

引用格式: HAO Shuai, LI Jiahao, MA Xu, et al. Infrared and Visible Image Fusion Algorithm Based on Feature Optimization and GAN[J]. Acta Photonica Sinica, 2023, 52(12):1210004

郝帅,李嘉豪,马旭,等.基于特征优化和GAN的红外与可见光图像融合算法[J].光子学报,2023,52(12):1210004

# 基于特征优化和GAN的红外与可见光 图像融合算法

郝帅,李嘉豪,马旭,何田,孙思雅,李彤

(西安科技大学 电气与控制工程学院, 西安 710054)

**摘要:**针对红外与可见光融合图像存在纹理细节不丰富、对比度较低及目标信息损失等问题,提出了一种基于特征优化和生成对抗网络的图像融合算法。首先,设计一种自适应特征优化模块以增强原始图像纹理细节及对比度;然后,为使融合图像保留更多的多模态信息,将生成对抗网络引入到融合框架中。在生成器模型中,考虑到红外与可见光图像成像机理差异,构建了双支路特征提取网络,并设计多尺度密集连接模块以提取异源图像丰富的特征信息;其次,在融合层构造通道和空间注意力模型以增强局部特征之间联系,减小融合图像中目标信息损失;最后,为使融合结果尽可能保留可见光纹理细节的同时又能够较好突出红外目标,构造了双判别器网络结构。为验证所提算法优势,在TNO数据集上进行实验,并与6种经典融合算法进行主观和客观比较。实验结果表明,所提算法无论在主观还是客观评价上均具有明显优势,生成的融合图像纹理细节更为丰富、边缘及目标更加清晰且具有更好的对比度,客观评价指标信息熵、空间频率、相关熵、视觉保真度和梯度信息分别提高了16.11%、65.46%、7.96%、42.67%和33.24%。

**关键词:**图像融合;特征优化;生成对抗网络;多尺度密集连接;注意力模型

中图分类号: TP391

文献标识码: A

doi:10.3788/gzxb20235212.1210004

## 0 引言

图像融合作为一种重要的图像增强方法,旨在从同一场景不同模态传感器捕获的图像中提取互补信息并进行融合,从而增强视觉效果,辅助人们对场景进行理解<sup>[1-2]</sup>。红外热成像和可见光成像作为两种最常用的探测手段已有广泛应用。红外传感器主要通过目标场景的热辐射信息成像,能够捕获物体的热显著目标,具有不易受外界天气、光照强度等环境因素影响的优点,但其图像分辨率往往较低,纹理细节信息相对较少。可见光传感器依据物体的反射光线成像,能够捕获丰富的细节及背景信息,其图像细节分辨能力强,但易受外界光照强度变化、异物遮挡等因素干扰<sup>[3-4]</sup>。因此,结合两类图像各自优势,将其进行融合,可以得到纹理细节丰富、目标显著及视觉效果好的图像。目前,该方法已在目标检测<sup>[5]</sup>、目标跟踪<sup>[6]</sup>、侦察<sup>[7]</sup>和医学图像分析<sup>[8]</sup>等领域展开广泛研究和应用。

根据特征信息提取方式不同,红外与可见光图像融合算法可以分为两类:传统融合算法和基于深度学习的融合算法。传统融合算法主要通过相关数学模型来获取原始图像不同层次信息,再通过手工设计合适的规则对信息进行融合,主要包括基于多尺度变换和基于表示学习的融合算法。基于多尺度变换的融合算法主要将原始图像在频域中分解为不同分辨率、不同尺度的若干层子图像,然后根据设定的融合规则对分解的子图像进行组合,最后通过多尺度逆变换得到融合图像。常见的多尺度变换方法有金字塔变换、小波

**基金项目:**国家自然科学基金(No. 51804250),中国博士后科学基金(No. 2020M683522),陕西省科技计划项目(Nos. 2021JQ-572, 2020JQ-757)

**第一作者:**郝帅,haoxust@163.com

**通讯作者:**马旭,maxu@xust.edu.cn

**收稿日期:**2023-05-10;**录用日期:**2023-07-24

<http://www.photon.ac.cn>

变换、轮廓波变换等。CHEN Jun等<sup>[9]</sup>利用拉普拉斯金字塔变换将原始图像分别分解为低频带和高频带,对高频带使用最大绝对值融合规则,低频带由分解的红外低频信息来确定融合权重,最后使用拉普拉斯逆变换来重建融合图像。基于表示学习的融合算法是直接空间域上处理图像像素,常用的主要包含稀疏表示和低秩表示。LU Xiaoqi等<sup>[10]</sup>利用核密度估计聚类 and 奇异值分解数学模型提取出目标区域,并对背景信息进行稀疏表示,从而实现红外与可见光图像融合。然而,这些传统融合算法通常需要手工设计融合规则,计算过程较为复杂。

近年来,由于卷积神经网络(Convolutional Neural Network, CNN)具有较强的特征提取能力,基于深度学习的融合算法被相继提出。LI Hui等<sup>[11]</sup>提出了用于红外和可见光图像融合的DenseFuse,该算法通过设计编码器和解码器网络来分别提高特征提取能力和重建能力。LIU Yu等<sup>[12]</sup>基于CNN构建了图像融合模型,通过活动水平测量和权重分配来实现图像融合。JIAN Lihua等<sup>[13]</sup>结合残差网络构建了一种对称编码器-解码器结构,该结构可以有效保留每层卷积提取的特征信息。然而,由于红外与可见光图像融合任务难以定义融合效果标准,且没有Ground Truth指导网络训练,导致该类方法在训练过程中无法有效估计原始图像特征分布,易产生特征分布不均衡的融合结果。

MA Jiayi等<sup>[14]</sup>提出了一种基于生成对抗网络(Generative Adversarial Network, GAN)<sup>[15]</sup>的融合算法,生成器负责提取红外与可见光图像特征并生成融合图像,而判别器用于将融合图像与可见光图像进行区分,使融合图像在对抗博弈中能够保留足够的梯度信息。在此基础上,MA Jiayi等<sup>[16]</sup>又构建了双判别器生成对抗网络模型,通过设计红外和可见光双判别器网络来区分融合图像与原始图像之间的结构差异,从而使融合结果能够同时保留原始图像特征信息。

基于GAN的融合算法在一定程度上解决了传统融合算法的不足,可以更高效地提取图像特征信息,且通过网络模型对抗训练有效地平衡了原始图像特征分布,但依然存在不足:1)现有的融合算法主要集中在融合模型的构建,并未考虑原始图像质量对最终融合效果的影响;2)生成器中大都采用单一尺度卷积提取图像特征,易造成图像特征提取不充分,进而导致融合图像无法全面保留原始特征信息;3)由于没有考虑原始图像局部特征的全局依赖性,使得局部特征没有得到细化和增强,从而导致融合图像丢失重要目标特征信息。

针对上述问题,本文提出一种基于特征优化和GAN的红外与可见光图像融合算法。考虑到原始图像质量对融合结果的影响,设计了一种基于变色龙算法(Chameleon Swarm Algorithm, CSA)<sup>[17]</sup>的目标函数自适应特征优化模块以增强可见光图像的纹理细节和红外图像的对比度。为解决生成器中采用单一尺度卷积层提取特征造成特征提取不充分的问题,构造了一种多尺度密集连接模块(Feature extraction module based on Multi-Scale Dense Connection, MSDC-Fem),从而增大网络感受野特征提取范围,以全面提取图像的深层语义特征和浅层纹理特征。为减小融合过程中重要目标特征信息损失,在特征融合层设计了基于空间和通道的并联型注意力模型。通过将红外与可见光图像特征信息分别同时送入空间和通道注意力模型中,捕捉不同模态特征之间的相关性和依赖关系,提高网络对关键信息的表达能力,从而更好地聚焦红外图像中的热显著目标和可见光图像中的纹理细节。

## 1 相关知识

### 1.1 潜在低秩表示

低秩表示(Low Rank Representation, LRR)<sup>[18]</sup>是在确定学习字典情况下,将原始数据矩阵表示为字典矩阵下的线性组合且表示系数矩阵低秩,以实现数据的空间分割与特征提取。但LRR无法保留图像局部结构信息,在LRR基础上,潜在低秩表示(Latent Low Rank Representation, LatLRR)<sup>[19]</sup>通过考虑隐藏数据信息对学习字典的影响来提取数据全局结构信息和局部结构信息,其相较于LRR具有更强的特征提取能力。LatLRR的数学模型可表示为

$$\min_{B, D, N} \|B\|_* + \|D\|_* + \lambda \|N\|_1 \quad \text{s.t.} \quad X = BX + DX + N \quad (1)$$

式中, $B$ 为原始数据最优的LRR系数矩阵, $D$ 为显著系数矩阵, $N$ 为稀疏噪声矩阵, $\|\cdot\|_*$ 为核范数, $\|\cdot\|_1$ 为L1范数, $X$ 为原始数据矩阵, $\lambda$ 为正则化平衡参数且大于0。

将式(1)通过增广拉格朗日乘法求解,得到系数 $B$ 和 $D$ 。以红外图像为例,假设红外图像为 $X_1$ ,经过LatLRR分解后可表示为

$$X_1 = BX_1 + DX_1 + N \quad (2)$$

式中, $BX_1$ 表示低秩分量, $DX_1$ 表示显著分量。

## 1.2 生成对抗网络

GAN是使用对抗性模型来估计样本分布并生成新数据的无监督网络模型,它主要由生成器(Generator,G)和判别器(Discriminator,D)两部分构成。生成器学习训练集数据特征,并在判别器指导下,将随机噪声分布尽量拟合为训练数据的真实分布,从而生成具有训练集特征的相似数据。判别器负责区分输入是真实数据或生成器生成的假数据,并将判断结果反馈给生成器。两个网络交替训练,直到生成器生成的数据能够以假乱真,并与判别器的能力达到一种纳什均衡状态。G与D的对抗关系为

$$\min_G \max_D V_{GAN}(G, D) = E_{x \sim P_{data}(x)} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (3)$$

式中, $x$ 表示输入样本, $z$ 表示输入到生成器的噪声, $P_{data}(x)$ 表示真实数据分布, $P_z(z)$ 表示噪声分布, $D(x)$ 表示判别器判断真实数据是否真实的概率, $D(G(z))$ 表示判别器判断生成数据是否真实的概率。

由于GAN具有在无监督情况下学习真实数据分布的能力,故利用生成器生成融合图像,并通过判别器以期使融合图像在对抗学习中能够尽可能地保留原始图像的丰富信息。

## 2 算法原理

本文所提算法框图如图1所示,主要由三部分构成:特征优化模块、生成器和判别器。特征优化模块在对原始红外和可见光图像进行增强,提高其特征表达能力;生成器的主要任务是对输入图像的特征进行提取、融合及重构,生成融合图像;判别器由可见光判别器(Discriminator-VIS)和红外判别器(Discriminator-IR)两部分构成,分别与生成器进行对抗学习,从而保证融合图像中既保留红外图像的对比度信息又具有可见光图像的梯度信息。

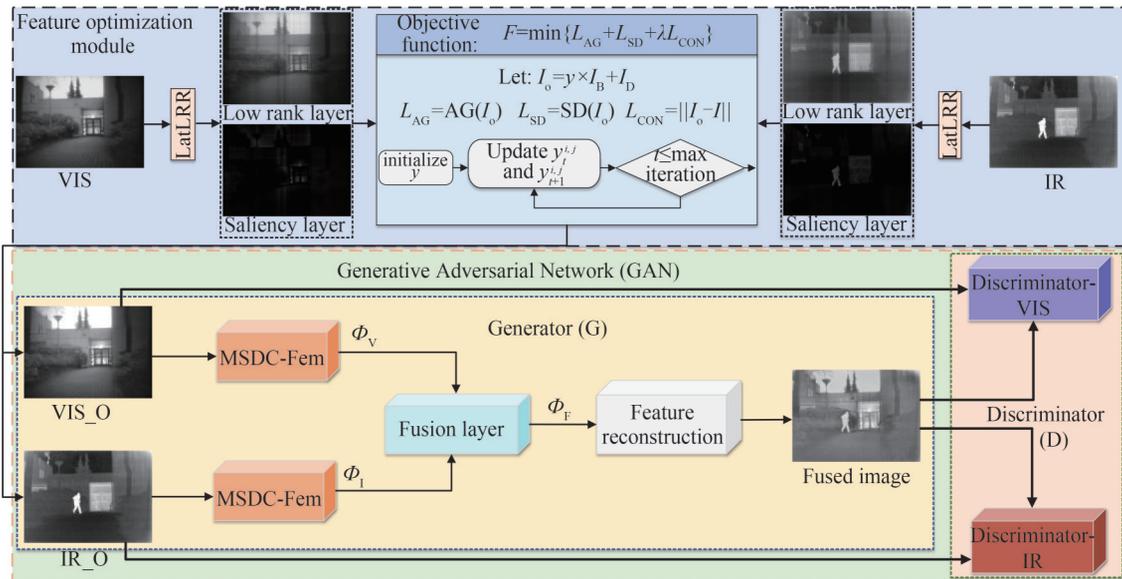


图1 本文算法框图

Fig.1 Block diagram of the proposed algorithm

## 2.1 自适应特征优化模块设计

由于受外界环境影响,可见光图像会因光照强度变化导致纹理细节不清晰,而红外图像受热辐射成像机理影响,易出现目标对比度低等问题,会影响最终的图像融合效果。为此,设计了一种自适应特征优化模块,以增强原始图像特征表达能力。首先,利用LatLRR对原始图像进行分解;然后,构建目标函数并设计优化模型对其进行迭代寻优;最后,得出优化因子 $y$ ,进而获得优化图像,为后续图像融合奠定基础。优化图像可表示为

$$I_o = yI_B + I_D \quad (4)$$

式中, $I_o$ 为优化后的图像, $y$ 为优化因子, $I_B$ 为图像低秩层, $I_D$ 为图像显著层。

### 2.1.1 基于LatLRR的图像分解

由于LatLRR充分考虑了图像数据的全局结构、局部结构和稀疏噪声等方面,不仅能够从源图像中提取全局结构信息和局部结构信息,还可以在在一定程度上分离图像中的噪声分量,减小视觉伪影。因此利用LatLRR分别对原始红外与可见光图像进行分解,并得到各自对应的低秩层和显著层,如图2所示。从图中可以看出,低秩层主要包含目标整体及背景细节信息,显著层则主要体现局部显著区域及边缘信息。

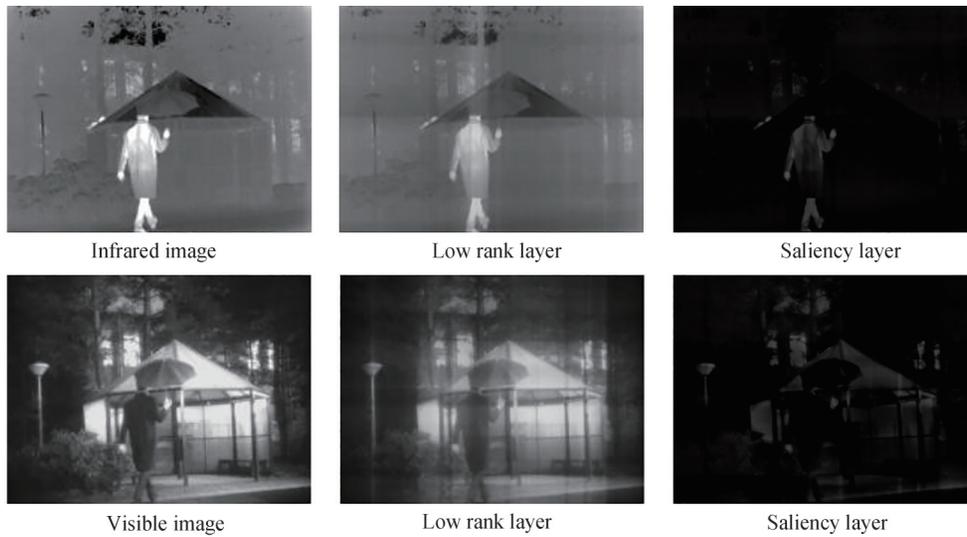


图2 LatLRR分解结果  
Fig.2 LatLRR decomposition results

### 2.1.2 目标函数构建

为了较好地增强原始图像纹理细节,突出显著目标,选取梯度变化、对比度变化和-content损失作为约束条件,构造目标函数 $F$ 为

$$F = \min(L_{AG} + L_{SD} + \lambda L_{CON}) \quad (5)$$

式中, $L_{AG}$ 为平均梯度函数, $L_{SD}$ 为标准差函数, $L_{CON}$ 为内容损失函数, $\lambda$ 表示保持两项平衡的正则化参数,取 $\lambda=1/300$ 。

平均梯度(Average Gradient, AG)用来度量图像梯度信息,且可以在一定程度上反映图像的细节纹理<sup>[20]</sup>。AG值越大,图像包含的梯度信息越多, $L_{AG}$ 定义为

$$L_{AG} = -\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \sqrt{\frac{\nabla I_{ox}^2(i,j) + \nabla I_{oy}^2(i,j)}{2}} \quad (6)$$

式中, $\nabla I_{ox}(i,j) = I_o(i,j) - I_o(i+1,j)$ , $\nabla I_{oy}(i,j) = I_o(i,j) - I_o(i,j+1)$ , $M$ 和 $N$ 为图像宽和高, $(i,j)$ 为图像横、纵坐标。

标准差(Standard Deviation, SD)主要用来反映图像对比度信息<sup>[21]</sup>,SD值越大,图像的对比度越高,视觉

效果越好, $L_{SD}$ 定义为

$$L_{SD} = - \sqrt{\sum_{i=1}^M \sum_{j=1}^N (I_o(i,j) - \phi)^2} \quad (7)$$

式中, $\phi$ 表示图像的平均值。

为了衡量优化过程中的目标信息损失程度,设计了内容损失函数 $L_{CON}$ ,定义为

$$L_{CON} = \|I_o - I\| \quad (8)$$

式中, $I$ 表示原始图像。

### 2.1.3 基于CSA的目标函数自适应优化

为得到优化因子 $y$ ,采用变色龙优化算法(CSA)进行求解。CSA主要模拟了变色龙在树木、沙漠和沼泽附近寻找食物时动态行为。该算法具有寻优能力强、收敛速度快、精度高等特点。为此,针对设计的目标函数 $F$ ,即式(5),将CSA引入到优化机制中,利用CSA进行寻优,从而求解优化因子 $y$ 。CSA步骤为:

1)初始化。初始种群是根据搜索空间中均匀随机初始化的变色龙数量和空间的维数创建的,即

$$y^i = l_j + r \times (u_j - l_j) \quad (9)$$

式中, $y^i$ 为第 $i$ 个变色龙的初始位置, $u_j$ 和 $l_j$ 分别表示搜索区域在第 $j$ 维的下界和上界, $r$ 为 $[0,1]$ 范围内均匀生成的随机数。

2)搜索猎物。变色龙觅食过程中的运动行为可以采用以下位置更新策略进行建模。

$$y_{t+1}^{i,j} = \begin{cases} y_{t+1}^{i,j} + p_1(p_{t+1}^{i,j} - G_t^j)r_2 + p_2(G_t^j - y_{t+1}^{i,j})r_1 & r_i \geq P_p \\ y_{t+1}^{i,j} + \mu \cdot l_b^j \text{sgn}(\text{rand} - 0.5) + \mu[(u_j - l_j)r_3] \text{sgn}(\text{rand} - 0.5) & r_i < P_p \end{cases} \quad (10)$$

式中, $y_{t+1}^{i,j}$ 为第 $i$ 只变色龙第 $j$ 维空间第 $t+1$ 次迭代位置, $y_t^{i,j}$ 为第 $i$ 只变色龙第 $j$ 维空间第 $t$ 次迭代位置, $p_1, p_2$ 为搜索能力控制系数, $P_{t+1}^{i,j}$ 为第 $i$ 只变色龙第 $j$ 维空间第 $t$ 次迭代最佳位置, $G_t^j$ 为变色龙第 $t$ 次迭代全局最优位置, $l_b^j$ 为变色龙在 $b$ 位置的第 $j$ 维的下界, $r_1, r_2, r_3$ 为 $[0,1]$ 范围内均匀生成的随机数, $P_p$ 为变色龙感知猎物概率。 $\text{sgn}(\text{rand} - 0.5)$ 对搜索和发现有影响,可以是1或-1, $\mu$ 为迭代函数的参数,随迭代次数的增加而减少,即

$$\mu = \gamma e^{(-\alpha t/T)^\beta} \quad (11)$$

式中, $T$ 为最大迭代次数, $t$ 为当前迭代次数, $\gamma, \alpha, \beta$ 用于控制搜索和发现的能力,分别设置为1、3.5、3。

3)眼睛旋转发现猎物。变色龙的眼睛能 $360^\circ$ 旋转进行猎物搜索,并根据猎物的位置来更新自己的位置,位置更新数学描述为

$$y_{t+1}^i = m \times (y_t^i - \bar{y}_t^i) + \bar{y}_t^i \quad (12)$$

式中, $y_{t+1}^i$ 为第 $i$ 只变色龙第 $t+1$ 次迭代位置, $y_t^i$ 为第 $i$ 只变色龙第 $t$ 次迭代位置, $\bar{y}_t^i$ 为第 $i$ 只变色龙第 $t$ 次迭代中心位置, $m$ 为旋转转矩,表示变色龙的旋转。

4)捕获猎物。当猎物离变色龙较近时,变色龙便利用舌头攻击并捕获猎物。位置更新数学描述为

$$y_{t+1}^{i,j} = y_t^{i,j} + [(v_t^{i,j})^2 - (v_{t-1}^{i,j})^2] / (2a) \quad (13)$$

式中, $v_t^{i,j}$ 为第 $i$ 只变色龙当前速度, $v_{t-1}^{i,j}$ 为第 $i$ 只变色龙上一次迭代速度, $a$ 为加速度,即

$$a = 2590 \times (1 - e^{-\log t}) \quad (14)$$

综上,通过搜索猎物、眼睛旋转发现猎物和捕获猎物三阶段的位置迭代更新可求得最佳位置 $y$ 。在迭代寻优过程中,针对变色龙的位置解,利用目标函数 $F$ 计算其适应度值,变色龙个体根据当前的位置和适应度值,采取一定的策略进行移动,并寻找更优的位置。根据新位置再次计算更新后的适应度值,进而更新全局最优解。如果达到最大迭代次数或满足目标函数 $F$ 的要求,则优化结束,输出最优解 $y$ ,否则继续进行迭代更新寻找最优解,进而得到优化因子。伪代码如表1所示,原始图像优化前后效果如图3所示。

从图3红色框标注区域可以看出,相较于原始图像,经过特征优化模块后,可见光图像的纹理细节和目标轮廓更清晰,红外图像的对比度和清晰度也更加明显,为后续图像融合奠定了良好基础。

表1 基于CSA的目标函数自适应优化  
Table 1 Adaptive optimization of objective function based on CSA

$n$ : number of chameleons in search space
$d$ : spatial dimension
$t$ : number of iterations, $T$ : max-number of iterations
Build adaptive optimization image: $I_o = yI_B + I_D$
Design objective function for optimizing parameter $y$
$F = \min(L_{AG} + L_{SD} + \lambda L_{CON})$
Initialize a population of $n$ chameleons within the search space
$y^i = l_j + r \times (u_j - l_j)$
Evaluate the fitness of each chameleon using the objective function $F$
While ( $t < T$ ) do
Step1: Search for prey
for $i=1$ to $n$ do
for $j=1$ to $d$ do
if $r_i \geq P_p$ then
$y_i^{i,j} + p_1(p_i^{i,j} - G_i^j)r_2 + p_2(G_i^j - y_i^{i,j})r_1$
else
$y_i^{i,j} + \mu \cdot l_j^i \text{sgn}(\text{rand} - 0.5) + \mu [(u_j - l_j)r_3] \text{sgn}(\text{rand} - 0.5)$
end if
end for
end for
Step2: Eyes' rotation reveals prey
for $i=1$ to $n$ do
$y_{i+1}^i = m \times (y_i^i - \bar{y}_i^i) + \bar{y}_i^i$
end for
Step3: Hunting process for prey
for $i=1$ to $n$ do
for $j=1$ to $d$ do
$y_{i+1}^{i,j} = y_i^{i,j} + [(v_i^{i,j})^2 - (v_{i-1}^{i,j})^2] / (2a)$
end for
end for
Evaluate the new positions of the chameleons
Update the position of the chameleons
Evaluate the fitness of each chameleon
Return best solution $y$
$t = t + 1$
end while
Obtain the global optimal solution $y$ , and then obtain the optimized image $I_o$ .

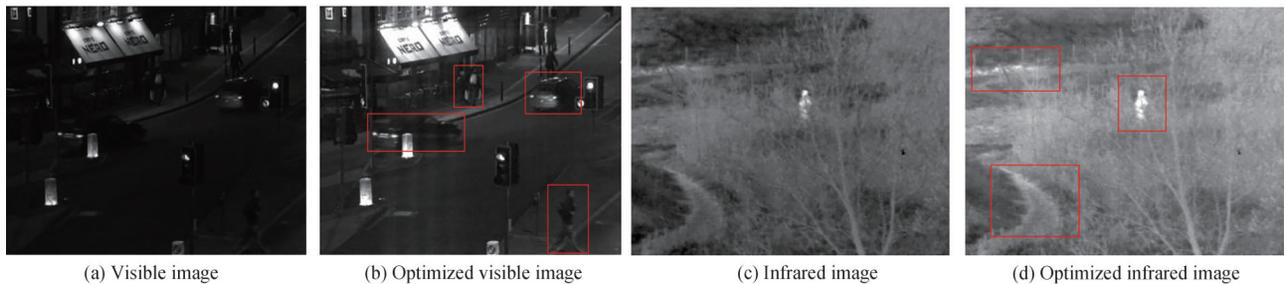


图3 优化前后对比

Fig.3 Comparison chart before and after optimization

## 2.2 融合注意力模型的GAN网络

### 2.2.1 生成器

生成器网络主要由特征提取、特征融合以及特征重构三部分构成。

## 2.2.1.1 基于多尺度密集连接的特征提取模块

红外和可见光图像具有不同的成像机理,其特征表现形式往往有所差异。为此,设计了双支路特征提取网络,分别提取可见光图像的梯度信息和红外图像的对比度信息,如图1所示。两条支路结构相同、参数独立,能够有效降低模型复杂度。同时,为解决单一尺寸卷积核特征提取不全面的问题,构造了一种基于多尺度密集连接的特征提取模块(MSDC-Fem),如图4所示。

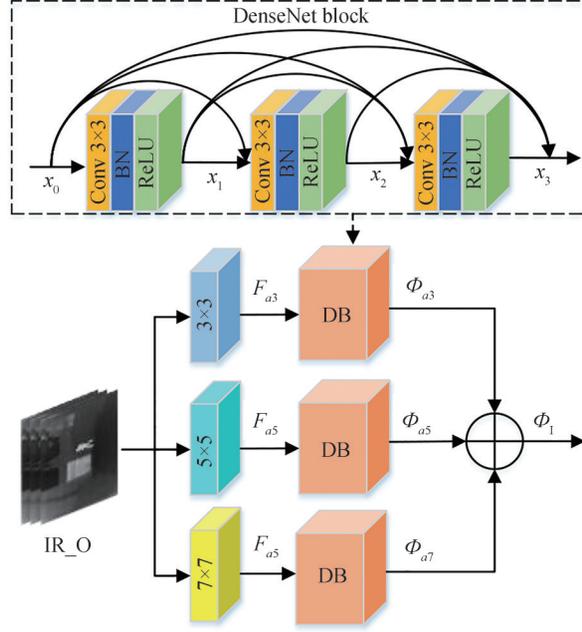


图4 MSDC-Fem 结构

Fig.4 MSDC-Fem structure diagram

以经过特征优化后的红外图像 IR\_O 为例,首先,分别使用三个卷积核大小为  $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$  的卷积层对输入图像进行多尺度特征提取。然后,为避免特征提取时中间层信息丢失问题,在每个不同尺度的卷积分支后分别接入密集连接网络(Densely Connected Convolutional Network, DenseNet)<sup>[22]</sup>进行更深层次特征提取,增强特征传递及重用,使生成器能够更有效地利用原始图像特征信息,同时还可以减轻网络梯度消失。该密集连接网络模块由3个密集连接的卷积层组成,各卷积层的卷积核大小均为  $3 \times 3$ 。最后,将提取的多尺度深层特征进行融合得到局部融合特征。因为在特征计算时所采用的补0策略都是“SAME”,所以不同尺度卷积得到的特征图尺度一样,可以直接进行相加。图4中,利用MSDC-Fem对输入图像进行特征提取时,可以表示为

$$F_{a3} = F_{in} * f_{3 \times 3} \quad (15)$$

$$F_{a5} = F_{in} * f_{5 \times 5} \quad (16)$$

$$F_{a7} = F_{in} * f_{7 \times 7} \quad (17)$$

式中,  $F_{a3}$ 、 $F_{a5}$ 、 $F_{a7}$  分别表示不同尺度的特征图,  $F_{in}$  为输入图像, \* 表示卷积操作,  $f_{3 \times 3}$ 、 $f_{5 \times 5}$ 、 $f_{7 \times 7}$  分别表示3个不同尺寸的卷积核。

$x_0$ 、 $x_1$  和  $x_2$  分别为三个卷积层的输入,  $x_3$  为密集块的末端输出,以  $3 \times 3$  卷积支路为例,  $x_0 \sim x_3$  分别表示为

$$x_0 = F_{a3} \quad (18)$$

$$x_1 = x_0 * f_{3 \times 3} \quad (19)$$

$$x_2 = (x_0 \oplus x_1) * f_{3 \times 3} \quad (20)$$

$$x_3 = (x_0 \oplus x_1 \oplus x_2) * f_{3 \times 3} \quad (21)$$

$$\Phi_{a3} = x_3 \quad (22)$$

式中,  $\oplus$  表示通道连接,  $\Phi_{a3}$  为  $3 \times 3$  支路得到的深层特征。同上,可分别得到另外两条支路的深层特征  $\Phi_{a5}$  和  $\Phi_{a7}$ 。最终,红外图像局部融合特征可表示为

$$\Phi_1 = \Phi_{a3} + \Phi_{a5} + \Phi_{a7} \quad (23)$$

同理,经过MSDC-Fem模块可获得可见光图像局部融合特征 $\Phi_V$ 。

### 2.2.1.2 基于双通道注意力特征融合网络

在融合层设计双通道注意力模型,以使网络从空间和通道两个方向同时聚焦重要目标特征信息,且通过并联连接,网络可以将不同层次的特征信息进行整合,更好地强化重要的空间位置信息和通道特征信息,从而提高网络对关键信息的表达能力。同时,相比于串联方式,并联的双通道注意力机制可以在不同的输入之间进行自由的信息交互,能够更好地捕捉输入特征之间的相关性和依赖关系。具体过程为:将可见光和红外图像对应的局部融合特征 $\Phi_V$ 和 $\Phi_I$ 经过空间注意力模型,可得到空间区域信息加强后的注意力融合特征图 $\Phi_F^{sa}$ ,经过通道注意力模型,可对局部融合特征的通道信息进行增强,得到通道注意力融合特征图 $\Phi_F^{ca}$ 。最后,采用加权平均融合规则将两种注意力融合特征图进行融合,得到全局融合特征图 $\Phi_F$ ,如图5所示。

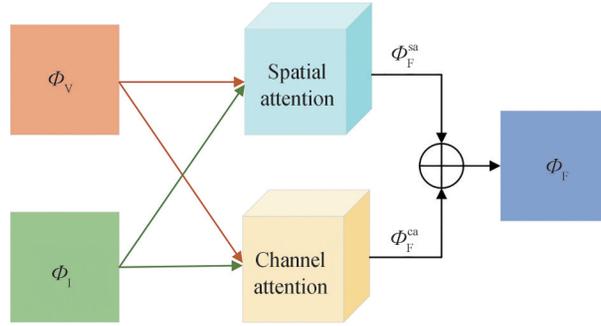


图5 注意力融合过程

Fig.5 Attention fusion process

#### 1)空间注意力模型

对 $\Phi_V \in R^{H \times W \times C}$ 和 $\Phi_I \in R^{H \times W \times C}$ 分别通过计算像素向量沿通道维数方向的 $L_p$ 范数得到其初始注意力图<sup>[23]</sup>,即

$$M_I^{sa}(i, j) = \left\| \Phi_I^{1:C}(i, j) \right\|_p \quad (24)$$

$$M_V^{sa}(i, j) = \left\| \Phi_V^{1:C}(i, j) \right\|_p \quad (25)$$

式中, $M_I^{sa}(i, j)$ 和 $M_V^{sa}(i, j)$ 分别表示红外和可见光图像在 $C$ 维通道中相应像素位置 $(i, j)$ 的初始注意力图, $\|\cdot\|_p$ 表示 $L_p$ 范数。再利用Softmax函数对初始注意力图进行计算,得到红外和可见光图像的空间加权图,即

$$\beta_I^{sa}(i, j) = \frac{\exp(M_I^{sa}(i, j))}{\exp(M_I^{sa}(i, j)) + \exp(M_V^{sa}(i, j))} \quad (26)$$

$$\beta_V^{sa}(i, j) = \frac{\exp(M_V^{sa}(i, j))}{\exp(M_I^{sa}(i, j)) + \exp(M_V^{sa}(i, j))} \quad (27)$$

将空间加权图与局部融合特征图进行相乘,分别得到红外和可见光图像的通道注意力图,即

$$S_I^{sa}(i, j) = \beta_I^{sa}(i, j) \times \Phi_I \quad (28)$$

$$S_V^{sa}(i, j) = \beta_V^{sa}(i, j) \times \Phi_V \quad (29)$$

最后,将两者进行线性叠加得到空间注意力特征图,即

$$\Phi_F^{sa}(i, j) = S_I^{sa}(i, j) + S_V^{sa}(i, j) \quad (30)$$

#### 2)通道注意力模型

首先,计算每个通道上特征映射的 $L_p$ 范数,得到初始通道注意向量,即

$$N_I^{ca}(c) = \left\| \Phi_I(c) \right\|_p \quad (31)$$

$$N_V^{ca}(c) = \left\| \Phi_V(c) \right\|_p \quad (32)$$

式中, $N_I^{ca}(c)$ 和 $N_V^{ca}(c)$ 分别表示红外与可见光图像的初始通道注意向量, $c = 1, 2, \dots, C$ 表示通道索引。然后,其通道加权图可通过式(33)和(34)计算得到。

$$\alpha_I^{ca}(c) = \frac{\exp(N_I^{ca}(c))}{\exp(N_I^{ca}(c)) + \exp(N_V^{ca}(c))} \quad (33)$$

$$\alpha_V^{ca}(c) = \frac{\exp(N_V^{ca}(c))}{\exp(N_I^{ca}(c)) + \exp(N_V^{ca}(c))} \quad (34)$$

将通道加权图与局部融合特征图进行相乘,分别得到对应的红外和可见光图像空间注意力图,即

$$C_I^{ca}(c) = \alpha_I^{ca}(c) \times \Phi_I \quad (35)$$

$$C_V^{ca}(c) = \alpha_V^{ca}(c) \times \Phi_V \quad (36)$$

最后,将两者线性叠加得到通道注意力特征图为

$$\Phi_F^{ca} = C_I^{ca} + C_V^{ca} \quad (37)$$

### 3) 加权融合

空间注意力特征图  $\Phi_F^{sa}$  和通道注意力特征图  $\Phi_F^{ca}$  在空间和通道两个方向获得了更高的重要性,局部特征也建立了全局依赖性且得到细化和增强。最后,使用加权平均融合规则获得最终的全局融合注意力特征图  $\Phi_F$ ,如式(38)所示,更好地保留了红外与可见光图像的显著特征,减小了融合过程中目标信息损失。

$$\Phi_F = \frac{\Phi_F^{sa}}{\Phi_F^{sa} + \Phi_F^{ca}} \cdot \Phi_F^{sa} + \frac{\Phi_F^{ca}}{\Phi_F^{sa} + \Phi_F^{ca}} \cdot \Phi_F^{ca} \quad (38)$$

#### 2.2.1.3 特征重构模块

在特征重构部分,通过4层卷积的解码网络对全局融合注意力特征图  $\Phi_F$  进行重构,进而得到融合图像,如图6所示。在解码网络中,各卷积层均采用(Conv3×3)+BN+ReLU结构,滤波器数分别设置为128、64、32和1。

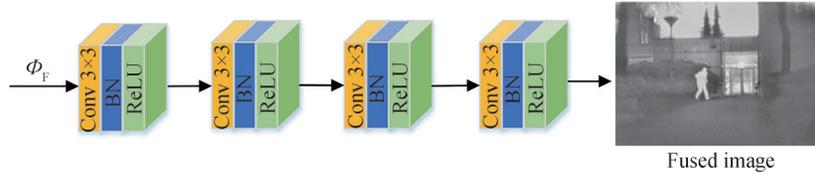


图6 特征重构模块

Fig.6 Feature reconstruction module

#### 2.2.2 双判别器

判别器网络主要用于判断生成器生成的图像是否服从真实样本分布。为了同时保留可见光的纹理细节信息和红外图像的显著目标信息,构造了双判别器结构,如图7所示。Discriminator-VIS用于区分融合图

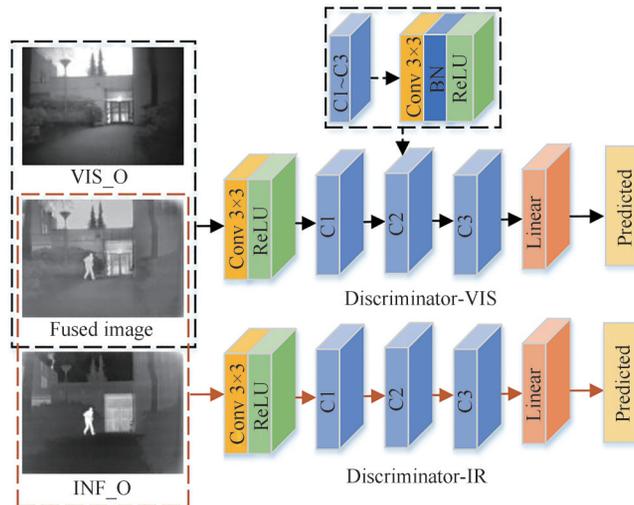


图7 判别器结构

Fig.7 Discriminator structure

像和可见光图像, Discriminator-IR用于区分融合图像和红外图像。考虑到两个判别器之间的平衡,使其具有相同的网络结构,均由4个卷积层和1个线性层组成,4个卷积层均使用 $3 \times 3$ 卷积核和ReLU激活函数,每个卷积层的步长设置为2,滤波器组参数分别为32、64、128和256。线性层将输入与前4个卷积层提取的特征进行区分,并输出分类概率。此外,卷积层C1~C3加入批归一化以缓解网络梯度消失问题。

### 2.3 损失函数设计

损失函数由生成器损失 $L_G$ 和判别器损失 $L_D$ 两部分构成,生成器损失 $L_G$ 包括对抗损失 $L_{adv}$ 和内容损失 $L_{content}$ ,即

$$L_G = L_{adv} + \lambda_1 L_{content} \quad (39)$$

式中, $\lambda_1$ 为权重系数,用于两者之间的平衡。

$$L_{adv} = \frac{1}{N} \sum_{n=1}^N \left\{ [D_{ir}(I_{fused}^n) - a_1]^2 + [D_{vis}(I_{fused}^n) - a_2]^2 \right\} \quad (40)$$

式中, $N$ 表示融合图像的数量, $I_{fused}^n$ 表示生成的融合图像, $D(\bullet)$ 表示判别器分类结果, $a_1$ 和 $a_2$ 分别表示两个判别器的输出概率。对于生成器,不期望判别器能够区分融合图像和真实图像,故取 $a_1 = a_2 = 0.7 \sim 1.2$ 之间随机。

内容损失由对比度损失 $L_{con}$ 和梯度损失 $L_{grad}$ 构成, $L_{con}$ 约束目标物体与背景的对比度, $L_{grad}$ 约束融合图像的纹理细节信息,即

$$L_{content} = \frac{1}{HW} (\xi_1 L_{con} + \xi_2 L_{grad}) \quad (41)$$

式中, $\xi_1$ 和 $\xi_2$ 是平衡两者的正则化参数。 $L_{con}$ 和 $L_{grad}$ 分别表示为

$$L_{con} = \|I_{fused} - I_{ir}\|_F^2 + \|I_{fused} - I_{vis}\|_F^2 \quad (42)$$

$$L_{grad} = \|\nabla I_{fused} - \nabla I_{vis}\|_F^2 + \|\nabla I_{fused} - \nabla I_{ir}\|_F^2 \quad (43)$$

式中, $I_{fused}$ 表示融合图像, $I_{ir}$ 表示原始红外图像, $I_{vis}$ 表示原始可见光图像, $\|\bullet\|_F$ 表示Frobenius范数, $\nabla$ 表示梯度计算。

通过判别器损失函数可以平衡判定融合图像与原始图像的真假性,进而与生成网络模型对抗博弈,使生成的融合图像更趋向于原始图像真实数据分布。判别器损失由红外判别器损失和可见光判别器损失两部分构成,即

$$L_D = L_{D-ir} + L_{D-vis} \quad (44)$$

$$L_{D-ir} = \frac{1}{N} \sum_{n=1}^N [D_{ir}(I_{ir}) - a_{ir}]^2 + \frac{1}{N} \sum_{n=1}^N [D_{ir}(I_{fused}) - d]^2 \quad (45)$$

$$L_{D-vis} = \frac{1}{N} \sum_{n=1}^N [D_{vis}(I_{vis}) - b_{vis}]^2 + \frac{1}{N} \sum_{n=1}^N [D_{vis}(I_{fused}) - d]^2 \quad (46)$$

式中, $a_{ir}$ 为输入的红外优化图像, $b_{vis}$ 为输入的可见光优化图像, $d$ 表示融合图像标签,在 $0 \sim 0.3$ 之间随机选取。

## 3 实验结果与分析

实验所使用的硬件平台配置:CPU为AMD Ryzen 5 5600X 6-Core Processor,主频3.70 GHz;GPU为NVIDIA GeForce RTX 3070 8GB。训练和测试在Windows10系统上,并采用Pytorch框架完成。

### 3.1 数据集与参数设置

采用TNO image fusion公开数据集<sup>[24]</sup>进行实验,从中任选32组已配准的不同场景下的红外和可见光图像,先经过特征优化模块得到优化后的图像,再将此图像作为训练数据集。为了训练一个良好的模型,增强模型的鲁棒性,需对数据集进行扩充。采用滑窗方式对32组优化后的图像进行裁剪,裁剪步长设为12,裁剪图像块尺寸为 $120 \times 120$ ,获取24 200组红外与可见光图像对,并将其灰度值范围转换为 $[0, 1]$ 。

生成器和判别器采用迭代训练方式,两者训练次数之比为 $p$ 。同时,batchsize大小为 $b$ ,总的训练epochs为 $M$ 。实验中,取 $p=1/2$ 、 $b=16$ 、 $M=300$ 。采用Adam优化器对网络模型参数进行优化,学习率设为 $1 \times 10^{-4}$ 。损失函数参数设置为 $\lambda_1 = 0.1$ 、 $\xi_1 = 2$ 、 $\xi_2 = 5$ ,具体训练过程见表2。

表2 网络模型的训练过程  
Table 2 Training process of network model

step1:	for $M$ epochs do
step2:	for $p$ times do
step3:	select $b$ visible image samples: $\{I_{vis}^1, I_{vis}^2, \dots, I_{vis}^b\}$ ;
step4:	select $b$ infrared image samples: $\{I_{ir}^1, I_{ir}^2, \dots, I_{ir}^b\}$ ;
step5:	select $b$ fusion image samples: $\{I_{fused}^1, I_{fused}^2, \dots, I_{fused}^b\}$ ;
step6:	Using the Adam optimizer to update discriminator parameters: $\nabla_D(L_D = L_{D-ir} + L_{D-vis});$
step7:	end for
step8:	select $b$ visible image samples: $\{I_{vis}^1, I_{vis}^2, \dots, I_{vis}^b\}$ ;
step9:	select $b$ infrared image samples: $\{I_{ir}^1, I_{ir}^2, \dots, I_{ir}^b\}$ ;
step10:	Using the Adam optimizer to update generator parameters: $\nabla_G(L_G = L_{adv} + \lambda_1 L_{content});$
step11:	end for

### 3.2 实验结果分析

#### 3.2.1 主观评价

为验证本文算法优势,从TNO数据集中任选6组红外和可见光图像进行主观评价。将本文算法与DenseFuse<sup>[11]</sup>、FusionGAN<sup>[14]</sup>、ResNet-ZCA<sup>[25]</sup>、MDLatLRR<sup>[26]</sup>、PMGI<sup>[27]</sup>以及RFN-Nest<sup>[28]</sup>进行对比,实验结果如图8所示。为了便于观察和分析,对融合结果局部细节用红框进行标注。



图8 主观实验结果对比

Fig.8 Subjective experimental results comparison

从图8可以看出:DenseFuse算法由于采用卷积网络作为特征提取和重建模块,融合图像中较好地保留了纹理细节,但其对红外图像的特征提取能力有一定的限制,所以红外目标信息存在一定的损失,如第4组图像中人物目标信息不突出;FusionGAN算法一定程度上保留了红外显著目标,但由于其采用单一判别器结构,造成可见光纹理细节信息的部分丢失,如第1组图像中树枝和第6组图像中飞机底支架的纹理细节不丰富、边缘轮廓较为模糊;MDLatLRR算法采用多级分解并使用不同的融合策略,能够较好地保留原始图像的细节信息,图像的轮廓信息也较为清晰,但红外目标不突出,如第3组图像中汽车前端红外信息损失严重;RFN-Nest算法采用两阶段训练,使用自动编码器更侧重于纹理细节的保留,而忽略了红外目标特征的保留,如第2组和第4组图像中人物热目标不突出、边缘较为模糊;PMGI算法基于均方误差构建了像素强度和梯度约束,可以生成较清晰的融合图像,但一些局部纹理细节信息不能有效保存,如第5组图像中路灯边缘较模糊;ResNet-ZCA算法通过计算权值图并与原始图像结合获得融合图像,取得了较好的融合效果,但目标区域没有突出显示,如第5组图像中人物背部红外信息不明显;而本文算法通过构造特征优化模块增强图像特征表达能力,并设计融合注意力模型的GAN网络,使得融合结果纹理细节信息丰富、红外目标突出、目标边缘清晰及视觉效果好,相比于其他对比算法具有明显优势。

### 3.2.2 客观评价

为了客观评价本文算法优势,从TNO数据集中随机选取21组图像进行定量分析。选取的客观评价指标<sup>[29]</sup>包括:信息熵(Entropy, EN)、空间频率(Spatial Frequency, SF)、相关熵(Joint Entropy, JE)、视觉保真度(Visual Information Fidelity, VIF)、结构相似性(Structure Similarity Index Measure, SSIM)和梯度信息指标(Gradient-based fusion performance,  $Q^{AB/F}$ )。上述评价指标的数值越大,图像的融合效果越好,对比实验得到的客观评价指标如图9所示。

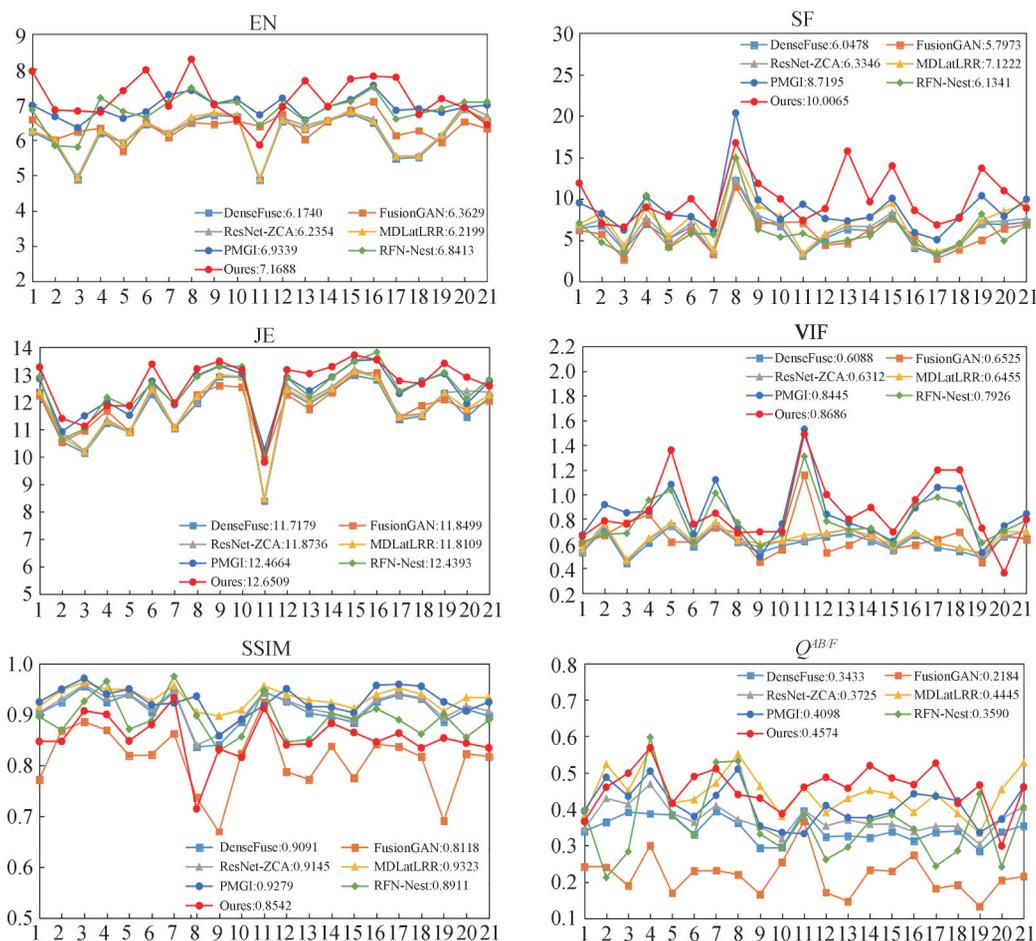


图9 客观实验结果对比

Fig.9 Objective experimental result comparison

从图9可以看出,本文所提算法在EN、SF、JE、VIF和 $Q^{AB/F}$ 5项指标中取得最优平均值,且相比于DenseFuse算法,客观指标分别提高了16.11%、65.46%、7.96%、42.67%和33.24%。SF、JE和VIF较高表明融合图像不仅具有较强的细节表达能力,而且符合人类视觉感知。 $Q^{AB/F}$ 和EN较高表明融合图像信息量更丰富。对于SSIM,其反映了融合图像与原始图像的结构相似性。由于所提算法引入了注意力融合模块,要同时保留红外图像的显著度信息和可见光图像的纹理细节信息,而融合过程通过优化平衡原始特征映射,会导致融合图像中的某些结构和边缘弱化,进而使SSIM相对较低。

### 3.2.3 算法运行时间比较

为了进一步评估算法的复杂性和运行效率,在测试集上对比了各算法的平均运行时间,比较结果如表3所示。实验结果表明,本文算法的平均运行时间低于DenseFuse和RFN-Nest,这是因为本文算法引入了多尺度密集连接模块和注意力融合模块,增加了模型的计算量。

表3 不同算法平均运行时间(单位:秒)  
Table 3 Average running time of different algorithms (units: s)

Algorithms	DenseFuse	FusionGAN	ResNet-ZCA	MDLatLRR	PMGI	RFN-Nest	Ours
Time	0.124	1.846	1.719	5.846	0.637	0.284	0.4519

### 3.2.4 算法收敛性分析

本文算法训练过程中的loss下降曲线如图10所示,可以看出,网络仅训练30轮损失值即下降至0.05,约在100轮之后达到平衡且最终稳定在0.024左右,表明本文算法取得了较好的训练效果,即损失函数收敛快,稳定数值小,网络训练稳定,图像融合精度高。

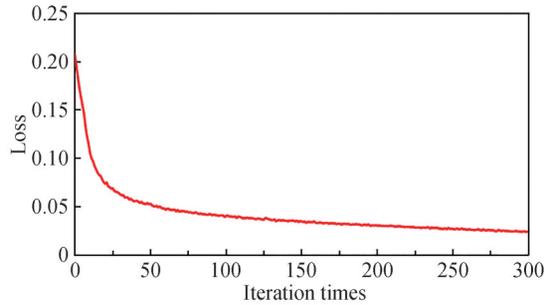


图10 损失函数曲线  
Fig.10 Loss function curve

## 3.3 消融实验

所提算法主要包括特征优化模块、MSDC-Fem模块及注意力融合模块。为验证各模块的有效性,通过加入不同模块对TNO数据集21组图像和6个评价指标进行消融实验,所有实验均使用相同的数据集和参数设置。为了描述方便,将3个模块分别简记为模块A、模块B和模块C。其中,在未使用特征优化模块时,网络的输入是原始可见光和红外图像;在未使用MSDC-Fem模块时,网络改用普通单一尺度卷积提取特征;在未使用注意力融合模块时,网络采用线性叠加的融合规则。21组图像消融实验客观指标平均值的对比结果如表4所示,最优值用粗体标注。

从表4可以看出,相比于原始融合网络(3个模块均不加入),引入特征优化模块,SF和 $Q^{AB/F}$ 分别提升了10.30%和14.29%,表明特征优化模块可使得融合结果在感知上更加清晰,细节更加丰富。引入MSDC-Fem特征提取模块,SF提升了21.59%,表明在特征提取过程中多尺度模块增大了网络感受野特征提取范围,密集连接模块加强了特征的有效传递和重用,进而增强了融合结果的细节和结构。引入注意力融合模块,SF和VIF分别提升了6.68%和3.85%,表明注意力模型使得红外与可见光图像的关键特征得到更好的保留和强化。不同模块协同训练的结果表明:模块A+模块B,EN、SF和SSIM分别提升了7.10%、35.67%和6.18%;模块A+模块C,VIF和 $Q^{AB/F}$ 分别提升了6.71%和22.44%;模块B+模块C,SF、JE和VIF分别提升了45.91%、4.22%和25.49%。本文算法通过融合3个设计模块,在EN、SF、JE、VIF和 $Q^{AB/F}$ 5个指标取得

表4 消融实验客观结果对比  
Table 4 Ablation experiments objectively results comparison

Module A	Module B	Module C	EN	SF	JE	VIF	SSIM	$Q^{AB/F}$
			6.389 5	5.864 2	11.956 1	0.664 2	0.813 4	0.307 0
✓			6.415 8	6.468 5	11.990 7	0.693 4	0.824 6	0.350 9
	✓		6.625 8	7.130 1	12.128 4	0.691 5	0.826 9	0.377 1
		✓	6.489 6	6.256 1	12.034 2	0.689 8	0.819 8	0.315 9
✓	✓		6.842 9	7.956 0	12.362 9	0.715 3	<b>0.863 7</b>	0.388 0
✓		✓	6.559 3	6.594 7	12.186 3	0.708 8	0.820 4	0.375 9
	✓	✓	6.952 4	8.556 3	12.461 2	0.833 5	0.822 5	0.411 8
✓	✓	✓	<b>7.168 8</b>	<b>10.006 5</b>	<b>12.650 9</b>	<b>0.868 6</b>	0.854 2	<b>0.457 4</b>

最优值,且分别提升了12.20%、70.64%、5.81%、30.77%和48.99%,验证了算法各模块的优势。

## 4 结论

本文提出了一种基于特征优化和GAN的红外与可见光图像融合算法。通过设计基于CSA的自适应特征优化模块,解决了原始图像纹理细节不清晰、目标对比度差等问题。在生成器中,构造了多尺度密集连接模块,可以使提取的特征更全面丰富;同时,在特征融合网络中设计并联式双通道注意力模型,有效地避免了原始重要信息的丢失。在判别器中,设计双判别器网络结构,既保留了可见光纹理细节信息又保留了红外热显著度信息。实验结果表明,与其他6种融合方法相比,本文方法不仅具有更好的主观效果,且在客观评价指标EN、SF、JE、VIF和 $Q^{AB/F}$ 中取得最优值。相比于DenseFuse,客观评价指标分别提高了16.11%、65.46%、7.96%、42.67%和33.24%,表明本文方法具有较好的融合效果。

### 参考文献

- [1] ZHANG Xingchen, DEMIRIS Y. Visible and infrared image fusion using deep learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 1: 1-20.
- [2] WANG Zhishe, SHAO Wenyu, YANG Fengbao, et al. Infrared and visible image fusion method via interactive attention-based generative adversarial network[J]. Acta Photonica Sinica, 2022, 51(4): 0410002.  
王志社, 邵文禹, 杨风暴, 等. 红外与可见光图像交互注意力生成对抗融合方法[J]. 光子学报, 2022, 51(4): 0410002.
- [3] HAO Shuai, HE Tian, AN Beiyi, et al. VDFEFuse: a novel fusion approach to infrared and visible images[J]. Infrared Physics & Technology, 2022, 121: 104048.
- [4] XIAO Wanxin, ZHANG Yafei, WANG Hongbin, et al. Heterogeneous knowledge distillation for simultaneous infrared-visible image fusion and super-resolution[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 5004015.
- [5] TU Zhengzheng, LI Zhun, LI Chenglong, et al. Multi-interactive dual-decoder for RGB-thermal salient object detection [J]. IEEE Transactions on Image Processing, 2021, 30: 5678-5691.
- [6] ZHANG Xingchen, YE Ping, LEUNG H, et al. Object fusion tracking based on visible and infrared images: a comprehensive review[J]. Information Fusion, 2020, 63: 166-187.
- [7] NAGARANI N, VENKATAKRISHNAN P, BALAJI N. Unmanned aerial vehicle's runway landing system with efficient target detection by using morphological fusion for military surveillance system[J]. Computer Communications, 2020, 151: 463-472.
- [8] GAO Yuan, MA Shiwei, LIU Jingjing, et al. Fusion of medical images based on salient features extraction by PSO optimized fuzzy logic in NSST domain[J]. Biomedical Signal Processing and Control, 2021, 69: 102852.
- [9] CHEN Jun, LI Xuejiao, LUO Linbo, et al. Infrared and visible image fusion based on target-enhanced multiscale transform decomposition[J]. Information Sciences, 2020, 508: 64-78.
- [10] LU Xiaoqi, ZHANG Baohua, ZHAO Ying, et al. The infrared and visible image fusion algorithm based on target separation and sparse representation[J]. Infrared Physics & Technology, 2014, 67: 397-407.
- [11] LI Hui, WU Xiaojun. DenseFuse: a fusion approach to infrared and visible images[J]. IEEE Transactions on Image Processing, 2018, 28(5): 2614-2623.
- [12] LIU Yu, CHEN Xun, CHENG Juan, 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] JIAN Lihua, YANG Xiaomin, LIU Zheng, et al. SEDRFuse: a symmetric encoder-decoder with residual block network for infrared and visible image fusion[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1-15.

- [14] MA Jiayi, YU Wei, LIANG Pengwei, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. *Information Fusion*, 2019, 48: 11-26.
- [15] GOODFELLOW I, POUGET-ABADIE J, MIRZA M. Generative adversarial networks [J]. *Advances in Neural Information Processing Systems*, 2014, 3: 2672-2680.
- [16] MA Jiayi, XU Han, JIANG Junjun, et al. DDcGAN: a dual-discriminator conditional generative adversarial network for multi-resolution image fusion[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4980-4995.
- [17] BRAIK M. Chameleon swarm algorithm: a bioinspired optimizer for solving engineering design problems [J]. *Expert Systems with Applications*, 2021, 174: 114685.
- [18] LIU Guangcan, LIN Zhouchen, YU Yong. Robust subspace segmentation by low-rank representation[C]. *Proceedings of the 27th International Conference on Machine Learning*, 2010: 663-670.
- [19] LIU Guangcan, YAN Shuicheng. Latent low-rank representation for subspace segmentation and feature extraction[C]. *IEEE International Conference on Computer Vision*, Piscataway, 2011: 1615-1622.
- [20] CUI Guangmang, FENG Huajun, XU Zhihai, et al. Detail preserved fusion of visible and infrared images using regional saliency extraction and multi-scale image decomposition[J]. *Optics Communications*, 2015, 341: 199-209.
- [21] MA Jiayi, MA Yong, LI Chang. Infrared and visible image fusion methods and applications: a survey[J]. *Information Fusion*, 2019, 45: 153-178.
- [22] HUANG Gao, LIU Zhuang, VAN D, et al. Densely connected convolutional networks[C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 4700-4708.
- [23] WANG Zhishe, WANG Junyao, WU Yuanyuan, et al. UNFusion: a unified multi-scale densely connected network for infrared and visible image fusion[J]. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, 32: 3360-3374.
- [24] TNO image fusion data set[EB/OL].<https://flgshare.com/articles/TNO Image Fusion Dataset/1008029>.
- [25] LI Hui, WU Xiaojun, DURRANI T. Infrared and visible image fusion with ResNet and zerophase component analysis[J]. *Infrared Physics & Technology*, 2019, 102: 103039.
- [26] LI Hui, WU Xiaojun, KITTLER J. MDLatLRR: a novel decomposition method for infrared and visible image fusion [J]. *IEEE Transactions on Image Processing*, 2020, 29: 4733-4746.
- [27] ZHANG Hao, XU Han, XIAO Yang, et al. Rethinking the image fusion: a fast unified image fusion network based on proportional maintenance of gradient and intensity [C]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34: 12797-12804.
- [28] LI Hui, WU Xiaojun, KITTLER J. RFN-Nest: an end-to-end residual fusion network for infrared and visible images[J]. *Information Fusion*, 2021, 73: 72-86.
- [29] ZHANG Xiaoli, LI Xiongfei, LI Jun. Validation and correlation analysis of metrics for evaluating performance of image fusion[J]. *Acta Automatica Sinica*, 2014, 40(2): 306-315.
- 张小利, 李雄飞, 李军. 融合图像质量评价指标的相关性分析及性能评估[J]. *自动化学报*, 2014, 40(2): 306-315.

## Infrared and Visible Image Fusion Algorithm Based on Feature Optimization and GAN

HAO Shuai, LI Jiahao, MA Xu, HE Tian, SUN Siyan, LI Tong

(College of Electrical and Control Engineering, Xi'an University of Science and Technology, Xi'an 710054, China)

**Abstract:** Aiming at the problems of insufficient texture details, low contrast and loss of target information in the infrared and visible fusion image, an image fusion algorithm based on feature optimization and a generative adversarial network is proposed. Firstly, considering the impact of original image quality on the fusion results, an adaptive feature optimization module, which is guided by an objective function and based on the chameleon swarm algorithm, is designed, which can enhance the texture details of visible image and the contrast of infrared image. Then, in order to preserve more multi-modal information in the fused image, the generative adversarial network is introduced into the fusion framework. In the generator model, considering the difference in imaging mechanisms between infrared and visible images, a dual branch feature extraction network is constructed. At the same time, in order to solve the problem of insufficient feature extraction caused by single scale convolutional layer, a multi-scale dense connection module is designed to increase the range of network receptive field feature extraction, so as to comprehensively

extract the deep semantic features and shallow texture features of images. To reduce the loss of important target feature information during the fusion process, a parallel attention model based on space and channel is designed in the feature fusion layer. By feeding the infrared and visible image feature information into the spatial and channel attention models, the correlation and dependency between different modal features are captured, and the expression ability of the network is improved, so as to better focus the thermal salient targets of infrared images and the texture details of visible images. In the discriminator model, to preserve the feature information of both infrared and visible images in the fusion results, a dual discriminator network structure for infrared and visible images is constructed, which can enable the fused image to retain as much rich information as possible from the original image in adversarial learning. Moreover, in the network training stage, in order to train a good model and enhance its robustness, 32 sets of images selected from the TNO image fusion data set are cropped and expanded to obtain 24 200 sets of infrared and visible image pairs. Finally, to verify the advantages of the proposed algorithm, subjective and objective results are compared with six classic fusion algorithms, such as DenseFuse, FusionGAN, ResNet-ZCA, MDLatLRR, PMGI and RFN-Nest. The experimental results show that the proposed algorithm has significant advantages in both subjective and objective evaluations. The fused images have richer texture details, clearer edges and targets, and better contrast. In the objective evaluation indexes, the entropy, spatial frequency, joint entropy, visual information fidelity, and gradient-based fusion performance obtain the optimal values, which are improved by 16.11%, 65.46%, 7.96%, 42.67%, and 33.24%, respectively, compared with the DenseFuse algorithm. In addition, to verify the effectiveness of the feature optimization module, multi-scale dense connection module and attention fusion module, ablation experiments are conducted on 21 sets of images and six evaluation indexes, all of which use the same parameter settings. The results of the ablation experiment indicate that compared with the original fusion network (i.e. none of the three modules is added), the six objective evaluation indexes are increased after the introduction of the three designed modules, which verifies the effectiveness of each module of the proposed algorithm.

**Key words:** Image fusion; Feature optimization; Generative adversarial network (GAN); Multi-scale dense connections; Attention model

**OCIS Codes:** 100.2000; 100.2960; 100.2980; 100.3020; 110.3000