基于空洞卷积与双注意力机制的红外与可见光图像融合

何 乐[1]，李忠伟[1]，罗 偲[1]，任 鹏[1]，隋 昊[2]
（1. 中国石油大学（华东）海洋与空间信息学院，山东 青岛 266580；
2. 中国石油大学（华东）计算机科学与技术学院，山东 青岛 266580）

摘要：针对红外与可见光图像融合算法中多尺度特征提取方法损失细节信息，且现有的融合策略无法平衡视觉细节特征和红外目标特征，出了基于空洞卷积与双注意力机制（Dilated Convolution and Dual Attention Mechanism, DCDAM）的融合网络。该网络首先通过多尺度编码器从图像中提取原始特征，其中编码器利用空洞卷积来系统地聚合多尺度上下文信息而不通过下采样算子。其次，在融合策略中引入双注意力机制，将获得的原始特征输入到注意力模块进行特征增强，获得注意力特征；原始特征和注意力特征合成最终融合特征，得在不丢失细节信息的情况下捕获典型信息，同时抑制融合过程中的噪声干扰。最后，解码器采用全尺度跳跃连接和密集网络对融合特征进行解码生成融合图像。通过实验表明，DCDAM 比其他同类有代表性的方法在定性和定量指标评价都有提高，体现良好的融合视觉效果。

关键词：图像融合；空洞卷积；多尺度结构；密集网络
中图分类号：TP391 **文献标志码**：A **文章编号**：1001-8891(2023)07-0732-07

Infrared and Visible Image Fusion Based on Dilated Convolution and Dual Attention Mechanism

HE Le[1]，LI Zhongwei[1]，LUO Cai[1]，REN Peng[1]，SUI Hao[2]
(1. *College of Oceanography and Space Informatics, China University of Petroleum (East China), Qingdao* 266580, *China*;
2. *College of Computer Science and Technology, China University of Petroleum (East China), Qingdao* 266580, *China*)

Abstract: The multiscale features extraction method in infrared and visible image fusion algorithms loses detail information. Existing fusion strategies also cannot balance the visual detail and infrared target features. Therefore, a fusion network via a dilated convolution and dual-attention mechanism (DCDAM) is proposed. First, the network extracts the original features from the image through a multiscale encoder. The encoder systematically aggregates the multiscale context information through dilated convolution instead of using downsampling operator. Second, a dual-attention mechanism is introduced into the fusion strategy, and the original features are input into the attention module for feature enhancement to obtain the attention features. The original and attention features were combined into the final fusion feature. The mechanism captured the typical information without losing details and suppressed the noise during the fusion process. Finally, the decoder used a full-scale jump connection and dense network to decode the fusion features and generate the fused image. The experimental results show that the DCDAM is better than other representative methods in qualitative and quantitative index evaluations and has a good visual effect.

Key words: image fusion, dilated convolution, multiscale structure, dense network

0 引言

图像融合是将同一场景的多模态图像中的重要信息集成到单张图像中，以实现最佳信息丰富度。高

收稿日期：2022-06-07；**修订日期**：2022-08-10.
作者简介：何乐（1997-），女，硕士研究生。主要研究方向为图像融合与目标检测。E-mail：hele0128@163.com。
通信作者：罗偲（1983-），男，副教授。主要研究方向为无人系统的仿生设计和控制。E-mail：tsai.lo.95@gmail.com。
基金项目：国家自然科学基金联合基金（U1906217）；国家自然科学基金（62071491）；国家重点研发计划（2021YFE0111600）；中央高校基本科研业务费专项资金资助（22CX01004A-1）。

分辨率、细节丰富的可见光图像有利于视觉观察，但当光照不足或物体被阴影、烟雾等遮挡时，会丢失重要的目标信息；而红外图像可以突出比背景温度更高或低的目标而不受外在条件约束[1-2]。因此，红外与可见光图像融合可以全面恢复场景信息。

目前图像融合算法可分为传统方法和深度学习方法。传统方法的分解和融合过程需要人工设计和大量计算，这限制了它在实时检测任务中的应用。因此，深度学习因其能保留高级语义信息和强大的自主学习能力而被广泛应用于图像融合。深度学习方法可分为卷积神经网络（Convolutional Neural Network, CNN）、生成对抗网络（Generative Adversarial Networks, GAN）和自动编码/解码器。而 CNN 网络模型结构简单，对学习较复杂融合模型时效果不佳；GAN 模型生成图像不稳定，容易造成融合图像整体亮度降低，且在融合过程中易引入噪声；而自动编码/解码器架构在没有监督学习的情况下具有良好的重构特性。2018 年 Li[3]等提出了一种端到端模型，将网络分为编码器、解码器和融合层，编码器中引入 Densenet 网络提取图像的深层特征，并在训练阶段丢弃融合层以获得更加灵活的网络。在此基础上，Jian[4]等在融合阶段引入了残差模块，通过元素选择获得的补偿特征被传递到相应的卷积层去重建图像。但是这种方法并未充分提取图像的多尺度特征。在图像处理中，不同尺度的特征映射得到不同的信息，底层特征图具有详细的空间信息和图像边缘信息；高级特征映射更多地是关于图像的位置信息。因此，采用多尺度特征提取会使图像包含信息更加丰富。2020 年，Li[5]等将 Unet＋＋结构用于图像融合，提出 NestFuse，在每层编码器与解码器间形成一个多尺度嵌套连接；为了减少层级之间的语义鸿沟，通过上采样与跳跃连接，引入更多参数将中间层的特征信息利用，最后重建图像。但是这种网络模型都使用下采样算子进行多尺度特征提取。在每次下采样操作中，详细信息逐渐被稀释；同时，多尺度结构在解码中没有得到充分利用。

为了进一步满足多尺度信息融合算法的要求，本文设计了一种新的图像融合模型 DCDAM。首先提出了一种新的多尺度特征提取网络，它在不改变图像分辨率的情况下增加感受野，避免由于多次下采样操作而丢失图像细节信息，从而最大限度地保留原始图像信息。同时，在特征融合中引入了双注意力机制模块进行特征加强。将原始特征与注意力特征相加后得到最终的融合特征，以平衡红外目标与可见光细节信息。最后，在特征重建时提出了一种密集连接解码网络，该网络通过全尺度密集网络连接，充分利用多尺度特征，对提取的特征最大程度重建。通过实验表明，DCDAM 比其他同类有代表性的方法在定性和定量指标评价都有提高，体现良好的融合视觉效果。

1 所提算法模型

本文算法模型主要包含特征提取编码器模块与图像融合模块，其中图像融合模块包括双注意力机制特征融合与全尺度密集连接解码器。I_{vi} 和 I_{ir} 分别表示输入可见光图像与红外图像，O_f 表示输出融合图像，网络框架图如 1 所示。

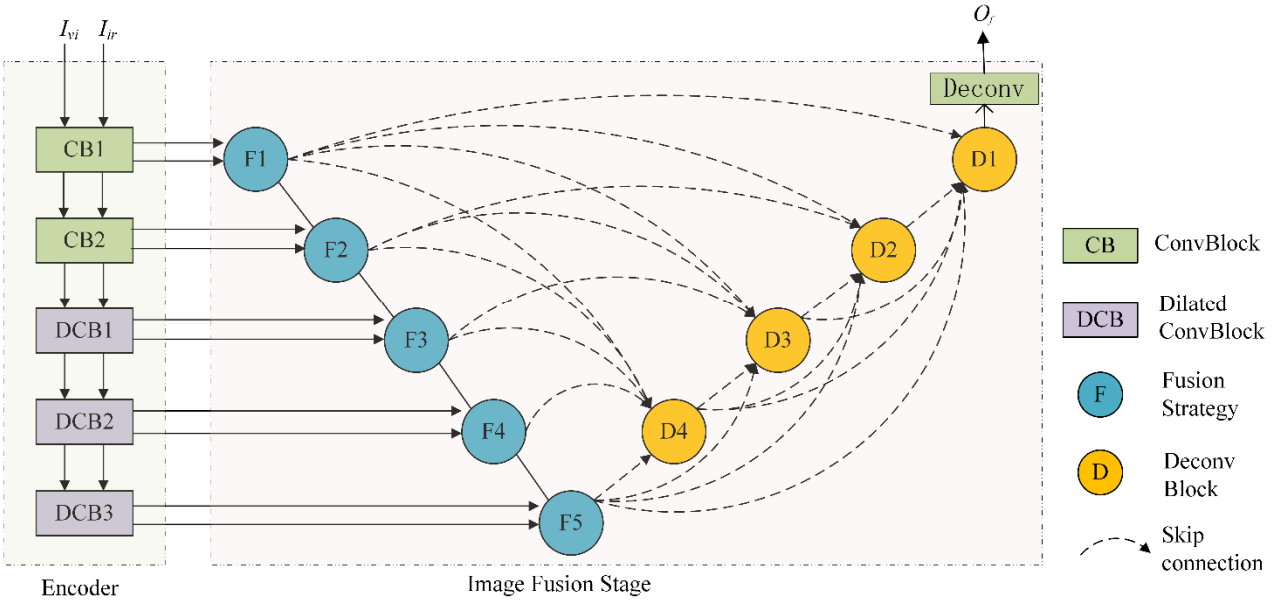


图 1 基于空洞卷积与双注意力机制的融合框架

Fig.1 Fusion network framework based on dilated convolution and dual attention mechanism

1.1 特征提取编码器

如表 1 所示，编码器结构深度为 5 层，用于图像特征的多尺度提取。由于空洞卷积提取图像特征可以保留内部数据结构，可避免多次下采样算子造成的像素损失。同时可以通过设置空洞因子的步长，在不改变分辨率情况下增加感受野[6]。DCDAM 网络中前两层为普通卷积层，后 3 层的卷积块用空洞卷积块替代。卷积块 CB1 和 CB2（ConvBlock）包含两个卷积与一个池化层；空洞卷积块 DCB1（Dilated ConvBlock）和 DCB2 包含 3 个空洞卷积与一个池化层。最后一次卷积提取特征后将进行特征融合，因此 DCB3 则比其他两个空洞卷积块少一个池化层。为消除空洞卷积造成的网格效应，并在图像特征提取中实现特征全覆盖。我们采用 HDC[7]结构，通过将空洞因子设置为锯齿状结构避免像素消失，即 3 个空洞因子分别设置为[1，2，5]。特征提取时，同时输入一组可见光和红外图像，这些特征的融合是可行的，因为相同的卷积层共享相同的权重，这可以输出相同类型的特征。

表 1 编码器的网络设置

Table 1 The encoder network settings

Layers	Channel（Input）	Channel（Output）	Output Size	Dilated Rated
CB1	1	64	1/2	-
CB2	64	128	1/4	-
DCB1	128	256	1/8	1，2，5
DCB2	256	512	1/16	1，2，5
DCB3	512	1024	1/16	1，2，5

1.2 图像融合模块

两幅源图像进行特征提取后输入到图像融合模块。图像融合模块包含两部分：一是双注意力机制的特征融合策略；二是全尺度密集连接解码器。下面我们将分别介绍特征融合和特征解码重建。

1.2.1 双注意力机制特征融合策略

大多数特征融合策略是采用平均加权方式来融合特征。但是这种融合方法无法突出源图像中的重要信息，如红外图像中的目标特征信息。为了获得更好的融合效果，保留重要细节和突出红外目标特征，我们引入双注意力融合策略。我们的特征融合模块与特征提取网络类似，同样具有 5 层。如图 2 所示为其中一层特征融合过程，将提取的可见光图像特征与红外图像特征分别输入通道注意力模块和空间注意力模块进行特征加强后获得注意力特征。本文中通道注意力模块，采用全局池化和 softmax 函数计算加权向量；在空间注意力模块中由 L1 范数和 softmax 函数计算加权向量。将加权向量与原始图相乘后获得通道注意力特征图，最后将注意力特征图与原始特征图相加获得每层融合特征图。

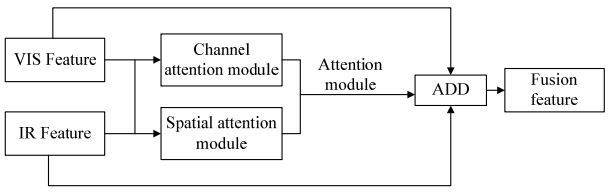


图 2 双注意力机制融合策略

Fig.2 Fusion strategy diagram of dual attention mechanism

1.2.2 全尺度密集连接解码器

解码器目的是从全尺度上探索足够的信息以重建融合图像。特征融合后需要解码器解码以重建融合图像。受到 UNet3+[8]的解码器启发，我们的解码器网络同样采用全尺度密集连接网络。我们将 5 层重建特征映射连接到解码器块中，在每个解码卷积路径上集成一个密集块，将浅层细节信息与高层语义信息无缝集成，为后续重建提供更丰富的特征。图 3 说明了构造密集块 D3 特征图过程。上面两条跳跃连接通过最大池化操作将 F1 和 F2 的特征进行池化下采样，以统一特征图的分辨率。下面两条跳跃连接则通过双线性插值法对解码器中的 D5 和 D4 进行上采样放大特征图的分辨率。统一分辨率后通过 64 个 3×3 大小的滤波器进行卷积，产生 64 个通道的特征图。将 5 个尺度的特征图进行拼接融合后，得到 320 个分辨率相同的特征图。再通过 320 个 3×3 滤波器卷积、BN 和 ReLU 函数后获得解码块 D3。其他解码块同理获得。最后将 D1 进行一次 1×1 卷积重建出融合图像。

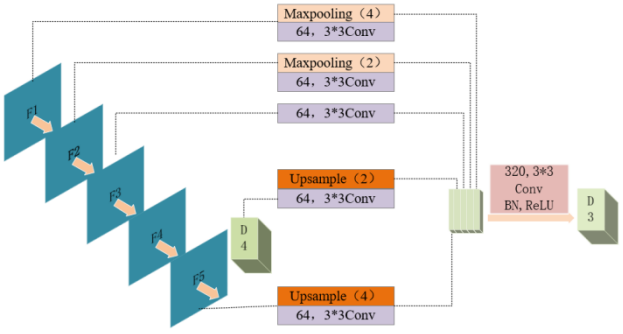


图 3 解码器聚合图

Fig.3 Decoder aggregation diagram

1.3 训练网络

由于红外和可见光图像融合属于异质图像融合，没有融合后的真值图像用于训练，而训练阶段是为了获得良好的网络模型进行特征提取和特征重构，因此我们在训练阶段丢弃融合层。如图 4 所示，输入单张源图像，在特征提取操作后跳过异源融合特征阶段，直接执行特征解码重建的过程。通过计算重建图像和原始图像之间的损失值来训练网络。

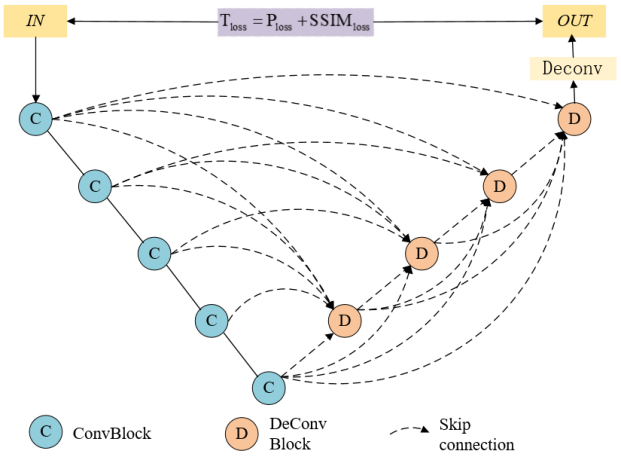


图 4 训练框架

Fig.4 Training network framework

在训练阶段，我们的损失函数由像素损失（P_{loss}）和结构相似性损失（$SSIM_{loss}$）作为总损失函数。这两个损失函数可以同时在像素和结构上约束重建图像与原始图像保持一致。像素损失 P_{loss} 计算公示如下：

$$P_{loss}=\sqrt{\sum_{i\in I,j\in J}\left[O(i,j)-I(i,j)\right]^2} \tag{1}$$

式中：O 与 I 分别代表输出图像和输入图像；(i,j)代表像素点位置。结构相似性函数 SSIM（structural similarity index measure）结合亮度、对比度和结构 3 方面对比图像相似性质量。用 I_A 和 I_B 表示两张图像，计算其结构相似性值表达式为：

$$\mathrm{SSIM}\left(I_A,I_B\middle|W\right)=\frac{2\mu_{I_A}\mu_{I_B}+C_1}{\mu_{I_A}^2+\mu_{I_B}^2+C_1}\times\frac{2\sigma_{I_AI_B}+C_2}{\sigma_{I_A}^2+\sigma_{I_B}^2+C_2} \tag{2}$$

式中：μ_{I_A} 和 μ_{I_B} 代表图像的像素平均值；σ_{I_A} 和 σ_{I_B} 表示图像的标准差；$\sigma_{I_AI_B}$ 表示 I_A 和 I_B 的协方差；C_1 和 C_2 都为常数；W 代表一个滑动窗口。由于红外图像的分辨率较低，为保持全局亮度一致性，本文丢弃亮度部分，则表达式改写为：

$$\mathrm{SSIM}_R\left(I_A,I_B\middle|W\right)=\frac{2\sigma_{I_AI_B}+C}{\sigma_{I_A}^2+\sigma_{I_B}^2+C} \tag{3}$$

在训练中我们将 W 设置为 11×11，C 为 9×10^{-4}。结构相似性损失（$SSIM_{loss}$）定义公式如下：

$$\mathrm{SSIM}_{loss}=1-\frac{1}{N}\sum_{W=1}^{N}\mathrm{SSIM}_R(I_A,I_B|W) \tag{4}$$

式中：N 表示滑窗的总个数。$SSIM_{loss}$ 越小代表融合图像与源图像越相似。网络总损失函数定义如下：

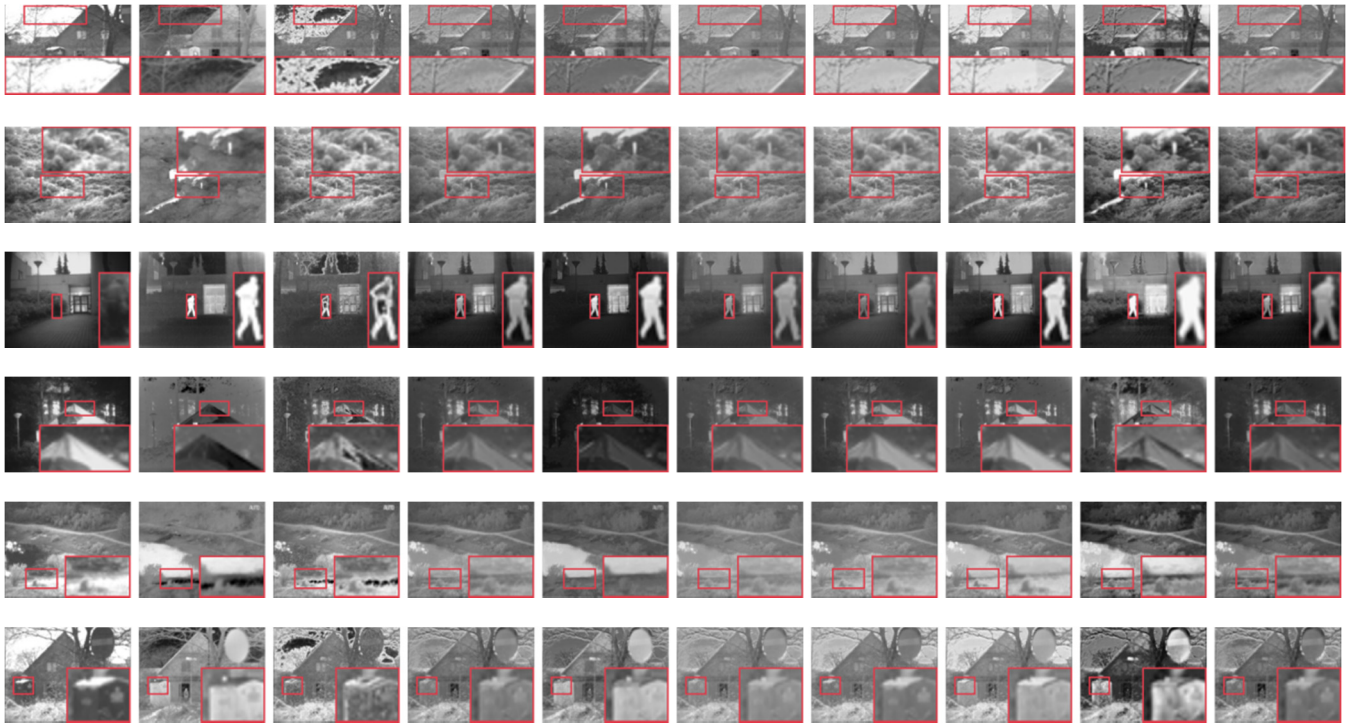
$$T_{loss}=P_{loss}+\mathrm{SSIM}_{loss} \tag{5}$$

2 实验设置与结果分析

我们从 MS-COCO 数据集中选择 80000 张可见光图像，从 KAIST 数据集中选择 20000 张红外图像作为我们的训练数据集。为了验证我们方法的有效性，我们选择了 7 种有代表性的融合方法进行测试实验分析，方法包括交叉双边滤波融合（CBF）[9]、Densefuse、Deeplearning[10]、FusionGAN[11]、Bayesian[12]、NestFuse 和 DDcGAN[13]。同时，因为视觉观测易受到主观因素影响，我们选择了 6 个客观评价指标评估实验结果，包括熵（En）、标准差（SD）、互信息（MI）、无参考图像的改进结构相似度（$SSIM_a$）[14]、视觉信息保真度（VIF）[15]和峰值信噪比（PSNR）。所有的客观评价指标值与融合图像质量成正比。本文实验平台为 NVIDIA GeForce GTX 1650 显卡。我们将部分融合结果的细节图放大到红色框内，便于主观视觉分析；客观指标中最优值用加粗字体，次优值用下划线标出。

2.1 TNO 数据集实验分析

如图 5 所示，我们选取 TNO[16]数据集中 21 组图像进行测试，并将其中 6 组代表性图像展示。从(a)到(j)分别为可见光源图像、红外源图像、CBF、Densefuse、Deeplearning、FusionGAN、Bayesian、NestFuse、DDcGAN 和 DCDAM。从图 5 中总体融合效果显示 CBF 的结果噪声干扰严重，结果产生较多虚假像素和边缘伪影，视觉效果差；Densefuse、Deeplearning 和 Bayesian 的融合结果更偏向可见光图像的细节信息没有突出红外目标且图像对比度低；而 FusionGAN 侧重红外图像而损失了纹理细节信息，并且融合图像产生平滑清晰度低。图中第四组 NestFuse 没有凸显出伞骨的轮廓细节且背景对比度低，视觉效果较不理想，而 DDcGAN 结果偏红外源图像，在树的重叠处出现融合失真，图像中产生边缘伪影。DCDAM 结果中伞的轮廓清晰且失真较小。图中第三组 DDcGAN 忽略了人物的细节同时融合结果有平滑效果导致图像不清晰，NestFuse 中的人物与红外源图像中的目标一致，没有重建衣物细节纹理信息；而 DCDAM 保留了人员衣物细节。图中第六组 DDcGAN 对于邮筒的轮廓重建失真不清晰，NestFuse 没有清晰显示邮筒上的图案，DCDAM 对邮筒的轮廓和细节都有很好的重建效果。图中第五组 NestFuse 目标不突出，DDcGAN 可以突出目标但无法将草的细节特征形态重建出来，DCDAM 的融合结果中不仅草的轮廓和细节纹理清晰，且红外目标与背景细节的对比度高。综上所述，DCDAM 在纹理细节和突出目标上都表现出强大的重构能力。



(a) VIS (b) IR (c) CBF (d) Densefuse (e) Deeplearning (f) FusionGAN (g) Bayesian (h) NestFuse (i) DDcGAN (j) DCDAM

图 5 TNO 数据集实验对比数据

Fig.5 Comparison of TNO dataset fusion results

客观分析中，我们对 21 组融合图像客观指标值取平均值得到表 2 的结果，加粗的数据为最优结果，下划线的数据为次优结果。从表 2 数据显示，我们的融合结果在 EN、MI、$SSIM_a$、VIF 和 PSNR 这 5 种指标均为表现最佳，说明 DCDAM 在信息丰富度和视觉保真度均优于其他方法。SD 指标值稍低是因为双注意力机制导致融合过程中存在特征平滑，导致损失清晰度。但是我们的方法 SD 数值仍处于前几列，并未过于损失清晰度。NestFuse 的多项指标获得次优值，是因为 NestFuse 也采用多尺度特征提取方法。但是不同的是它融合时没有加入原始可见光图像中的细节特征，而且特征提取时下采样算子操作会损失细节信息，导致融合结果中红外目标的细节纹理丢失。DDcGAN 在 EN 指标中获得次佳值是因为其方法产生边缘伪影虚假像素，这与我们主观分析一致。

2.2 INO 数据集实验分析

INO[17]是加拿大光学所录制的视频监控数据集，内容涉及各种生活日常场景。我们对 INO 数据集的视频帧提取后选取 36 组图像作为 INO 测试集。将其中一组典型融合结果扩大展示如图 6，其融合结果客观指标取平均值于表 3 所示。从图 6 中可以看出，CBF、Densefuse、Deeplearning 方法对于重建路灯的轮廓和细节信息都有损失；FusionGAN 中建筑细节信息模糊，边缘信息缺失；Beyesian 在可见光细节重建方面效果较好，但是树枝重建时丢失了红外的轮廓信息；NestFuse 中路灯细节有损失且人物重建结果偏红外不利于视觉观测；DDcGAN 中建筑的细节模糊，同时融合图像背景融入红外源图像中的噪声点；DCDAM 在路灯和建筑的轮廓细节都有很好的重建效果，同时对路灯的轮廓重建也清晰。

表 2 TNO 数据集评价指标

Table 2 Evaluation indexes of TNO dataset

Fusion Methods	EN	SD	MI	$SSIM_a$	VIF	PSNR
CBF[9]	7.38	71.024	10.727	0.5412	0.6301	59.820
Densefuse[3]	5.46	58.520	11.505	0.6876	0.6614	57.825
Deeplearning [10]	5.46	63.866	9.806	0.7488	0.7092	58.583
FusionGAN[11]	5.40	55.654	10.995	0.6173	0.6309	59.251
Bayesian[12]	6.84	65.658	11.343	0.7487	0.6112	<u>59.950</u>
NestFuse[5]	7.28	**78.918**	<u>13.039</u>	<u>0.7634</u>	<u>0.7347</u>	59.097
DDcGAN[13]	<u>7.45</u>	74.808	13.176	0.7224	0.6868	58.684
Proposed	**7.58**	<u>78.722</u>	**13.673**	**0.8035**	**0.7936**	**60.866**

从表 3 的评价指标可以看出，我们的方法在 EN、MI、$SSIM_a$、VIF 和 PSNR 指标都获得了最佳值，说明 DCDAM 在此数据集上也实现了较好的融合结果。值得说明的是 INO 数据集的 PSNR 指标值相较于其他两个数据集的 PSNR 指标值较低，是因为 INO 数据

(a)VIS (b) IR (c) CBF (d) Densefuse (e) Deeplearning

(f)FusionGAN (g) Bayesian (h) NestFuse (i) DDcGAN (j) Proposed

图 6 INO 数据集实验对比数据

Fig.6 Shows of INO dataset fusion result

集是从视频中进行提取帧图像，含有较多噪声点，融合结果皆会包含噪声较多所以导致此指标值较低。DCDAM 的 SD 指标值较低是由于我们的注意力机制在抑制噪声点的同时对图像有平滑效果，而此测试集中噪声较多，导致 DCDAM 融合过程中图像产生平滑导致清晰度不高。而 NestFuse 的 $SSIM_a$ 和 VIF 指标值居第二但是其他指标值低，表明其结构信息重建很好，但是它的红外目标的纹理信息缺失；DDcGAN 方法的 EN 和 MI 指标值高，是因为其融合图像中含有较多噪点，与源红外图像像素保持较多相似，这与主观分析一致。

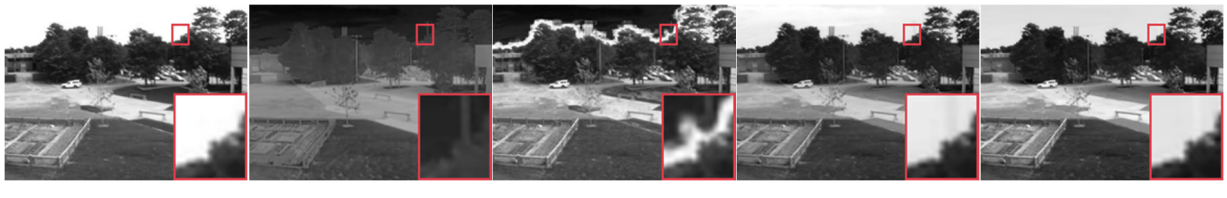
2.3 VOT-RGBT 数据集实验分析

VOT-RGBT[18]数据集是爱尔兰大学利用热成像摄像机和彩色摄像机采用同步锁相方式拍摄。我们选取了 18 组图片作为测试集，将一组融合结果展示如图 7。从图 7 中看出，CBF 融合结果产生失真，融入了噪声干扰；Densefuse、Deeplearning、NestFuse 和 DDcGAN 融合结果没有突出目标特征；FusionGAN 的融合结果出现边缘伪影；Bayesian 融合结果在放大框的结果目标不够突出，且在融合背景天空云朵的特征时有所忽略，边缘细节丢失；DCDAM 不仅红外目标轮廓清晰且边缘信息保留，实现了红外与可见光图像良好的平衡。

表 3 INO 数据集评价指标

Table 3 Evaluation indexes of INO dataset

Methods	EN	SD	MI	$SSIM_a$	VIF	PSNR
CBF	5.241	43.651	10.487	0.622	0.771	55.322
Densefuse	6.914	44.144	13.771	0.692	0.691	55.715
Deeplearnin	6.885	46.649	13.236	0.724	0.863	55.953
FusionGAN	5.248	35.784	10.497	0.652	0.480	55.822
Bayesian	7.355	**60.627**	14.755	0.694	1.048	56.209
NestFuse	6.973	50.317	13.946	0.724	1.203	55.860
DDcGAN	7.427	49.743	14.810	0.679	0.929	55.958
Proposed	**7.593**	57.980	**15.186**	**0.726**	**1.388**	**56.571**



(a)VIS (b) IR (c) CBF (d) Densefuse (e) Deeplearning

(f) FusionGAN (g) Bayesian (h) NestFuse (i) DDcGAN (j) Proposed

图 7 VOT-RGBT 数据集实验对比数据

Fig.7 Shows of VOT-RGBT dataset fusion result

从表 4 可以看出，DDcGAN 在 EN、VIF 和 PSNR 指标值较高但是 SD、MI 和 $SSIM_a$ 指标低表示 DDcGAN 信息丰富度高，但是与源图像相似度低，结果出现了失真。这与主观视觉中融合图像的天空云彩重建效果较好，但是出现了边缘阴影分析一致。NestFuse 的 SD 和 MI 指标较高表示细节纹理重构结构和清晰度较好，与我们主观分析一致。DCDAM 在 6 种评价指标中均实现了最佳，表明 DCDAM 在红外与可见光特征实现了良好的平衡，在保留细节的同时突出了红外目标。

表 4 VOT-RGBT 数据集评价指标

Table 4 Evaluation indexes of VOT-RGBT dataset

Methods	EN	SD	MI	$SSIM_a$	VIF	PSNR
CBF	6.9837	62.287	11.856	0.6542	0.7039	57.367
Densefuse	6.5287	69.880	11.782	0.7384	0.6239	58.117
Deeplearnin	6.8882	63.520	12.792	0.7929	0.6983	58.094
FusionGAN	5.8544	57.944	11.048	0.7777	0.5822	57.757
Bayesian	6.9072	67.608	13.044	0.8084	0.6222	59.971
NestFuse	6.9498	78.408	13.795	0.8031	0.7533	59.040
DDcGAN	7.0221	71.131	12.324	0.7947	0.7562	60.080
Proposed	**7.2823**	**78.517**	**14.043**	**0.8110**	**0.7652**	**61.559**

3 结束语

本文针对红外与可见光图像融合领域对于深层特征提取和利用欠缺，融合图像无法平衡目标与细节信息，提出了基于空洞卷积与双注意力机制的红外与可见光图像融合方法。通过空洞卷积对图像进行多尺度信息提取，将原始特征输入到双注意力模块得到注意力特征，与原始特征聚合成最终融合特征，最后通过一系列密集连接对融合特征加以解码，在 3 个数据集上的主观与客观双重指标证明了我们的网络获得良好的效果。但是当源图像中含有较多噪声点时，注意力机制会噪声抑制对图像进行平滑，导致清晰度欠佳，下一步我们将进一步解决此问题。

参考文献：

[1] LI S, KANG X, FANG L, et al. Pixel-level image fusion: a survey of the state of the art[J]. *Information Fusion*, 2017, **33**: 100-112.

[2] ZHAO W, LU H, WANG D. Multisensor image fusion and enhancement in spectral total variation domain[J]. *IEEE Transactions on Multimedia*, 2017, **20**(4): 866-879.

[3] HUI L, WU X J . DenseFuse: a fusion approach to infrared and visible images[J]. *IEEE Transactions on Image Processing*, 2018, **28**(5): 2614-2623.

[4] JIAN L, YANG X, LIU Z, 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.

[5] 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.

[6] YU F, Koltun V. Multi-scale context aggregation by dilated convolutions[J/OL]. arXiv preprint arXiv:1511.07122, 2015.

[7] WANG P, CHEN P, YUAN Y, et al. Understanding convolution for semantic segmentation[C]//2018 *IEEE Winter Conference on Applications of Computer Vision* (WACV). *IEEE*, 2018: 1451-1460.

[8] HUANG H, LIN L, TONG R, et al. Unet 3+: A full-scale connected unet for medical image segmentation[C]//ICASSP 2020-2020 *IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP). *IEEE*, 2020: 1055-1059.

[9] Shreyamsha Kumar B K. Image fusion based on pixel significance using cross bilateral filter[J]. *Signal, Image and Video Processing*, 2015, **9**(5): 1193-1204.

[10] LI H, WU X J, Kittler J. Infrared and visible image fusion using a deep learning framework[C]//2018 24th *International Conference on Pattern Recognition* (ICPR). *IEEE*, 2018: 2705-2710.

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

[12] ZHAO Z, XU S, ZHANG C, et al. Bayesian fusion for infrared and visible images[J]. *Signal Processing*, 2020, **177**: 107734.

[13] MA J, XU H, JIANG J, 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.

[14] WANG Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. *IEEE Transactions on Image Processing*, 2004, **13**(4): 600-612.

[15] HAN Y, CAI Y, CAO Y, et al. A new image fusion performance metric based on visual information fidelity[J]. *Information Fusion*, 2013, **14**(2): 127-135.

[16] Toet Alexander. TNO Image Fusion Dataset [EB/OL]. 2014, https://doi.org/10.6084/m9.figshare.1008029.v1.

[17] INO. INO's Video Analytics Dataset[EB/OL]. [2022-06-07]. https://www. ino.ca/en/technologies/video-analytics-dataset/.

[18] Conaire C Ó, O'Connor N E, Cooke E, et al. Comparison of fusion methods for thermo-visual surveillance tracking[C]//2006 9th *International Conference on Information Fusion. IEEE*, 2006: 1-7.