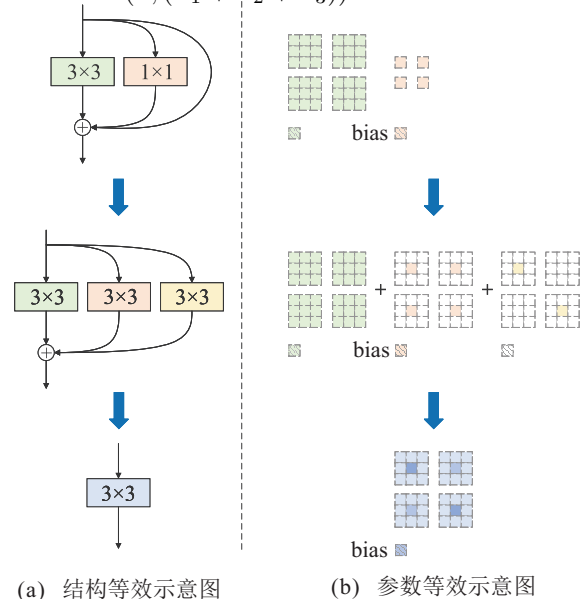


图1 多分支卷积等效转换过程

其中: Conv为卷积运算,  $x$  为输入特征,  $\omega_1$ 、 $\omega_2$ 、 $\omega_3$  为大小相同的卷积核.

这种结构重参数化的操作使得网络模型在训练时具备残差连接型网络强大特征提取能力,又在测试推理时拥有单路直连型网络模型计算速度快、耗费内存资源少的特点.

### 1.2 双边滤波

由于红外成像设备容易受到热噪声和校准噪声的影响,采集的红外图像中会含有大量的噪声. 为了实现更好的融合效果,本文对红外图像进行了双边滤波处理,使得融合图像减少噪声以达到更优越的视觉效果. 双边滤波是同时考虑像素点的空域信息和值域信息,达到保持边缘、降噪平滑的效果<sup>[20]</sup>. 双边滤波利用周边像素亮度值的加权平均代表某个像素的强度,所用的加权平均基于高斯分布<sup>[21]</sup>. 最重要的是,双边滤波的权重不仅考虑了像素的欧氏距离,还考虑了像素范围域中的辐射差异<sup>[22]</sup>. 双边滤波的公式为

$$I_E = \frac{1}{k_p} \sum_{i,j \in \Omega_p} g_{\sigma_d}(\|(x,y) - (i,j)\|) \times$$

$$g_{\sigma_r}(\|I(x,y) - I(i,j)\|)I(x,y), \quad (2)$$

$$g_{\sigma_d}(\|(x,y) - (i,j)\|) = \exp\left(-\frac{(x-i)^2 + (y-i)^2}{2\sigma_d^2}\right), \quad (3)$$

$$g_{\sigma_r}(\|I(x,y) - I(i,j)\|) = \exp\left(-\frac{\|I(x-i) + I(y-i)\|^2}{2\sigma_r^2}\right)k_p. \quad (4)$$

其中:  $k_p$  是归一化常数,  $g_{\sigma_d}$  和  $g_{\sigma_r}$  分别表示空间位置和像素大小的滤波核,  $\sigma_d$  和  $\sigma_r$  表示空间位置和像素大小的标准偏差,  $I(i,j)$  表示像素点坐标为  $I(x,y)$  的邻域像素,  $\Omega$  表示用于图像  $I$  的滤波窗口. 带有噪声的红外图像经过双边滤波后,可以平滑噪声,并能保持其边缘特征,有利于后续图像融合任务的进行.

## 2 所提出算法

### 2.1 网络结构

本文使用端到端的卷积神经网络来实现红外与可见光图像的特征提取、特征融合和图像重建,可简洁、高效地完成红外与可见光图像融合任务. 所提出算法的整体网络结构如图2所示.

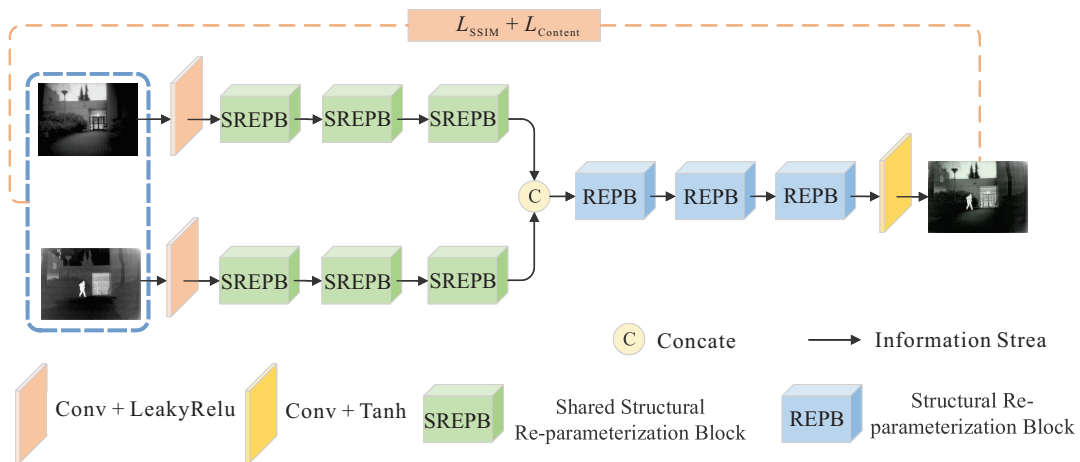


图2 算法框架

特征提取网络由红外与可见光这两个特征提取分支构成. 在训练阶段,网络第1层用  $1 \times 1$  卷积提取图像浅层的特征,接着用3个结构重参数化模块(SREPB)进一步提取浅层特征中的细节和深层特征. 网络模块在对称分支中每一个结构重参数化模块的最后一层对应的  $3 \times 3$  卷积的参数进行权重共享. 然后,将红外与可见光特征进行通道级联后送入重建网络,这使得网络所提取的特征能够全部保留,从而促使融合图像拥有更多的源图像信息. 最后,重建网络由3个结构重参数化模块(REPB)和  $1 \times 1$  卷积构成,激活函数为 ReLU,经过 Tanh 层得到融合图像. 训练完成后,使用所提出的结构重参数化操作对训练网络

进行等效转换,从而得到推理测试阶段的最终网络.

与单分支卷积神经网络的方法相比,本文中双分支卷积神经网络的方法能够更好地提取不同模态图像的特征信息,而双分支的方法又会带来不同模态信息的信息域差距较大、融合不同模态特征较难的问题. 为此,本文算法在特征提取过程还设计了对称分支的部分卷积层权重共享操作.

### 2.2 结构重参数化模块

本文所提出算法的主体部分都是由图3所示的结构重参数化模块构成. 在训练阶段,一个结构重参数化模块由两个包含  $3 \times 3$  卷积层、 $1 \times 1$  卷积层与恒等映射层并联的块和一个  $1 \times 1$  卷积层、共享权重的

$3 \times 3$ 卷积层的块组成.而在测试阶段,根据结构重参数化可将训练阶段的 $1 \times 1$ 卷积和恒等映射层看作特殊的 $3 \times 3$ 卷积,然后根据卷积的可加性得到测试阶段的单分支.特征提取部分的SREPB与图像重建部分的REPB略有不同,前者具有后者没有的对称分支权重共享操作,如灰色虚线所示.该策略可以减少网络

的参数量,并且红外与可见光图像是两种模态差异较大的图像,在对称分支中采用权重共享,还能使红外与可见光图像特征以相同的方式提取更深层的特征并补偿到主信息流中,这能够拉近二者不同模态图像特征的信息域,加强不同模态信息层次间的交流,方便后续特征融合的进行.

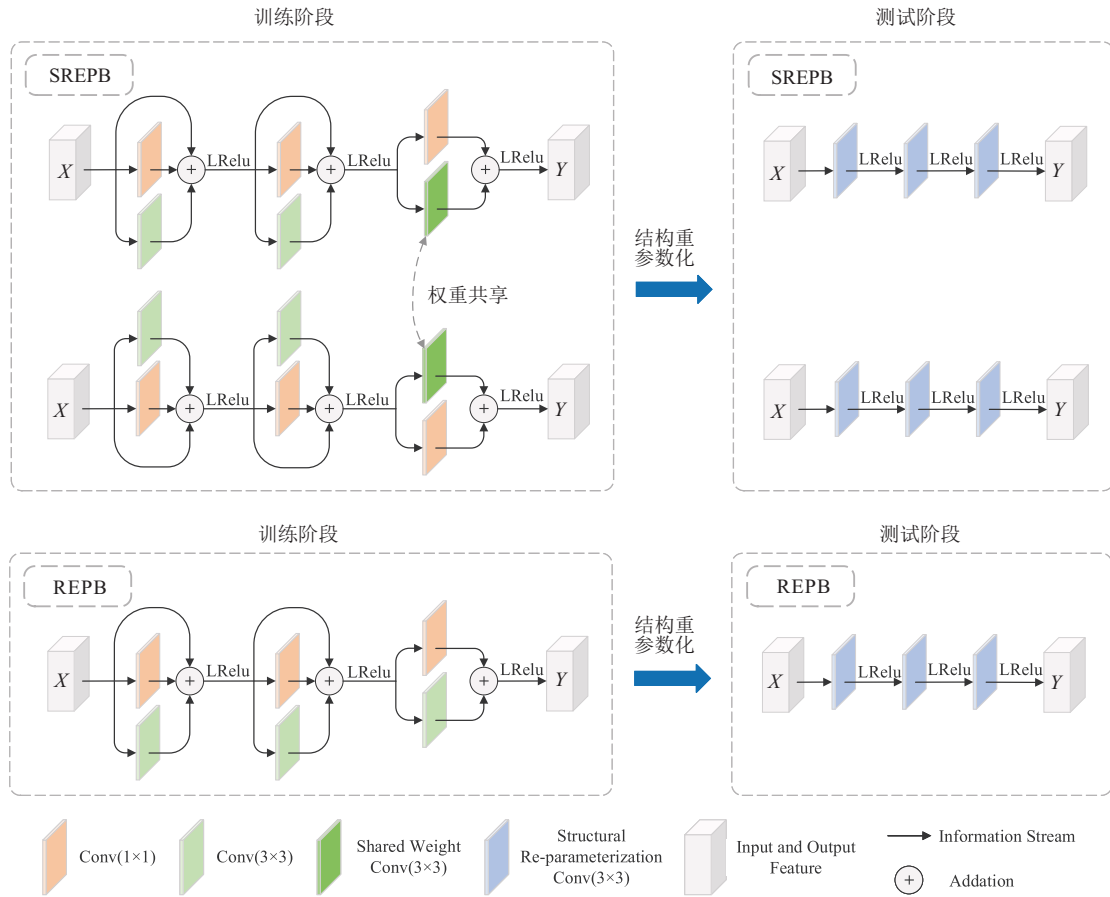


图3 结构重参数化模块

### 2.3 损失函数

损失函数是指导深度神经网络的训练达到预期效果的关键.结构相似性损失常用来保持融合图像清晰的强度分布,为了促进纹理细节的恢复,本文还设计了双边滤波去噪的内容损失来促使网络关注融合结果的显著性与边缘纹理.本文用结构相似性损失 $L_{SSIM}$ 和内容损失 $L_{Content}$ 共同约束网络的训练,总损失公式为

$$Loss = \lambda L_{SSIM} + L_{Content}, \quad (5)$$

其中 $\lambda$ 为平衡这两个损失的权重系数.

#### 2.3.1 结构相似性损失

结构相似性(SSIM)可以有效地评估两幅图像与融合图像之间的结构相似性.在融合任务中,让融合图像与红外图像和可见光图像之间达到平衡,使得图像的视觉效果比较符合人眼感知的视觉效果.结构

相似度损失公式如下:<sup>2</sup>

$$L_{SSIM} = 1 - \frac{SSIM(I_f, I_{vi}) + SSIM(I_f, I_{ir})}{2}, \quad (6)$$

$$L_{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (7)$$

其中: $I_f$ 表示融合图像, $I_{ir}$ 和 $I_{vi}$ 分别表示红外图像和可见光图像; $L_{SSIM}(x, y)$ 计算两个图像间的结构相似度, $\mu_x$ 和 $\mu_y$ 是两个源图像中的所有像素的平均值; $\sigma_x$ 和 $\sigma_y$ 是两个源图像的像素值的方差; $C_1$ 、 $C_2$ 是常数,用来保证函数的稳定性.

#### 2.3.2 内容损失

损失函数仅用结构相似性损失会导致融合图像的显著性被削弱,为了促进融合模型保留红外图像里的显著性以及源图像的边缘纹理信息,本文设计带有双边滤波去噪的内容损失.内容损失由强度损失和双边滤波边缘梯度损失两部分组成,其公式为

$$L_{\text{Content}} = \mu_1 L_{\text{in}} + \mu_2 L_{\text{grad}}, \quad (8)$$

其中  $\mu_1$ 、 $\mu_2$  为平衡两个损失的权重系数。

强度损失约束融合图像的整体表现强度,为了更好地保留显著目标,融合图像的像素强度要偏向于红外与可见光图像的强度最大值. 强度损失的公式为

$$L_{\text{in}} = \frac{1}{HW} \|I_f - \text{Max}(I_{ir}, I_{vi})\|_1. \quad (9)$$

其中:  $H$ 、 $W$  分别是输入图像的高度和宽度,  $\|\cdot\|_1$  代表范数。

为了让网络模型更好地保留融合图像的边缘纹理细节,已有方法中会使用源图像的边缘梯度最大值来约束网络的训练,但这种损失很容易受到红外图像中噪声的影响. 为此,本文使用能够保留边缘梯度的双边滤波对红外图像去噪,从而减少融合图像的噪声. 双边滤波处理前后的梯度最大图像对比示例如图4所示. 双边滤波边缘梯度损失计算公式为

$$L_{\text{grad}} = \frac{1}{HW} \|\|\nabla I_f| - \text{Max}(|\nabla \text{Bila}(I_{ir})|, |\nabla I_{vi}||)\|_1. \quad (10)$$

其中:  $\nabla$  是测量图像纹理信息的梯度算子,在本文中使用了 Sobel 算子计算梯度;  $\text{Bila}$  表示双边滤波器。



(a) 双边滤波前的最大梯度



(b) 双边滤波后的最大梯度

图4 双边滤波处理前后的对比结果

### 3 实验分析

#### 3.1 实验设置

本文实验均在 GeForce RTX 2080Ti 11GB 和 Intel Core i5-12600KF 上进行,深度学习框架为 PyTorch. 在实验训练阶段,由于现有红外与可见光图像融合数据集较小,将已配准的红外与可见光图像公共数据集 MSRS 作为训练集部分,从 1 083 个图像对扩充到

26 112 个图像对,切割后的图像对尺寸大小为 64 pixel × 64 pixel,基本能满足训练要求. 网络优化器采用 Adam, epoch = 8, batch\_size = 64, 学习率为  $1 \times 10^{-4}$ , 损失函数参数为  $\lambda = 1$ 、 $\mu_1 = 10$ 、 $\mu_2 = 20$ , 双边滤波的参数为  $\sigma_d = 0.05$ 、 $\sigma_r = 8.0$ 、窗口大小 =  $11 \times 11$ . 优化器 Adam 具有占用内存小、适合处理稀疏梯度的特点,是目前广泛用于各领域且表现较佳的优化算法. 选用公共数据集 TNO、RoadScene、MSRS 作为测试集部分,各选择 42、20、25 对图像进行算法对比实验. 实验中所有对比算法均以原论文的设置进行实验. 本文算法训练流程的伪代码如算法 1 所示。

#### 算法1 训练流程.

Input: 红外图像  $I_{ir}$  和可见光图像  $I_{vi}$ ;

Output: 融合图像  $I_f$ .

- 1 for  $M$  epochs do
- 2   for  $p$  steps do
- 3      $n$  张红外图像  $\{I_{ir}^1, I_{ir}^2, \dots, I_{ir}^n\}$ ;
- 4      $n$  张可见光图像  $\{I_{vi}^1, I_{vi}^2, \dots, I_{vi}^n\}$ ;
- 5     用本文算法生成  $n$  张融合图像  $\{I_f^1, I_f^2, \dots, I_f^n\}$ ;
- 6     通过本文损失计算网络损失;
- 7     通过 Adam 优化器更新网络参数;
- 8     保存第  $M$  轮的网络模型权重;
- 9   // 模型训练结束,结构重参数化开始;
- 10   for  $q$  steps do
- 11     读取当前网络模块及其参数;
- 12     if 该模块有 “switch\_to\_deploy” 方法 then
- 13       计算重参数化后的卷积核和偏置参数值;
- 14       将参数赋值给新创建的卷积块;
- 15       删除原网络结构;
- 16     保存结构重参数化后的网络模型和权重.

#### 3.2 对比实验

为全面地评估本文方法的有效性,在 TNO<sup>[23]</sup>、RoadScene<sup>[24]</sup>、MSRS<sup>[13]</sup> 3 个公共数据集上进行定性和定量对比实验. 将本文算法与 7 种融合算法进行比较,包括 DenseFuse<sup>[11]</sup>、FusionGAN<sup>[15]</sup>、RFN-Nest<sup>[25]</sup>、IFCNN<sup>[26]</sup>、PMGI<sup>[27]</sup>、U2Fusion<sup>[24]</sup>、SDNet<sup>[28]</sup>. 在 3 个数据集上的可视化实验结果如图 5 所示。

由实验结果可看出: DenseFuse 采用自编码器网络和简单的融合策略,融合图像既未能突出显著目标,纹理细节也比较差; RFN-Nest 采用残差结构进行融合,然而融合图像中的红外目标对比度较弱; FusionGAN 采用生成对抗网络无监督地实现融合任务,在一定程度上凸显红外显著目标,但丢失了大量背景细节纹理; IFCNN 以端到端的形式实现融合,



图5 在TNO、RoadScene和MSRS数据集上的算法对比结果

突出显著目标,但目标边缘模糊;PMGI和U2Fusion会削弱红外目标;SDNet能突出目标,却削弱了背景纹理特征.本文所提出算法生成的融合图像能够在突出显著目标的同时更好地保留了背景细节纹理.

然而,可视化的实验结果具有主观性,且由于目前的红外与可见光图像融合数据集均无参考图像,无法直接通过参考图像评价融合结果的质量情况,故引入客观的通用图像质量评价指标进行评估.本

文选用6种指标对上述算法进行评估,分别是标准差(standard deviation, SD)、互信息(mutual information, MI)、视觉保真度(visual information fidelity, VIF)、差异相关和(sum of correlation differences, SCD)、信息熵(entropy, EN)、空间频率(spatial frequency, SF). SD反映图像的对比度;MI衡量源图像到融合图像的信息量;VIF则以人类视觉系统量化源图像与融合图像的信息量;SCD度量源图像和融合图像的差异;EN衡量图像包含的信息量;SF则是测量图像的梯度.以上指标均是正向指标,值越大意味着融合结果越好.

对比指标结果如表1所示,其中:红色代表最优

结果,蓝色次之.本文方法在3个数据集MI、VIF、EN上的值均为最优:MI值最优说明融合图像中包含最多源图像的信息;VIF值最优意味着本文方法融合的图像更符合人类视觉系统;EN值最优则能说明本文方法包含的信息量最多.表1中SD值、SF值在对比方法中处于最优或次优的位置,说明融合图像的对比度较高,且能反映图像背景的纹理细节,较为符合人类的视觉感知.此外,本文还对比了各算法在相同条件下的平均运行时间,结果如表2所示.本文所提出算法基于结构重参数化的思想,在推理阶段网络简洁高效,模型复杂度低,可大幅减少融合时间.

表1 对比指标结果

数据集	算法	指标					
		SD	MI	VIF	SCD	EN	SF
TNO	DenseFuse	8.5765	2.1987	0.6704	1.5916	6.3422	0.0248
	RFN-Nest	9.3153	2.1268	0.8103	1.7711	6.9285	0.0226
	FusionGAN	8.6058	2.3426	0.6457	1.3688	6.5199	0.0240
	IFCNN	9.0058	2.4154	0.7996	1.6850	6.7413	0.0471
	PMGI	9.5824	2.3627	0.8612	1.6895	6.9864	0.0337
	SDNet	9.0398	2.2710	0.7523	1.5488	6.6670	0.0448
	U2Fusion	8.8553	1.8730	0.6787	1.5862	6.4230	0.0327
	Ours	9.5492	3.1619	0.9669	1.6169	7.1455	0.0431
RoadScene	DenseFuse	9.2881	2.5912	0.6022	1.5109	6.7180	0.0330
	RFN-Nest	9.7761	2.4605	0.6347	1.6549	7.1447	0.0290
	FusionGAN	9.6237	2.5381	0.5286	1.2061	6.8825	0.0300
	IFCNN	10.1064	2.7068	0.6892	1.5700	7.0790	0.0593
	PMGI	9.6297	3.0025	0.6879	1.4430	7.1064	0.0377
	SDNet	9.6120	3.0125	0.6747	1.3855	7.0858	0.0518
	U2Fusion	9.4077	2.4246	0.5855	1.3753	6.7512	0.0440
	Ours	10.2965	3.5981	0.8505	1.4185	7.3261	0.0602
MSRS	DenseFuse	7.0692	2.5409	0.6752	1.3296	5.8397	0.0235
	RFN-Nest	6.9939	2.3430	0.5364	1.2881	5.7514	0.0181
	FusionGAN	5.4307	1.9160	0.4234	0.7948	5.2179	0.0146
	IFCNN	7.5947	2.7399	0.8283	1.6658	6.3109	0.0456
	PMGI	7.2221	2.1656	0.6071	1.2242	6.0063	0.0274
	SDNet	5.3143	1.7396	0.3745	0.8298	4.88512	0.0270
	U2Fusion	5.6231	1.8953	0.3967	1.0034	4.7525	0.0256
	Ours	7.8719	3.8022	0.8609	1.7483	6.3351	0.0411

表2 各算法平均运行时间

算法	DenseFuse	RFN-Nest	FusionGAN	IFCNN	PMGI	SDNet	U2Fusion	Ours
运行时间/s	0.874	0.473	0.612	0.097	0.416	0.341	0.855	0.006

### 3.3 验证实验

目标检测算法是一种重要的计算机视觉方法,其性能可以很好地反映融合图像中的语义信息.为了验证融合图像对后续视觉任务的促进作用,本节将官方预训练的目标检测模型YOLOv5s用于评估融合图像的目标检测性能,目标检测结果如图6所示.

从实验结果看,融合结果图像的预测结果均比红外与可见光图像要好一些,尤其是夜间光照条件比较

差时,可见光传感器很难捕捉到足够的信息,因此目标检测模型往往很难检测出行人;而在融合图像中,可以较好地识别出行人,在一定程度上说明图像融合后可对后续视觉任务有提升作用.

### 3.4 消融实验

为了验证网络中加入权重共享结构重参数化模块与带双边滤波去噪的内容损失的有效性,本节设计了如下的消融实验:实验1为不加重共享结构重参

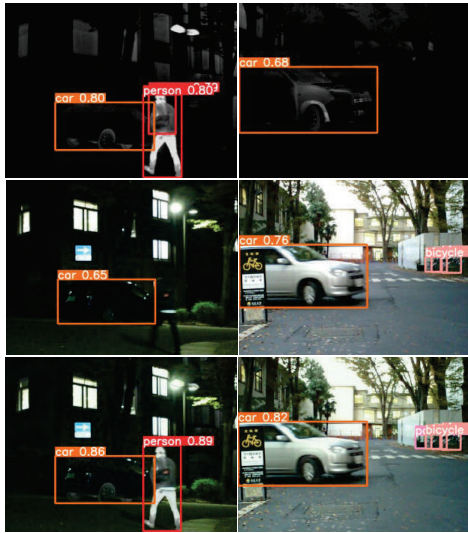


图6 目标检测验证实验结果

数化模块的直连型网络、加双边滤波内容损失的方法;实验2为加权重共享结构重参数化模块的训练网络、不加双边滤波内容损失的方法;实验3为加权重共享的结构重参数化训练网络、加双边滤波内容损失的本文方法. 本实验在TNO数据集上进行,消融实验结果如图7所示.

从实验结果可以看出,实验3和实验1整体视觉效果会比实验2好,说明双边滤波内容损失能有效指导网络生成视觉感官较好的结果,而实验3比实验1在细节纹理上更平滑,视觉感官更自然. 表3是消融实验的客观评价指标结果,本文的方法在各项指标的值较实验1和实验2有所提升,表明了本文方法的优越性. 因此,这在指标层面也侧面说明了权重共享结构重参数化模块与双边滤波内容损失的有效性.

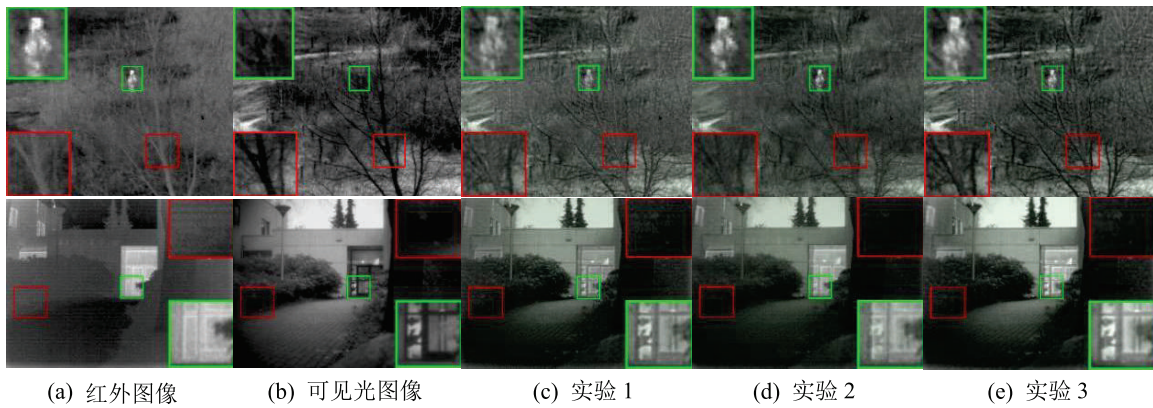


图7 消融实验结果

表3 消融实验指标结果

消融实验	指标					
	SD	MI	VIF	SCD	EN	SF
实验1	9.4191	2.8446	0.9358	1.5968	7.1028	0.0402
实验2	9.3790	2.6630	0.8940	<b>1.6212</b>	7.0504	0.0370
实验3	<b>9.5492</b>	<b>3.1619</b>	<b>0.9669</b>	1.6169	<b>7.1455</b>	<b>0.0431</b>

此外,参考MSRS数据集中所用的图片尺寸,使用 $640 \times 480 \times 1$ 的数据作为网络前向传播的输入,通过TorchSummary库计算网络模型的数量和权重大小,对结构重参数化前后的运行时间、参数量、权重大小以及融合结果逐像素点累计偏差进行实验对比,对比结果如表4所示. 由对比结果可知,在结构重参数化前后网络的融合结果几乎没有差别,说明结构重参数化可以在极低误差的情况下,有效地减少运行时间、参数量和权重大小.

表4 结构重参数化前后对比结果

是否结构重参数化	运行时间/s	参数量	权重大小	偏差
×	0.013	653 657	2.49 MB	/
√	0.006	587 457	2.24 MB	$4.65 \times e^{-6}$

## 4 结论

本文提出了一种基于结构重参数化的红外与可见光图像融合算法. 首先,网络主体部分通过结构重参数化模块对特征进行提取,同时提取图像的细节纹理特征和显著性特征,并对所提取的特征进行图像重建. 接着,在训练完成后,对网络进行结构重参数化操作,将残差连接型网络等效转换成直连型网络,在保持原有训练效果的基础上,使得推理速度和推理所需存储空间得到优化. 最后,主观和客观实验结果表明,所提出算法与其他算法相比,其融合结果能达到更优的效果,并具有更好的实时性和更少的推理内存占用,在实际工程应用中,更适合定制型硬件加速电路的设计. 此外,本文算法是基于结构重参数化的典型结构进行改进的,这种相对简洁的结构并未能充分提取图像特征,随着结构重参数化方法的深入发展,越来越多学者将会探寻效果更佳的复杂结构训练网络与新的结构重参数化的方式,未来会考虑采用更优的结构重参数化的网络结构,让网络模型更好地学习到多尺度图像特征,从而在较低计算资源消耗下达到更

佳的融合效果.

### 参考文献(References)

- [1] Zhang H, Xu H, Tian X, et al. Image fusion meets deep learning: A survey and perspective[J]. *Information Fusion*, 2021, 76: 323-336.
- [2] 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*, DOI: 10.13195/J.kzyjc.2022.0808.
- [3] 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.
- [4] Liu J, Lei Y J, Xing Y Q, et al. Innovative image fusion method based on improved NSST[J]. *Control and Decision*, 2017, 32(2): 275-280.
- [5] 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.
- [6] Wang J, Peng J Y, Feng X Y, et al. Fusion method for infrared and visible images by using non-negative sparse representation[J]. *Infrared Physics & Technology*, 2014, 67: 477-489.
- [7] 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.
- [8] Liu Y, Liu S P, Wang Z F. A general framework for image fusion based on multi-scale transform and sparse representation[J]. *Information Fusion*, 2015, 24: 147-164.
- [9] 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.
- [10] 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.
- [11] Li H, Wu X J. DenseFuse: A fusion approach to infrared and visible images[J]. *IEEE Transactions on Image Processing*, 2019, 28(5): 2614-2623.
- [12] 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*, 2022, 32(1): 105-119.
- [13] 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.
- [14] Liu R S, Liu Z, Liu J Y, et al. Searching a hierarchically aggregated fusion architecture for fast multi-modality image fusion[C]. *Proceedings of the 29th ACM International Conference on Multimedia*. New York, 2021: 1600-1608.
- [15] 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.
- [16] Ma J Y, Xu H, Jiang J J, et al. DDcGAN: A dual-discriminator conditional generative adversarial network for multi-resolution image fusion[J]. *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society*, 2020: 4980-4995.
- [17] Karim S, Tong G, Li J Y, et al. Current advances and future perspectives of image fusion: A comprehensive review[J]. *Information Fusion*, 2023, 90: 185-217.
- [18] Ding X H, Zhang X Y, Ma N N, et al. RepVGG: Making VGG-style ConvNets great again[C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, 2021: 13728-13737.
- [19] Ding X H, Zhang X Y, Han J G, et al. Diverse branch block: Building a convolution as an inception-like unit[C]. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville, 2021: 10881-10890.
- [20] Tomasi C, Manduchi R. Bilateral filtering for gray and color images[C]. *The 6th International Conference on Computer Vision*. Bombay, 2002: 839-846.
- [21] Sudeep P V, Issac Niwas S, Palanisamy P, et al. Enhancement and bias removal of optical coherence tomography images: An iterative approach with adaptive bilateral filtering[J]. *Computers in Biology and Medicine*, 2016, 71: 97-107.
- [22] Zhang Y, Wang M, Xia X P, et al. Medical image fusion based on quasi-cross bilateral filtering[J]. *Biomedical Signal Processing and Control*, 2023, 80: 104259.
- [23] Toet A. TNO image fusion dataset[DB/OL]. [2023-03-05]. [https://figshare.com/articles/dataset/TNO\\_Image\\_Fusion\\_Dataset/1008029](https://figshare.com/articles/dataset/TNO_Image_Fusion_Dataset/1008029).
- [24] 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.
- [25] 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.
- [26] Zhang Y, Liu Y, Sun P, et al. IFCNN: A general image fusion framework based on convolutional neural network[J]. *Information Fusion*, 2020, 54: 99-118.
- [27] 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.
- [28] 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.

### 作者简介

陈昭宇(1996—), 男, 硕士生, 从事图像融合、图像处理等研究, E-mail: 444019139@qq.com;

范洪博(1982—), 男, 讲师, 博士, 从事图像融合、图像处理等研究, E-mail: qingchen233@foxmail.com;

马美燕(1996—), 女, 硕士生, 从事图像融合的研究, E-mail: 1845976781@qq.com;

赵艺博(1998—), 男, 硕士生, 从事图像分割、图像处理等研究, E-mail: 1248784084@qq.com.