多尺度和卷积注意力相结合的红外与可见光图像融合

祁艳杰，侯钦河
（太原科技大学 电子信息工程学院，山西 太原 030024）

摘要：针对红外与可见光图像融合时，单一尺度特征提取不足、红外目标与可见光纹理细节丢失等问题，提出一种多尺度和卷积注意力相结合的红外与可见光图像融合算法。首先，设计多尺度特征提取模块和可变形卷积注意力模块相结合的编码器网络，多感受野提取红外与可见光图像的重要特征信息。然后，采用基于空间和通道双注意力机制的融合策略，进一步融合红外和可见光图像的典型特征。最后，由 3 层卷积层构成解码器网络，用于重构融合图像。此外，设计基于均方误差、多尺度结构相似度和色彩的混合损失函数约束网络训练，进一步提高融合图像与源图像的相似性。本算法在公开数据集上与 7 种图像融合算法进行比较，在主观评价和客观评价方面，所提算法相较其它对比算法具有较好的边缘保持性、源图像信息保留度，较高的融合图像质量。

关键词：红外与可见光图像；混合损失函数；多尺度特征提取；注意力机制；图像融合
中图分类号：TP391 **文献标志码**：A **文章编号**：1001-8891(2024)09-1060-10

Infrared and Visible Image Fusion Combining Multi-scale and Convolutional Attention

QI Yanjie，HOU Qinhe
(*School of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan* 030024, *China*)

Abstract: A multiscale and convolutional attention-based infrared and visible image fusion algorithm is proposed to address the issues of insufficient single-scale feature extraction and loss of details, such as infrared targets and visible textures, when fusing infrared and visible images. First, an encoder network, combining a multiscale feature extraction module and deformable convolutional attention module, is designed to extract important feature information of infrared and visible images from multiple receptive fields. Subsequently, a fusion strategy based on spatial and channel dual-attention mechanisms is adopted to further fuse the typical features of infrared and visible images. Finally, a decoder network composed of three convolutional layers is used to reconstruct the fused image. Additionally, hybrid loss function constraint network training based on mean squared error, multiscale structure similarity, and color is designed to further improve the similarity between the fused and source images. The results of the experiment are compared with seven image-fusion algorithms using a public dataset. In terms of subjective and objective evaluations, the proposed algorithm exhibits better edge preservation, source image information retention, and higher fusion image quality than other algorithms.

Keywords: infrared and visible images, hybrid loss function, multi-scale feature extraction, attention mechanism, image fusion

0 引言

图像融合是一种图像增强技术，它的目标是将多个传感器采集到的有效信息结合到一起，得到一幅信息较全面的图像，以供后续处理或辅助决策。红外与可见光融合是近几年较为热门的一种图像融合技术。其中红外成像传感器能根据热辐射的不同，可将目标与背景区域区分开，具有全天时全天候工作的能力，即使在雨雪等恶劣条件下仍具有良好的目标检测识别能力，但图像分辨率低、对比度差、边缘模糊；可

收稿日期：2023-06-21；**修订日期**：2023-08-22.
作者简介：祁艳杰（1979-），女，蒙古族，博士，副教授，研究方向为信号与信息处理，图像处理。E-mail：qiyanjie1979@163.com。
通信作者：侯钦河（1998-），男，硕士研究生，研究方向为图像融合、深度学习。E-mail：776094677@qq.com。
基金项目：山西省基础研究计划项目（202203021221144）。

见光图像可以提供与人眼视觉相似的高分辨率，能获取场景、纹理等信息，但容易受外界光照、天气等因素的影响。因此，将可见光与红外图像的信息互补融合在一起，可生成目标显著、纹理细节丰富的高质量图像，广泛应用于军事侦察、实时监控、汽车自动驾驶等领域[1-2]。

早期研究人员一般采用基于稀疏表示（Sparse Representation，SR）[3]、低秩表示（Low Rank Representation，LRR）[4-5]、多尺度变换等传统算法实现红外与可见光图像的融合。基于 SR 和 LRR 的融合方法[6-9]中，利用滑动窗把原始影像分割成影像块，再把影像块构建成矩阵，该矩阵被反馈送到 SR（或 LRR）中计算 SR（或 LRR）系数，利用这些系数表征图像特征。通过该运算，将图像融合问题转化为系数融合问题。融合系数由适当的融合策略生成，然后在 SR（或 LRR）框架中重构融合图像。多尺度变换方法[10-14]首先对源图像进行多尺度分解，然后设计相应的融合规则对不同尺度的图像进行融合，最后进行多尺度逆变换重构融合图像。这些图像融合算法的融合性能高度依赖于所使用的特征提取方法，且需人工设计融合规则，计算复杂度高，缺乏通用性。

近年来，由于卷积运算强大的特征提取能力，基于深度学习的图像融合算法在图像融合领域得到了飞速发展。2018 年，Liu 等人[15]提出一种基于卷积神经网络的多聚焦图像融合方法，打破了传统融合算法手动设计图像活动水平测量的约束，但该算法网络层数较少，特征提取能力不足，融合图像存在信息缺失。2019 年，Ma 等人[16]提出 FusionGAN，将生成对抗网络引入图像融合领域，但该算法在对抗训练时，判别器仅以可见光图像作为参照，使得融合图像对比度强但细节纹理不明显。2020 年，Ma 等人[17]又提出一种 GANMcC 算法，利用多分类约束生成对抗网络进一步将图像融合问题转化为多分类限定问题，但该算法缺少对源图像非典型特征的抑制。Prabhakar 等人[18]提出了一种无监督的深度学习框架 DeepFuse，实现多曝光图像的融合，其自编码网络思想被很多研究者采纳，但其网络结构简单，图像深度特征提取不充分。Li 等人[19]将密集连接模块引入编码器结构中，以获取图像深层特征，但该算法网络结构简单，不能提取图像多尺度特征，融合图像对比度不足。Zhang 等人[20]提出了一种基于卷积神经网络的融合框架，这是一种简单而有效的图像融合架构，但其仅用单一尺度对图像进行特征提取，导致部分特征缺失。

针对上述问题，本文提出一种多尺度和卷积注意力相结合的红外与可将光图像融合算法。首先，编码器采用多尺度卷积操作提取红外和可见光图像不同感受野的特征信息，以克服单一尺度卷积核特征提取不足的问题，同时为了获取全局关联信息，引入改进的可变形卷积注意力模块（Deformable-Convolutional Block Attention Module, D-CBAM）[21]，把网络生成的特征图和通过空间注意力和通道注意力得到的注意力特征图进行加权，增强网络对红外和可见光图像重要特征的表达能力。其次，将编码器提取到的红外和可见光的图像特征输入融合层，融合策略中引入空间注意力和通道注意力机制，以融合红外与可见光的典型目标和纹理细节等重要特征。最后，构建三层卷积块组成的解码器，对融合后的特征进行重构，得到最终的融合图像。训练阶段舍弃融合层，并利用混合损失函数进行约束，提升模型学习图像均方误差、结构和色彩等图像特征的能力。

1 本文算法

1.1 网络总体框架

多尺度和卷积注意力相结合的红外与可见光图像融合方法总体框架如图 1 所示。

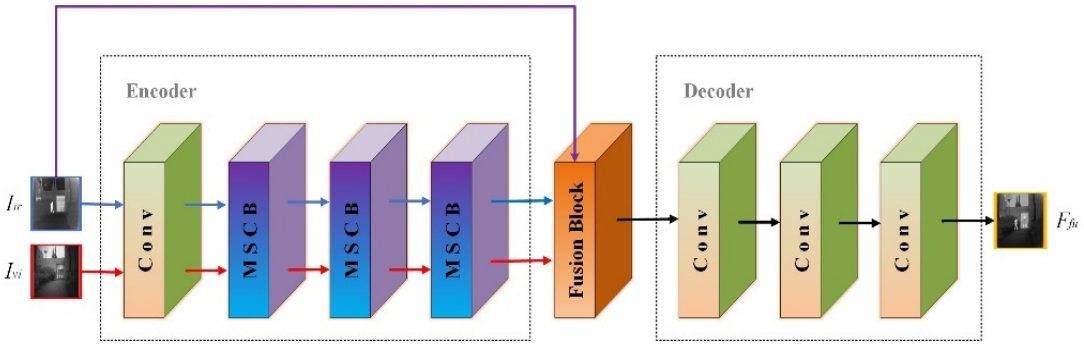


图 1 所提网络结构

Fig.1 Structure of proposed network

整体框架由 3 部分组成：编码器、融合层和解码器。融合时，首先将红外与可见光图像作为源图像输入编码器，通过一层卷积核大小为 3×3 的卷积层和三层多尺度可变形卷积注意力模块（Multi-scale Deformable-Convolutional Block Attention Module, MSCB）组成的编码器提取源图像的多通道显著特征信息；然后，引入基于空间注意力和通道注意力的双重注意力机制融合策略对编码器提取到的特征进行融合；最后，在解码器中对融合后的特征信息进行重构，输出最终的融合图像。

1.2 编码器

本文编码器由一个单一尺度卷积核的卷积块和 3 个 MSCB 组成，每个 MSCB 包含 4 个独立分支和一个 D-CBAM。4 个独立分支采用不同尺度的卷积核，可以提取图像不同感受野的特征信息，丰富图像信息。卷积注意力模块可以捕获红外与可见光图像的全局依赖关系，增强红外与可见光轮廓及纹理细节等信息。

1.2.1 多尺度可变形卷积注意力网络

红外与可见光图像融合旨在将红外目标和可见光的场景纹理信息更好地结合在一起，因此需要提取源图像多尺度的区域特征，以更好地表征红外目标和可见光的纹理细节信息。而在常见的基于卷积神经网络的深度学习方法中，大都采用单一尺度卷积核的卷积块提取图像特征，导致无法对源图像的特征信息进行全面的提取。Szegedy 等人[22]提出深度卷积神经网络 Inception module 模型，该模型通过使用不同大小卷积核的卷积块对源图像不同感受野的特征信息进行提取，从而获得图像不同尺度的特征信息，成功应用于图像分类和图像检测等多种任务中。受其启发，本文提出一种多尺度卷积注意力模块 MSCB，结构如图 2 所示。

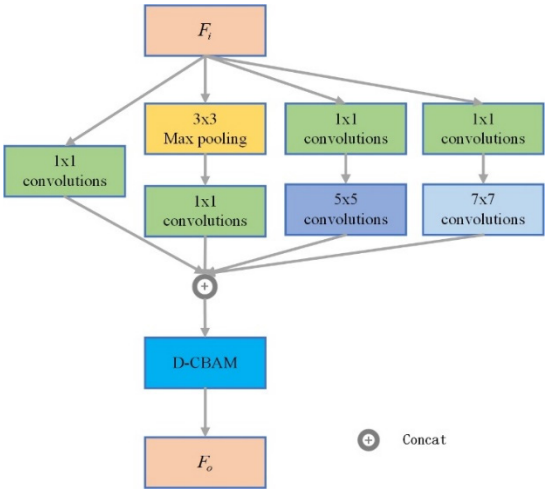


图 2 多尺度卷积注意力模块

Fig.2 Multi-scale convolutional attention block

其中，F_i 表示输入特征，F_o 表示经 MSCB 提取加强后的特征，即 MSCB 的输出。每个 MSCB 包含 4 个独立分支和一个注意力模块，每个分支结构由不同卷积核的卷积层组成，MSCB 模块参数如表 1 所示。

分支一可以减少中间层信息的损失；分支二使网络能够更好地提取源图像的背景信息；分支三与分支四增加网络感受野，提取多个尺度的特征信息，丰富融合图像信息。其中，分支三用两个卷积核大小为 3×3 的卷积层替代卷积核大小为 5×5 的卷积层，分支四用一个卷积核大小为 1×7 的卷积层和一个卷积核大小为 7×1 的卷积层替代卷积核大小为 7×7 的卷积层，每一分支使用卷积核大小为 1×1 的卷积层为该分支降维，以降低模型参数量和计算量、增加网络深度，加快计算速度并增强网络的非线性特性。之后，将 4 个分支的输出进行级联操作，然后将其输入到 D-CBAM 中，对每一通道信息赋值权重，使更具作用的信息被赋予更大权重，大大提升了对图像特征的提取能力，从而提升融合图像的质量。

表 1 MSCB 模块参数设置

Table 1 MSCB module parameter settings

	Kernel size	Outputs channel	Activation function
Branch1	1×1	16	R-Relu
Branch2	3×3Maxpooling	16	R-Relu
	1×1	16	
Branch3	1×1	32	R-Relu
	3×3	64	R-Relu
	3×3	16	R-Relu
Branch4	1×1	64	R-Relu
	1×7	128	R-Relu
	7×1	16	R-Relu

1.2.2 可变形卷积注意力机制

在深度学习构建图像融合的众多方法中，注意力机制是最有效的建模方法之一。目前常用的注意力机制主要有通道注意力、空间注意力、通道与空间注意力等；通道注意力机制旨在显示不同通道之间的相关性，空间注意力机制旨在提升关键区域的特征表达，通道与空间注意力机制结合了通道注意力和空间注意力的形式形成一种更加综合的特征注意力机制，例如卷积注意力模块（Convolutional Block Attention Module, CBAM）[23]。

CBAM 包含通道注意力（Channel Attention Module, CAM）和空间注意力（Spatial Attention Module, SAM）两个子模块，所提算法选用可变形卷积注意力模块（D-CBAM），就是将 CBAM 中的 SAM 子模块中的卷积层换成可变形卷积块，从而增大卷积

块的感受野，使重要信息更易被关注到，并予以更大权重，使编码器提取到更加重要的特征信息，并降低信息冗余。其结构示意图如图 3 所示。

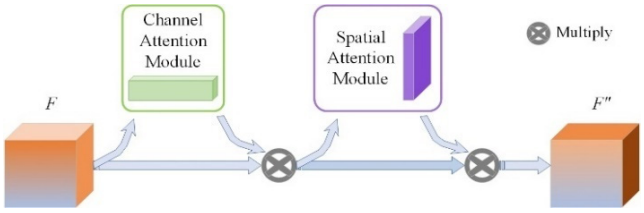


图 3 卷积注意力模块

Fig.3 Convolutional block attention module

1.3 融合层

简单的加权平均融合策略没有对提取的特征图进行筛选，容易引入噪声造成融合图像存在伪影[24]。空间注意力和通道注意力可以同时在空间和通道维度上对深度网络提取的深度特征进行提取，从而增强红外与可见光轮廓和纹理细节等特征信息。因而，本文使用基于空间和通道注意力双重注意力机制的融合策略，融合策略结构如图 4 所示。

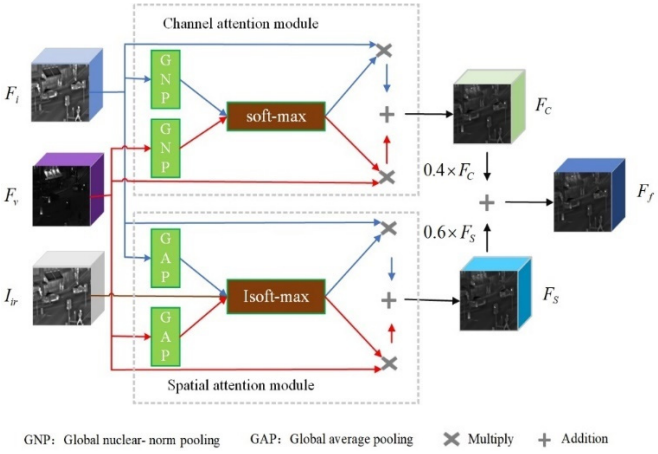


图 4 融合策略结构

Fig.4 The structure of the fusion strategy

图 4 中，F_{v} 和 F_{i} 为由编码器分别从可见光图像和红外图像中提取的多尺度深度特征，F_{S} 和 F_{C} 分别为通过空间注意力模型和通道注意力获得的融合特征，F_{f} 为经融合层融合得到的多尺度深度特征，将其作为解码器网络的输入。其中，由 F_{S} 和 F_{C} 得到 F_{f} 的表达式为：

$$F_{\mathrm{f}}=0.6\times F_{\mathrm{S}}+0.4\times F_{\mathrm{C}} \tag{1}$$

1.3.1 空间注意力模块

空间注意力模型是在图像融合任务中利用基于空间的融合策略，因红外图像中的显著目标亮度较大，为增强融合图像的显著目标对比度，将 soft-max 算子进行改进，输入特征 F_{v} 和 F_{i} 通过全局平均池化层和改进后的 soft-max 算子（Isoft-max）计算获得权重图 α_{v} 和 α_{i}，其计算表达式为：

$$\alpha_{\mathrm{v}}(x,y)=\begin{cases}0, I_{\mathrm{ir}}(x,y)>220\\ \dfrac{\mathrm{e}^{\|F_{\mathrm{v}}(x,y)\|_1}}{\mathrm{e}^{\|F_{\mathrm{i}}(x,y)\|_1}+\mathrm{e}^{\|F_{\mathrm{v}}(x,y)\|_1}}, \quad 25\le I_{\mathrm{ir}}(x,y)\le 220\\ 1, I_{\mathrm{ir}}(x,y)<25\end{cases}$$

$$\alpha_{\mathrm{i}}(x,y)=\begin{cases}1, I_{\mathrm{ir}}(x,y)>220\\ \dfrac{\mathrm{e}^{\|F_{\mathrm{i}}(x,y)\|_1}}{\mathrm{e}^{\|F_{\mathrm{i}}(x,y)\|_1}+\mathrm{e}^{\|F_{\mathrm{v}}(x,y)\|_1}}, \quad 25\le I_{\mathrm{ir}}(x,y)\le 220\\ 0, I_{\mathrm{ir}}(x,y)<25\end{cases} \tag{2}$$

式中：$\|\cdot\|_1$ 表示 L1 范数；(x, y)表示像素对应位置坐标。

然后，将输入特征(F_{v} 和 F_{i})与权重图(α_{v} 和α_{i})做相乘操作得到增强后的可见光图像特征 $\hat{F}_{\mathrm{v}}$ 和红外图像特征 $\hat{F}_{\mathrm{i}}$。最后，将增强后的特征相加得到空间注意力模型增强后的特征 F_{s}，计算表达式为：

$$F_{\mathrm{s}}=\hat{F}_{\mathrm{v}}+\hat{F}_{\mathrm{i}} \tag{3}$$

1.3.2 通道注意力模块

通道注意力模型是在图像融合任务中利用基于信道信息的融合策略，输入特征 F_{v} 和 F_{i} 通过全局池化算子计算获得初始加权向量，这里，全局池化算子选用核范数算子，它是一个通道的奇异值之和，通道所包含重要信息越多奇异值之和越大；最后通过 soft-max 算子计算得到加权向量β_{v} 和β_{i}，计算表达式为：

$$\beta_n(m)=\frac{G\left(F_n(m)\right)}{G\left(F_{\mathrm{i}}(m)\right)+G\left(F_{\mathrm{v}}(m)\right)} \tag{4}$$

式中：$n\in\{\mathrm{v},\mathrm{i}\}$，$m$ 表示输入特征中通道的对应索引；G 表示全局池化算子。

然后，将输入特征 F_{v} 和 F_{i} 与加权向量β_{v} 和β_{i} 做相乘操作得到增强后的可见光图像特征 $\tilde{F}_{\mathrm{v}}$ 和红外图像特征 $\tilde{F}_{\mathrm{i}}$。最后，将增强后的特征相加得到通道注意力模型增强后的特征 F_{C}，计算表达式为：

$$F_{\mathrm{C}}=\tilde{F}_{\mathrm{v}}+\tilde{F}_{\mathrm{i}} \tag{5}$$

1.4 解码器

解码器网络结构由三层卷积核大小为 3×3 卷积块组成，步长均为 1，输出通道分别为 32、16、1，将融合层的输出作为解码器网络的输入，经最后一层卷积重构出灰度融合图像。网络卷积块均舍弃批量归一化层（Batch Normalization），以减少融合图像伪影，提高计算网络计算速率。激活函数均为 R-Relu。

1.5 损失函数

设计了一种训练阶段的损失函数 L，由均方误差

L_{MSE}、多尺度结构相似性度量误差 $L_{MS\text{-}ssim}$ 和色彩感知误差 L_C 共同约束，保证网络进行合理的优化迭代，其表达式为：

$$L=L_{MSE}+\mu L_C+\lambda L_{Ms\text{-}ssim} \tag{6}$$

式中：λ和μ为权重系数。

均方误差是利用融合图像与源图像之间像素差的均方值衡量两幅图像间的差异，计算表达式为：

$$L_{MSE}=\frac{1}{W\times H}\Sigma\left(F_{to}-I_{ti}\right)^2 \tag{7}$$

式中：F_{to} 表示重构图像；I_{ti} 表示输入图像；W 表示图像的宽；H 表示图像的高。

色彩感知误差是通过计算图像的颜色直方图误差来增强融合图像的亮度对比度，从而保证融合图像能够突出可见光图像的纹理以及红外图像的热辐射信息，其计算表达式为：

$$L_C=\frac{1}{255}\|\text{Histogram}(O)-\text{Histogram}(I)\|_2 \tag{8}$$

式中：Histogram(·)表示颜色直方图；$\|\cdot\|_2$ 表示二范数。

多尺度结构相似性度量误差通过亮度因子 $L(x,y)$ 对比度因子 $C(x,y)$和结构因子 $S(x,y)$衡量输入图像与重构图像的相似程度，其计算表达式为：

$$L_{Ms\text{-}ssim}=1-\text{MS\_SSIM}(O,I) \tag{9}$$

$$\begin{cases} L(x,y)=\dfrac{2\mu_x\mu_y+c_1}{\mu_x^2+\mu_y^2+c_1} \\ C(x,y)=\dfrac{2\sigma_x\sigma_y+c_2}{\sigma_x^2+\sigma_y^2+c_2} \\ S(x,y)=\dfrac{\sigma_{xy}+c_3}{\sigma_x\sigma_y+c_3} \\ \text{MS\_SSIM}(x,y)=\left[L_M(x,y)\right]^{\alpha_M}\cdot \prod_{j=1}^{M}\left[C_j(x,y)\right]^{\beta_j}\cdot\left[S_j(x,y)\right]^{\gamma_j} \end{cases} \tag{10}$$

式中：MS_SSIM(x,y)表示两个图像间的多尺度结构相似度；(x,y)表示像素坐标；μ_x 和μ_y 分别表示 x 和 y 的均值；σ_x 和σ_y 分别表示 x 和 y 的标准差，σ_{xy} 表示 x、y 的协方差，$\alpha_j=\beta_j=\gamma_j$，$j=\{1,\cdots,M\}$，$c_1,c_2,c_3$ 是常数，用于保证函数稳定性。

2 实验结果

2.1 实验设置

训练阶段时舍弃融合层，只训练编码器网络和解码器网络，使模型能更精确地重建输入图像，减少重建图像的损失，训练网络结构如图 5 所示。

训练数据集选用 MS-COCO 数据集[25]，选择 80000 张图像转换为灰度图像，并调整为 256 pixel×256 pixel 作为输入图像，网络优化器选用 Adam, epoch=2, batch size=4，学习率为 1×10^{-4}，填充方式为反射填充，填充数 p 的计算表达式如式(11)所示：

$$p=\text{ron}(\frac{\text{kernel}}{2}) \tag{11}$$

式中：ron()为取整函数，kernel 为卷积核大小。损失函数参数λ=700，μ=0.01，硬件配置环境为 NVIDIA GeForce RTX 3090 24GB、12th Gen Intel(R) Core(TM) i7-12700。

为验证所提方法的有效性，选择 7 种近几年提出的经典融合算法进行比较，包括统一无监督网络（U2Fusion）[26]、压缩分解网络（SDNet）[27]、密集连接网络（DenseFuse）[19]、生成对抗网络（FusionGAN）[16]、多分类约束（GANMcC）[17]、通用的有监督图像融合网络（IFCNN）[20]、嵌套连接网络（NestFuse）[28]，并通过主客观评价指标进行评价分析。

2.2 融合图像主观评价

在 TNO [29]和 RoadScene 数据集[30]中分别选取两组（Scene 1～2）和 4 组（Scene 3～6）图像进行实验分析，实验结果如图 6 所示，采用实线框标记背景纹理、虚线标记红外显著目标。

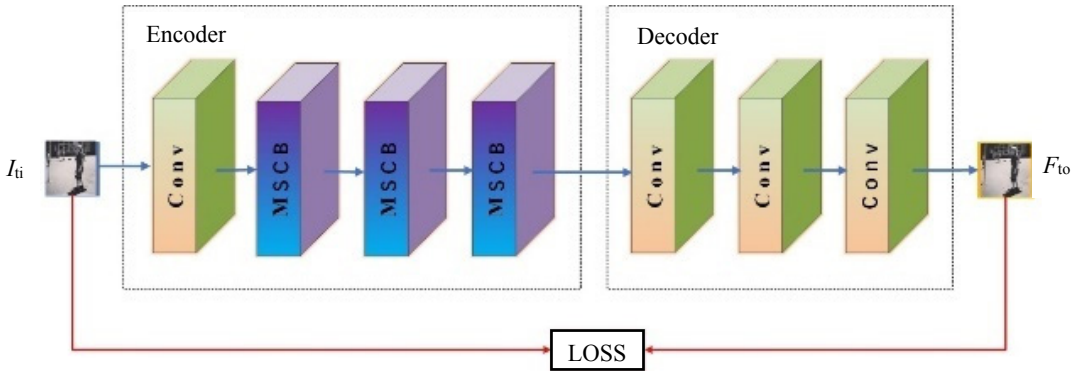


图 5 训练网络结构

Fig.5 The structure of the training network

图 6 部分结果视觉效果对比

Fig.6 Visual comparison of partial results

可以看出，SDNet 算法采用压缩分解网络实现红外与可见光图像融合，一定程度上保留了红外显著目标，但存在伪影，树叶、天空等背景纹理方面表现较差；NestFuse 算法采用嵌套连接网络实现图像融合，对树叶等背景纹理信息处理表现较好，但云层等背景信息对比度差；DenseFuse 算法采用密集连接网络保留红外图像与可见光图像特征，融合图像在保留背景纹理方面表现较好，但未能突出红外目标且树叶等背景纹理信息处理欠佳；FusionGAN 算法采用了生成对抗网络实现红外与可见光图像融合，保留了红外显著目标，但树叶等背景纹理信息严重丢失；GANMcC 算法在突出红外辐射信息方面较好，但融合图像存在少量伪影，细节纹理清晰度较低；IFCNN 算法生成的融合图像未能很好区分红外显著目标与背景纹理，对比度较低；如图 6 中 Scene1 所示，U2Fusion 算法未能突出红外显著目标，对比度较差；所提算法生成的融合图像更能凸显红外显著目标，同时保留背景纹理细节，更符合人类视觉系统特征。

2.3 融合图像客观评价

主观评价存在人为主观因素，具有一定的随机性和片面性。为了更好地分析融合图像的质量，选取 4 种基于融合图像质量的客观评价指标均方误差（mean

squared error，MSE）、信息熵（entropy，EN）、标准差（standard deviation，STD）、空间频率（spatial frequency，SF）和 3 种基于融合图像与源图像的客观评价指标互信息（mutual information，MI）、边缘保持度（$Q^{AB/F}$）、结构相似度（structural similarity，SSIM）对融合图像质量进行对比实验分析。

从 TNO 和 VOT 数据集[31]中共选取 40 对红外与可见光图像，并以 5 对图像为一组，分为 8 组作为对比测试集。表 2 中示出了通过现有融合算法和所提算法获得的所有融合图像的 7 个评价指标得分的平均值。将得出的指标得分按分组取均值后以折线图的形式进行可视化，图 7 为不同算法的客观评价指标折线图。

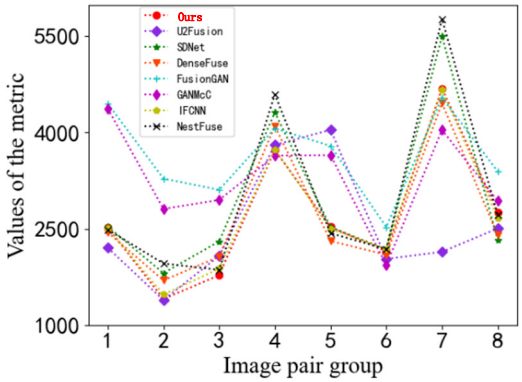
可以看出，所提算法在实验中，7 项指标中有 5 项指标为最优值。SSIM 指标得分与 DenseFuse 得分仅有较小差距；尤其较 GANMcC 算法，$Q^{AB/F}$ 指标提高了约 100%，SF 指标提高了约 77.69%，说明用所提图像融合算法融合的图像中纹理与边缘信息更加清晰丰富。同时，MSE、SF、$Q^{AB/F}$、STD4 项指标较其他 7 种对比算法的平均值提高了 10.3%、23.6%、48.5%、23.5%。说明所提图像融合算法相较于其它对比算法具有较好的边缘保持性、源图像信息保留度、视觉效果及较高的融合图像质量。

表 2 TNO 数据集与 VOT 数据集对比实验客观评价指标均值

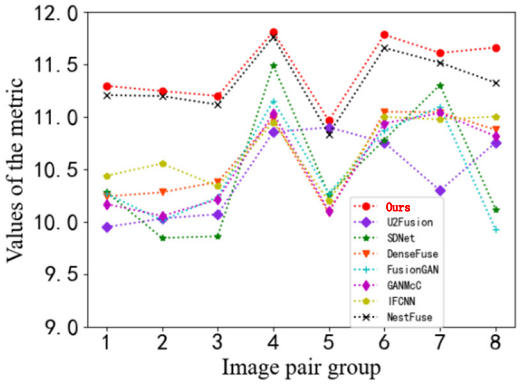
Table 2 Mean values of objective evaluation indicators in comparative experiments between TNO dataset and VOT dataset

Algorithms	MSE	MI	SF	SSIM	$Q^{AB/F}$	STD	EN
U2Fusion	2704	10.45	11.54	0.68	0.45	36.33	6.95
SDNet	2936	10.49	11.82	0.70	0.45	33.14	6.69
DenseFuse	2696	10.61	8.77	**0.72**	0.45	34.83	6.78
NestFuse	2999	11.33	10.02	0.71	0.53	41.67	6.98
IFCNN	2701	10.68	**12.42**	0.71	0.53	35.43	6.74
FusionGAN	3645	10.48	6.08	0.66	0.22	29.61	6.52
GANMcC	3290	10.55	6.14	0.69	0.28	33.33	6.72
Ours	**2657**	**11.67**	10.91	0.71	**0.56**	**42.71**	**7.01**

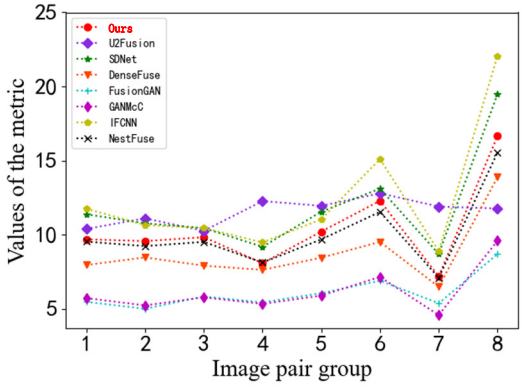
Note: Bold font is the optimal value for each column



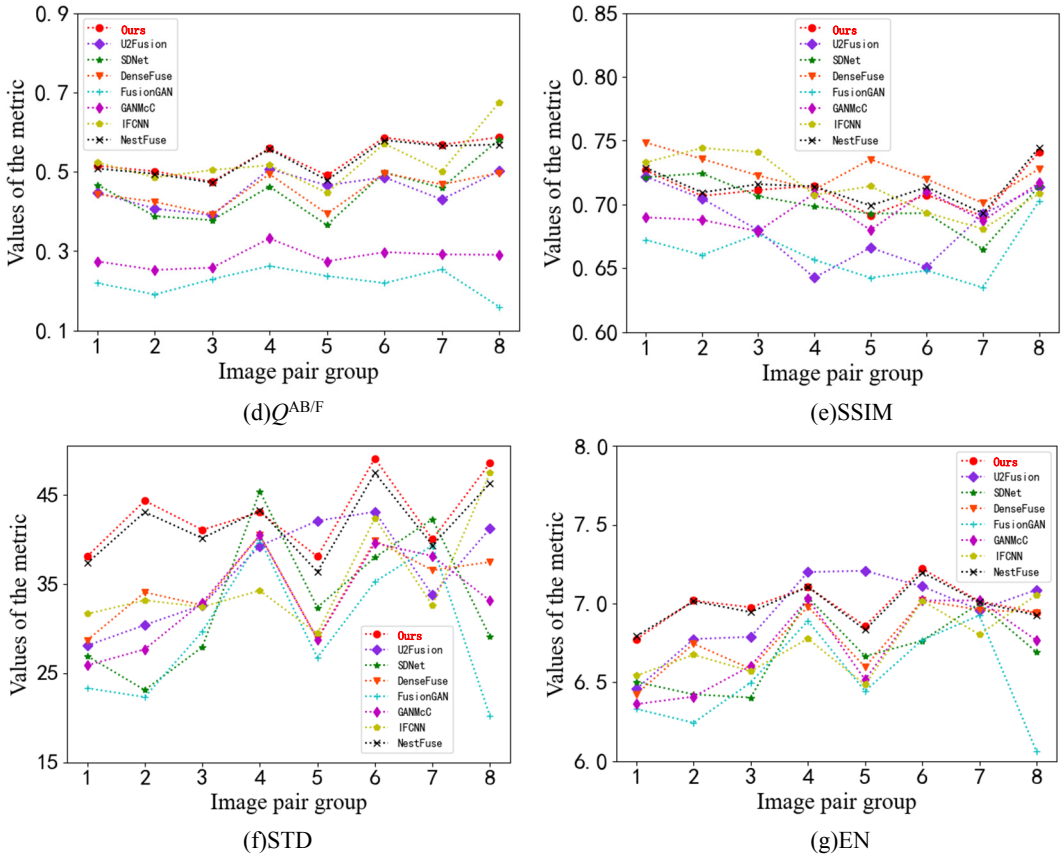
(a)MSE



(b)MI



(c)SF



(d)$Q^{AB/F}$ (e)SSIM

(f)STD (g)EN

图 7 不同算法的客观指标对比折线图

Fig.7 Comparison of objective indicators of different algorithms line chart

为进一步验证所提算法的性能，选择含有 221 对红外与可见光图像对的 RoadScene 数据集与其它 7 种融合算法进行比较，现有融合算法和所提出的融合算法获得的所有融合图像的 7 个评价指标得分的平均值如表 3 所示，其中加粗字体为最优值。从实验各项评价指标得分的均值可以看出，所提算法在 7 项指标中的 5 项指标均为最优值。其中 MSE、SF、$Q^{AB/F}$、STD 指标分别平均提高了 21%、16.6%、28.6%、16.6%；从而进一步表明所提算法相较于其它 7 种对比算法具有更好的图像融合效果。

表 3 RoadScene 数据集对比实验客观评价指标均值

Table 3 Mean of objective evaluation indicators for comparative experiments on the RoadScene dataset

Algorithms	MSE	MI	SF	SSIM	$Q^{AB/F}$	STD	EN
U2Fusion	2273	11.77	15.01	0.68	0.51	42.87	7.26
SDNet	2866	12.10	15.03	0.70	0.51	44.97	7.31
DenseFuse	2919	11.82	12.32	0.69	0.48	42.57	7.22
NestFuse	2319	12.45	13.28	0.67	0.50	49.97	7.38
IFCNN	2328	11.77	**15.07**	**0.70**	0.51	39.18	7.12
FusionGAN	4460	11.65	8.32	0.59	0.26	38.98	7.06
GANMcC	3807	11.80	8.99	0.65	0.35	43.76	7.23
Ours	**2231**	**12.57**	13.90	0.69	**0.54**	**50.03**	**7.40**

Note: Bold font is the optimal value for each column

以上实验结果说明所提算法在红外与可见光图像融合任务中不仅可以保留丰富信息，还有更好的结构和清晰度，且融合图像视觉效果更加自然。

2.4 消融实验

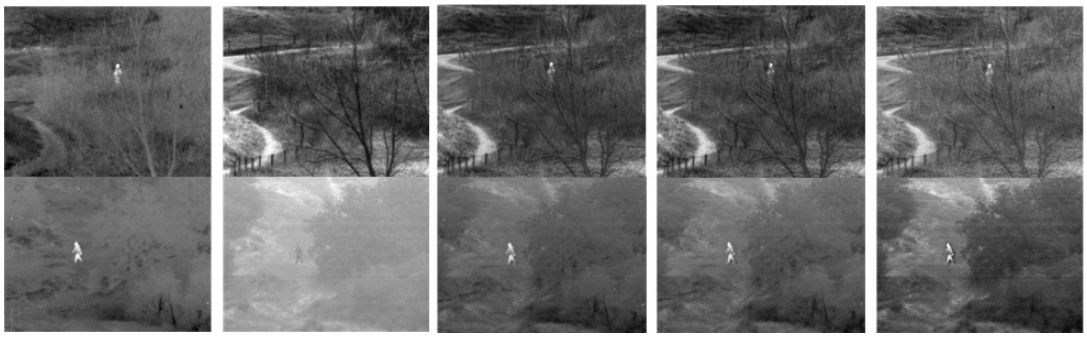
为进一步验证所提算法中提出的各模块的有效性，进行以下消融实验：实验 1 编码器使用 3 个单一尺度卷积块（Conv）；实验 2 编码器使用 3 个单一尺度卷积和卷积注意力机制（Conv+D-CBAM）；实验 3 编码器使用 3 个多尺度卷积注意力模块（MSCB）。

在 TNO 数据集中随机选择一组图像结果作为消融实验结果，实验结果如图 8 所示。

通过观察发现，实验 2 融合图像亮度信息和细节

纹理信息较好，但对比度较差，实验 1 相较于实验 2 对比度有较多提升，但细节纹理信息不足。实验 3 融合结果改善了上述缺点，结合了二者优点，融合图像很好地保留了红外与可见光图像中的特征信息，纹理信息丰富，有较好对比度。融合效果良好。

另外，表 4 是实验融合图像的客观评价指标结果，最优值用加粗标记。所提算法取得了 6 个最优值，在 SF、MI、STD 和 $Q^{AB/F}$ 指标上相较于实验 2 提升较多，这说明融合图像在信息保留度、细节纹理以及视觉方面效果更优。



(a) IR (b) VIS (c) Experiment 1 (d) Experiment 2 (e) Ours

图 8 消融实验结果视觉对比

Fig.8 Visual comparison of ablation experiment results

表 4 消融实验客观指标

Table 4 Objective indicators of ablation experiments

Experiment	MSE	MI	SF	SSIM	$Q^{AB/F}$	STD	EN
Conv	2002	10.58	8.59	0.70	0.45	35.88	6.77
Conv + D-CBAM	**1832**	10.61	8.86	0.69	0.47	36.21	6.80
MSCB	2122	**11.51**	**10.08**	**0.71**	**0.51**	**39.51**	**6.95**

Note: Bold font is the optimal value for each column

3 结论

本文提出了一种多尺度和卷积注意力相结合的红外与可见光图像融合算法。首先，在自编码器网络中采用多尺特征提取模块和卷积注意力机制对源图像特征进行提取，同时提取源图像的浅层细节特征和深层显著特征；其次，采用一种基于两阶段注意力模型的融合策略，融合可见光与红外图像的典型目标特征和纹理细节特征，舍弃无用特征，经编码器网络重构最终的融合图像。在两组对比实验中，相较于其他对比实验，所提算法在 7 种客观评价指标中均有 5 种评价指标取得最佳。其中，在 TNO 与 VOT 数据集对比实验中，在 MSE、SF、$Q^{AB/F}$、STD 指标相较于其他 7 种对比算法均值分别提高了 10.3%、23.6%、48.5%、23.5%；在 RoadScene 数据集对比实验中，MSE、SF、$Q^{AB/F}$、STD 指标相较于其他 7 种对比算法均值分别提高了 21%、16.6%、28.6%、16.6%。通过对比实验结果证明，所提算法在红外与可见光图像融合任务中不仅可以保留丰富的信息，还有更好的结构和边缘信息清晰度，在客观方面和主观方面均取得较好的融合效果。

参考文献：

[1] 代立杨, 刘刚, 肖刚. 基于 FRC 框架的红外与可见光图像融合方法[J]. **控制与决策**, 2021, **36**(11): 2690-2698.
DAI L Y, LIU G, XIAO G. Infrared and visible image fusion based on FRC algorithm[J]. *Control and Decision*, 2021, **36**(11): 2690-2698.

[2] MA J, MA Y, LI C. Infrared and visible image fusion methods and applications: a survey[J]. *Information Fusion*, 2019, **45**: 153-178.

[3] LI X S, WAN W J, ZHOU F Q, et al. Medical image fusion based on sparse representation and neighbor energy activity[J]. *Biomedical Signal Processing and Control*, 2023, **80**(2): 104353.

[4] LIU G, LIN Z, YAN S, et al. Robust recovery of subspace structures by low-rank representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, **35**(1): 171-184.

[5] 孙彬, 诸葛吴为, 高云翔, 等. 基于潜在低秩表示的红外和可见光图像融合[J]. **红外技术**, 2022, **44**(8): 853-862.
SUN Bin, ZHUGE Wuwei, GAO Yunxiang, et al. Infrared and visible image fusion based on latent low-rank representation[J]. *Infrared Technology*, 2022, **44**(8): 853-862.

[6] LI H, WU X J. Multi-focus image fusion using dictionary learning and low-rank representation[C]//*Proceedings of the* 9th *International Conference on Image and Graphics*, 2017: 675-686.

[7] LIU C H, QI Y, DING W R. Infrared and visible image fusion method

based on saliency detection in sparse domain[J]. *Infrared Physics and Technology*, 2017, **83**: 94-102.

[8] GAO R, Vorobyov S A, ZHAO H. Image fusion with cosparse analysis operator[J]. *IEEE Signal Processing Letters*, 2017, **24**(7): 943-947.

[9] LI Y H, LIU G, Bavirisetti D P, et al. Infrared-visible image fusion method based on sparse and prior joint saliency detection and LatLRR-FPDE[J]. *Digital Signal Processing*, 2023, **134**: 103910.

[10] 蒋杰伟, 刘尚辉, 金库, 等. 基于 FCM 与引导滤波的红外与可见光图像融合[J]. **红外技术**, 2023, **45**(3): 249-256.
JIANG Jiewei, LIU Shanghui, JIN Ku, et al. Infrared and visible-light image fusion based on FCM and guided filtering[J]. *Infrared Technology*, 2023, **45**(3): 249-256.

[11] 李文, 叶坤涛, 舒蕾蕾, 等. 基于高斯模糊逻辑和 ADCSCM 的红外与可见光图像融合算法[J]. **红外技术**, 2022, **44**(7): 693-701.
LI W, YE K T, SHU L L, et al. Infrared and visible image fusion algorithm based on Gaussian fuzzy logic and adaptive dual-channel spiking cortical model[J]. *Infrared Technology*, 2022, **44**(7): 693-701.

[12] LI S, KANG X, HU J. Image fusion with guided filtering[J]. *IEEE Transactions on Image Processing*, 2013, **22**(7): 2864-2875.

[13] 霍星, 邹韵, 陈影, 等. 双尺度分解和显著性分析相结合的红外与可见光图像融合[J]. **中国图象图形学报**, 2021, **26**(12): 2813-2825.
HUO X, ZOU Y, CHEN Y, et al. Dual-scale decomposition and saliency analysis based infrared and visible image fusion[J]. *Journal of Image and Graphics*, 2021, **26**(12): 2813-2825.

[14] 刘明葳, 王任华, 李静, 等. 各向异性导向滤波的红外与可见光图像融合[J]. **中国图象图形学报**, 2021, **26**(10): 2421-2432.
LIU M W, WANG R H, LI J, et al. Infrared and visible image fusion with multi-scale anisotropic guided filtering[J]. *Journal of Image and Graphics*, 2021, **26**(10): 2421-2432.

[15] LIU Y, CHEN X, WANG Z, et al. Deep learning for pixel-level image fusion: recent advances and future prospects[J]. *Inf. Fusion*, 2018, **42**: 158-173.

[16] MA J, WEI Y, LIANG P, et al. FusionGAN: a generative adversarial network for infrared and visible image fusion[J]. *Inf. Fusion*, 2019, **48**: 11-26.

[17] MA J, ZHANG H, SHAO Z, et al. GANMcC: a generative adversarial network with multiclassification constraints for infrared and visible image fusion[J]. *IEEE Transactions on Instrumentation and Measurement*, 2021, **70**: 1-14.

[18] Prabhakar K R, Srikar V S, Babu R V. DeepFuse: a deep unsupervised approach for exposure fusion with extreme exposure imagepairs[C]//*IEEE International Conference on Computer Vision* (ICCV), 2017: 4724-4732.

[19] 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.

[20] 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.

[21] 陈永, 张娇娇, 王镇. 多尺度密集连接注意力的红外与可见光图像融合[J]. **光学精密工程**, 2022, **30**(18): 2253-2266.
CHEN Yong, ZHANG Jiaojiao, WANG Zhen. Infrared and visible image fusion based on multi-scale dense attention connection network[J]. *Optics and Precision Engineering*, 2022, **30**(18): 2253-2266.

[22] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, et al. Rethinking the inception architecture for computer vision[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 2016: 2818-2826.

[23] WOO S, PARK J, LEE J, et al. CBAM: Convolutional block attention module[C]//*European Conference on Computer Vision*, 2018, **06521**: 3-19.

[24] 李霖, 王红梅, 李辰凯. 红外与可见光图像深度学习融合方法综述[J]. **红外与激光工程**, 2022, **51**(12): 20220125.
LI L, WANG H M, LI C K. A review of deep learning fusion methods for infrared and visible images[J]. *Infrared and Laser Engineering*, 2022, **51**(12): 20220125.

[25] LIN T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context[C]//*Proceedings of the* 13th *European Conference on Computer Vision*, 2014: 740-755.

[26] XU H, MA J, JIANG J, et al. U2Fusion: a unified unsupervised image fusion network[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, **44**(1): 502-518.

[27] ZHANG H, MA J. SDNet: a versatile squeeze-and-decomposition network for real-time image fusion[J]. *International Journal of Computer Vision*, 2021, **129**: 2761-785.

[28] LI H, WU X J, Durrani T. NestFuse: an infrared and visible image fusion architecture based on nest connection and spatial/channel attention models[J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, **69**(12): 9645-9656.

[29] TOET A. TNO image fusion dataset [EB/OL]. [2021-02-20]. https://figshare.com/articles/TN Image Fusion Dataset/1008029.

[30] XU Han. Roadscene database[DB/OL]. [2020-08-07]. https://github.com/hanna-xu/RoadScene.

[31] Kristan M, Matas J, Leonardis A, et al. The eighth visual object tracking VOT2020 challenge results[C]//*Proceedings of the* 16th *European Conference on Computer Vision*, 2020, **12539**: 547-601.