文章编号：1673-3193（2025）05-0584-08　　J. North Univ. China, Nat. Sci. Ed., 2025,46(5):584-591.

基于类激活映射的红外与可见光图像融合方法

完　琦，秦品乐，曾建潮

（中北大学 计算机科学与技术学院，山西 太原 030051）

摘　要：针对当前图像融合算法信息选择策略较为固定单一导致源图像重要信息丢失且无效信息干扰融合图像质量等问题，本文提出了一种基于类激活映射的可解释红外与可见光图像融合方法。根据类激活映射机制获取不同源图像的类激活权值（反映了网络对于源图像不同特征的重要性的关注程度），利用类激活权值分配不同通道的特征权重，根据特征权重对提取到的深度特征进行加权融合，以保留源图像更丰富的显著目标和纹理细节等重要信息并抑制噪声信息。实验结果表明，本文所提出的方法在TNO和RoadScene数据集上的表现优于现有的大多数先进算法，其中，TNO数据集上信息熵和视觉保真度分别达到7.327 2和0.692 7，远高于其他方法，这表明本文方法能够在充分保留源图像关键特征信息的同时兼具优秀的视觉感知性能。

关键词：图像融合；信息选择；类激活映射；权重分配；深度学习

中图分类号：TP391.41　　**文献标识码**：A　　**doi:** 10.62756/jnuc.issn.1673-3193.2025.03.0003

引用格式：完琦，秦品乐，曾建潮. 基于类激活映射的红外与可见光图像融合方法[J]. 中北大学学报（自然科学版），2025，46(5)：584-591.

WAN Qi，QIN Pinle，ZENG Jianchao. Method for infrared and visible image fusion based on class activation mapping[J]. Journal of North University of China(Natural Science Edition)，2025，46(5)：584-591.

Method for Infrared and Visible Image Fusion Based on Class Activation Mapping

WAN Qi，QIN Pinle，ZENG Jianchao

（School of Computer Science and Technology，North University of China，Taiyuan 030051，China）

Abstract：To address the issues of fixed and monotonous information selection strategies in current image fusion algorithms，which lead to the loss of critical source image information and interference from invalid noise degrading fusion quality，this paper proposed an interpretable infrared and visible image fusion method based on Class Activation Mapping（CAM）. By leveraging the CAM mechanism，class activation weights were derived from different source images，reflecting the network's attention to feature importance. These weights were utilized to assign channel-specific feature priorities，enabling weighted fusion of deep features to preserve richer salient targets，texture details，and critical information from source images while suppressing noise. Experimental results demonstrate that the proposed method outperformed most state-of-the-art algorithms on the TNO and RoadScene datasets. On the TNO dataset，it achieves superior information entropy（EN）and visual fidelity（VIF）scores of 7.327 2 and 0.692 7，respectively，significantly surpassing existing approaches. This indicates that the proposed method effectively retains key features of source images while exhibiting exceptional visual perception performance.

Key words：image fusion；information selection；class activation mapping；weight allocation；deep learning

收稿日期：2025-03-07

基金项目：山西省科技重大专项计划（202101010101018）

作者简介：完　琦（2000－），男，硕士生，主要从事图像融合的研究。

通信作者：秦品乐（1978－），男，教授，博士，主要从事机器视觉、大数据的研究。E-mail：qpl@nuc.edu.cn。

0　引　言

由于成像设备或成像环境的限制，使用单一类型的传感器得到的图像往往只有场景的部分信息，无法对场景进行全面表征。为此，图像融合技术应运而生。图像融合可以整合源图像中的互补特征，从而生成一幅具有丰富信息的融合图像。红外图像能够突出显著目标但通常忽略了纹理细节，并且容易受噪声影响。相反，可见光图像通常包含了丰富的纹理和结构信息，但容易受光照、遮挡等环境因素干扰。这种互补性需要将红外图像与可见光图像融合在一起，以产生既能突出显著目标，又能展现丰富纹理细节的图像。因此，红外与可见光图像融合技术在多个领域中被广泛应用，如目标检测、语义分割[1]、行人重识别等[2]。

当前红外与可见光图像融合技术有传统方法和基于深度学习的方法两种。传统方法通常采用特定变换提取特征，制定融合规则，最后通过逆变换重构融合图像，可以分为基于多尺度变换的方法[3]、基于稀疏表示的方法[4]、基于子空间的方法[5]、基于显著性的方法[6]、混合方法[7]等，传统融合方法可解释性强、计算资源需求低且无需标注数据，适用于轻量化部署与实时场景，但由于人工设计的融合规则粗糙，如最大值策略、均值策略等，难以适应特征特异性，在复杂场景下易导致图像失真[8]。基于深度学习的红外与可见光图像融合方法通过自动学习多层次特征，有着强大的泛化能力和多模态兼容性，尤其在复杂场景下融合效果更好，但需依赖大量数据与高算力，也存在可解释性差等问题。深度学习融合方法包括基于自编码器、卷积神经网络（Convolutional Neural Network，CNN）和生成对抗网络（Generative Adversarial Network，GAN）的融合方法三种。基于自编码器的方法中，DenseFuse[9]通过密集连接保留了网络中间的有用信息；RFN-Nest[10]通过残差结构和两阶段训练策略优化了细节保留和特征增强；DRF[11]通过将源图像信息来源分解，缓解了特殊信息提取不当的问题。尽管基于自编码器的融合方法可解释性强，但是人工设计的融合规则仍然会导致图像失真。基于CNN[12]的方法中，STDFusionNet[13]使用显著目标掩模针对性地提取源图像中的显著特征进行融合。Tang等[14]提出的PIAFusion中通过感知光照情况解决了极端光照情况下的图像融合。Zhao等[15]提出的MetaFusion通过元特征嵌入融合网络使融合特征和生成的对象语义特征自然兼容，实现了融合任务和检测任务之间的相互促进学习。基于CNN的融合方法实现了端到端融合，局部特征提取能力强且具有一定的自适应能力，但存在局部感受野限制全局信息整合，忽略跨区域重要信息，多模态特征区分能力不足混淆模态特有信息以及可解释性较差等问题。基于GAN的图像融合方法中，Ma等[16]提出的FusionGAN首次将GAN引入到图像融合领域中。为了解决单一鉴别器融合不平衡的问题，Ma等[17]又提出了具有双鉴别器的DDcGAN以保持图像融合的平衡。GANMcC[18]则通过将图像融合转化为多分类约束问题，一定程度上解决了融合结果倾向于红外或可见光某一模态的问题。由于端到端融合算法通常忽略全局依赖关系的融合，Rao等[19]提出了一种基于Transformer和GAN网络的融合算法学习空间与通道维度的全局融合关系以提升复杂场景下的融合效果。基于GAN网络的方法在一定程度上可以在保持有效信息的同时避免图像失真，但是也存在训练稳定性相对较差且无法彻底摆脱对内容损失的依赖等局限性。

综上所述，大多图像融合方法仍然存在以下问题：1）对于不同模态源图像的特征信息选择策略单一，例如仅提取红外图像中的梯度信息或可见光图像的强度信息，导致部分重要的互补信息丢失，且信息选择过程可解释性弱[20]，不确定性强；2）对源图像的不同区域采取相同的处理方式，导致融合过程中除了重要特征还引入了大量的冗余甚至无效的信息[12]。

为了解决这些问题，本文提出了一种通过类激活映射权值进行特征权重分配的可解释红外与可见光图像融合方法。首先，训练一个自编码器来提取深度特征映射并重构源图像。然后，设计一个基于预训练编码器的分类器对红外图像与可见光图像进行区分，根据类激活映射理论，最后一层线性层的激活权值反映了每个特征通道对于分类结果的贡献[21]，即该通道提取到的特征对于该类源图像的重要程度，实现了所有特征的可解释重要性评估。之后，将不同模态源图像的激活权值作为融合阶段信息选择的度量标准，有针对性地对不同通道提取到的深度特征进行权重分配，保留尽可能多的有意义信息而不仅仅是某一类信息，同时抑制干扰或噪声信息，为最终的融合图像保留了丰富的有价值信息，减少了图像失真，也为后续高级视觉任务的应用奠

定了基础。类激活热力图可以直观展示网络对于不同源图像各自重要特征的关注区域，帮助理解模型的关注点，这也大大提高了融合方法的可解释性与确定性。此外，为了提升网络对于不同类别源图像的判别能力，突出源图像类别特征，本文精心设计了损失函数，构建类激活映射与类别无关映射之间的距离损失，驱动主干网络表达目标类别特征，抑制非目标类别特征等冗余信息，从而获得更具判别性的特征表示。

为了验证方法性能，本文在TNO[22]和RoadScene数据集上进行了对比实验以及消融实验，验证了本文方法的先进性和有效性。

1 相关工作

1.1 自编码器网络

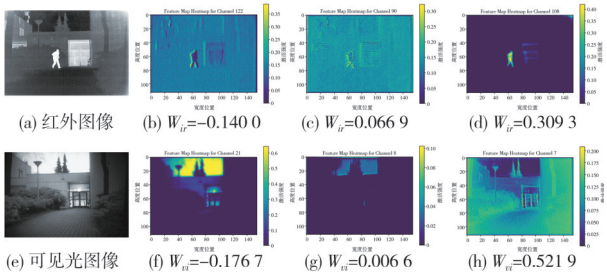
自编码器由编码器和解码器构成，是一种在半监督和无监督学习中使用的人工神经网络。其功能是通过将输入信息作为学习目标，对输入信息进行表征学习，通过输入数据的压缩与重构学习数据表示。其结构通常包含一个输入层、一个或多个隐藏层和一个输出层。自编码器的目标是学习将输入数据压缩到隐藏层中，并能够通过解码器将其从压缩表示中重建出输入数据。编码器将输入数据映射到隐藏层，通常通过一系列的非线性变换和特征提取来实现。解码器则将隐藏层的表示映射回到重构的输入数据，同时也通过非线性变换和特征提取来实现。自编码器通过将输入数据进行压缩和重建的过程来学习有效的数据表示，因此可以应用于依赖特征提取与图像重构的图像融合等领域。

1.2 分类网络

分类网络是计算机视觉领域中最基础也是最重要的任务之一，其目标是将输入的图像分配到预定义的类别中。近年来，随着深度学习技术的快速发展，基于CNN的分类网络取得了显著的成果，并在图像分类、目标检测、图像分割等任务中得到了广泛应用[23]。CNN是一种专门用于处理网格状数据(如图像)的深度学习模型，其核心思想是利用卷积操作提取图像的局部特征，并通过多层堆叠的卷积层和池化层逐步提取更高层次的语义信息。CNN通常由卷积层、池化层、激活函数以及全连接层组成。近年来，研究者们提出了许多经典的CNN模型，如AlexNet、VGGNet、ResNet等，并在图像分类任务上取得了突破性的进展。

1.3 类激活映射

类激活映射[24]是一种用于可视化深度学习模型决策过程的技术，可以反映不同的特征通道对于最终分类决策的重要性，帮助我们理解CNN在分类任务中是如何做出决策的。由类激活权值生成的类激活图直观地展示了图像中哪些区域对于特定分类决策最为重要。具体来说，首先网络通过正常的前向传播处理图像，并在最后一个卷积层得到特征图。然后对这些特征图应用全局平均池化，得到一个向量，该向量的每个元素对应于一个特征通道的全局平均值。接着，这个向量通过全连接层与最终的类别分数相连接。最后，通过将全连接层的权重(每个类别对应一组权重)与最后一个卷积层的特征图加权，生成类激活映射。类激活映射机制可以评估原始信息的重要程度，增强重要特征，并抑制噪声或无关信息，计算过程简单高效，因此可以被应用于非常依赖信息选择的图像融合领域。举例来说，在红外图像与可见光图像的分类结果中，如果通道k的特征包含了红外图像的重要特征，比如显著目标，那么与其对应的红外权值W_{ir}^{k}就会很大，但如果通道k几乎没有可见光图像的重要特征，那么可见光权值W_{vi}^{k}就会非常低。而在一些包含了红外图像与可见光图像共同特征的通道中，W_{ir}^{k}与W_{vi}^{k}会呈现近似的值，从而避免信息丢失。如图1所示，通过可视化热力图，可以直观地看到权重较高的特征通道提取到了该源图像中重要的特征信息，较低权重则对应于该类别源图像的次要或无关信息，而负权重则对应于干扰或噪声信息。



(a) 红外图像 (b) W_{ir}=−0.140 0 (c) W_{ir}=0.066 9 (d) W_{ir}=0.309 3

(e) 可见光图像 (f) W_{vi}=−0.176 7 (g) W_{vi}=0.006 6 (h) W_{vi}=0.521 9

图1 可视化热力图

Fig. 1 Visual heatmap

2 本文方法

本文所提方法通过类激活映射机制评估源图像特征信息的重要程度，将类激活权值作为融合阶段红外图像与可见光图像特征信息选择的度量

标准进行特征权重分配，增强较高权值所对应的重要特征信息并抑制负权重所对应的噪声信息，实现了特征信息的灵活选择，在融合图像中保留了更丰富的重要互补信息并一定程度上减少了干扰信息对融合图像质量的影响。

2.1　网络结构

模型的整体框架如图 2 所示，包括两个部分，特征权重分配网络（Feature Weight Allocation Network，FWA-Net）和红外与可见光图像融合网络（Infrared and Visible Image Fusion Network，IVF-Net）。其中，FWA-Net 由编码器和分类器网络构成，IVF-Net 主要由自编码器组成。

1）分类器网络结构。分类器网络如图 2 中FWA-Net 所示，由预训练好的编码器和分类块构成。分类块由六个深度卷积层和全局平均池化层以及全连接层组成。深度卷积层中使用四个并行分支以提高模型性能。深度卷积具有参数量少，内存占用小，计算效率高等特点。最为重要的是，深度卷积对每个通道使用独立的卷积核，不执行通道求和，因此通道顺序不会进行重新排列，这保证了每个特征权重与通道之间的对应关系保持不变。

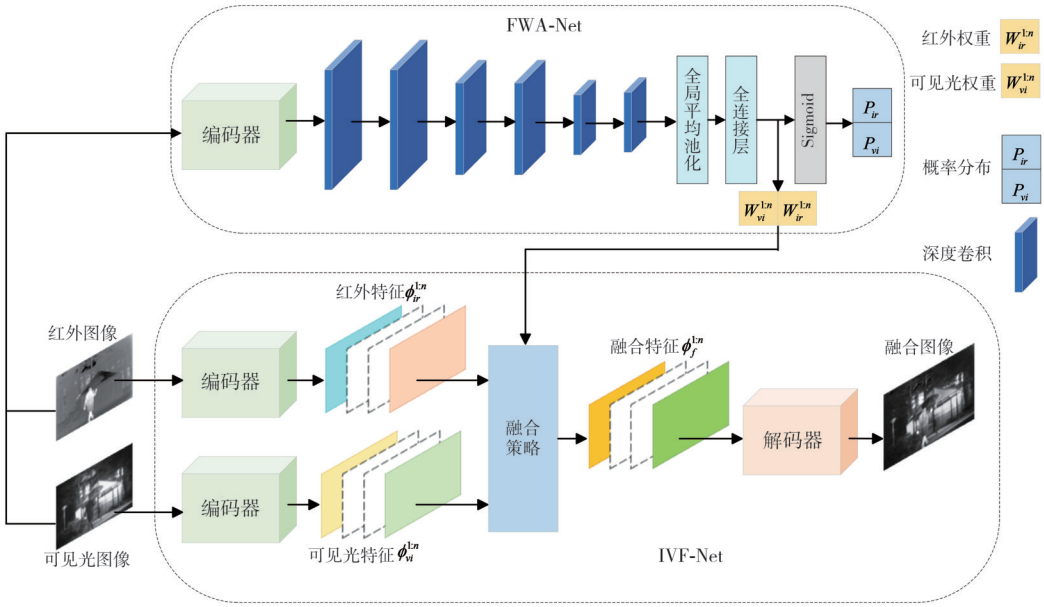


图 2　网络框架图

Fig. 2　Network architecture diagram

2）自编码器网络结构。自编码器网络结构如图 3 所示，包括编码器和解码器两部分。编码器使用 5 个卷积层进行特征提取。每层均使用 3×3 卷积核和 ReLU 激活函数。在第二层和第四层使用步长为 2 的卷积层代替最大池化层进行下采样操作，可以保留更多的特征信息。解码器中依旧使用 5 个卷积层，在最后一层中使用 1×1 卷积以及 tanh 激活函数从深度特征映射中重构图像[17]。同时，使用双线性插值进行上采样操作。

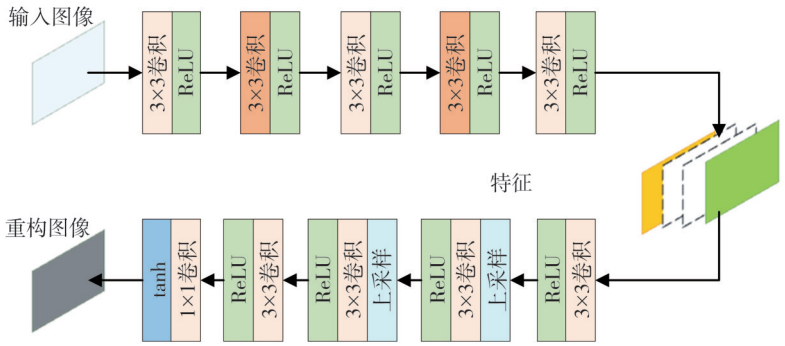


图 3　自编码器网络图

Fig. 3　Autoencoder network diagram

2.2 FWA-Net

FWA-Net主要用于获取红外图像与可见光图像分别对应的类激活权值，并进行特征权重分配

$$\boldsymbol{W}_c^{1:n}=C(I_c), \tag{1}$$

式中：$C(\cdot)$为分类函数；I_c为c类的输入图像；$c\in\{ir,vi\}$，ir和vi分别表示红外类别和可见光类别。通过对分类器最后一层的输出$\hat{\phi}^{1:n}$执行全局平均池化操作得到$\boldsymbol{W}_c^{1:n}$，$\boldsymbol{W}_c^{1:n}$表示分类器中最后一个线性层的类激活权值，对应c类源图像的特征权重。类激活映射机制保证了$\hat{\phi}^k$与$\boldsymbol{W}_c^k(k\in[1,n])$之间的对应关系，$n$表示通道数。之后，将网络末端线性层的输出传递给Sigmoid激活函数得到概率分布

$$P_c=\sigma\left(\sum_k \boldsymbol{W}_c^k\cdot GAP(\hat{\phi}^k)\right), \tag{2}$$

式中：P_c为I_c属于c类的概率；$\sigma(\cdot)$为Sigmoid函数；$GAP(\cdot)$表示全局平均池化。这样线性权值$\boldsymbol{W}_c^k$可以直接代表通道k对于分类结果的贡献。$\boldsymbol{W}_c^{1:n}$可以揭示每个通道所包含该类别源图像信息的丰富度与重要性。之后进行特征权重分配，表示为

$$\hat{\boldsymbol{W}}_c^{1:n}=0.5*\tanh(\boldsymbol{W}_c^{1:n})+0.5, \tag{3}$$

式中：$\hat{\boldsymbol{W}}_c^{1:n}$表示将类激活权值归一化之后的特征权重。归一化由具有更好的数值稳定性和梯度流的非线性函数$\tanh(\cdot)$实现。

2.3 IVF-Net

IVF-Net用于接收FWA-Net分配的特征权重并根据特征权重进行图像融合。使用相同的预训练编码器进行特征提取，结合深度卷积的特性以确保$\phi_c^{1:n}$与$\hat{\boldsymbol{W}}_c^{1:n}$之间的对应关系。给定一对配准的红外图像与可见光图像，分别定义为I_{ir}，I_{vi}。首先使用编码器$E_F(\cdot)$将源图像映射到特征空间，表示为

$$\{\phi_{ir}^{1:n},\phi_{vi}^{1:n}\}=\{E_F(I_{ir}),E_F(I_{vi})\}, \tag{4}$$

式中：$\phi_{ir}^{1:n}$和$\phi_{vi}^{1:n}$分别表示从I_{ir}和I_{vi}中提取的深度特征映射；n表示通道数。然后，使用FWA-Net分配的特征权重对深度特征进行加权融合，表示为

$$\phi_f^{1:n}=\sum_c \hat{\boldsymbol{W}}_c^{1:n}\otimes\phi_c^{1:n}, \tag{5}$$

式中：$\otimes$表示加权操作。最后，利用解码器$D_F(\cdot)$从融合特征映射$\phi_f^{1:n}$中生成融合图像，表示为

$$I_f=D_F(\phi_f^{1:n})。 \tag{6}$$

2.4 损失函数

1）自编码器损失函数。自编码器使用损失函数L_{ae}训练，定义为

$$L_{ae}=L_{pixel}+\lambda L_{ssim}, \tag{7}$$

式中：L_{pixel}和L_{ssim}分别表示输入图像I_i和重构图像I_r之间的像素损失和结构相似性(SSIM)损失；λ表示L_{pixel}和L_{ssim}之间的权衡值。

L_{pixel}在像素级上约束输入图像与重构图像之间的相似性，其定义为

$$L_{pixel}=\frac{1}{HW}\|I_r-I_i\|_F^2, \tag{8}$$

式中：H和W分别为输入图像的高度和宽度；$\|\cdot\|_F$代表矩阵的Frobenius范数。

结构相似性损失L_{ssim}定义为

$$L_{ssim}=1-SSIM(I_r,I_i), \tag{9}$$

式中：$SSIM(\cdot)$表示结构相似性度量[9]，$SSIM(\cdot)$的值越大，说明输入图像I_i与重构图像I_r在结构上的相似性越大。

2）分类器损失函数。在分类器训练阶段，预训练好的编码器是固定的。分类器中其它层的参数更新依赖于损失函数L_{class}，其定义为

$$L_{class}=L_{ce}+\alpha L_{df}, \tag{10}$$

式中：L_{ce}和L_{df}分别表示二元交叉熵损失和距离损失；α表示合并比。

交叉熵损失L_{ce}定义为

$$L_{ce}=-(y\log(\hat{y})+(1-y)\log(1-\hat{y})), \tag{11}$$

式中：y表示输入图像的标签；$\hat{y}$即分类器网络的输出P_c。

类激活映射可以用于识别特定类别的空间判别区域，通过累加特征映射，也可以得到一个类别无关的激活映射(CAAM)[24]，它表明了特征在空间上的分布。可以发现，CAAM通常比目标类别的激活映射具有更大的激活区域和更丰富的特征，但它同时也具有非常多冗余的非目标特征。如果将CAAM约束得更为接近目标类别的激活映射，就可以在抑制非目标类别特征的同时更有效地突出目标类别的特征。因此，通过最小化每个训练图像的CAAM与目标类别的激活映射之间的距离构造损失函数L_{df}，以驱动主干网络从空间角度学习更多的判别特征表示，其定义为

$$L_{df}=\frac{1}{HW}\sum_{x,y}\|m_i(x,y)-m_r(x,y)\|_{l1}, \tag{12}$$

式中：$m_i(x,y)$和$m_r(x,y)$分别表示类别无关映射和类激活映射；$\|\cdot\|_{l1}$表示使用l_1距离测量像素空间距离。

$$m_i(x,y)=MMN\left(\sum_k \phi^k(x,y)\right), \tag{13}$$

$$m_r(x,y)=MMN\left(\sum_k W_c^k \phi^k(x,y)\right), \tag{14}$$

式中：$\phi^k(x,y)$表示特征单元k在空间位置(x,y)的激活值；$MMN(\cdot)$表示离差标准化。

3　实验及结果分析

在 TNO[22]数据集和 RoadScene 数据集上进行实验，以验证本文所提方法的有效性。

3.1　数据集及实验设置

TNO[22]数据集是一个多波段图像融合数据集，提供了多种军事和监视场景的图像。RoadScene 数据集提供了 221 个配准的红外与可见光图像对，包含了诸如行人、车辆以及道路等丰富的场景。为了得到更加强大的特征提取和图像重构能力，在自编码器的训练阶段，不能局限于使用红外与可见光图像。本文使用包含大量复杂日常场景的 MS-COCO 数据集训练自编码器，选择 8 000 张图像，调整图像尺寸为 256×256，并归一化为[−1，1]。针对 TNO[22]和 RoadScene 数据集分别训练两个单独的分类器以适应两个数据集的独特分布特征，确保每个分类器在特定数据集上的性能最优化。由于两个数据集中的图像数量都很少，裁剪原始图像并通过滑动窗口获得 256×256 的图像补丁以支持分类器网络的训练。

本文方法通过在 64 位的 NVIDIA 服务器上使用 Pytorch 实现，服务器内存为 128 G，拥有 4 张显存为 32 G 的 TeslaV100 显卡。模型使用 Adam 优化器更新参数，学习率设置为 10^{-6}，batchsize 设置为 16，epoch 设置为 30。为了平衡损失，λ 设置为 100。关于 α 的设置，由于最开始的 epoch 中 m_r 过于离散，所以将其视作一个简单的阶跃函数，当 epoch 小于 20，α 设置为 0，反之为 3。

3.2　结果对比

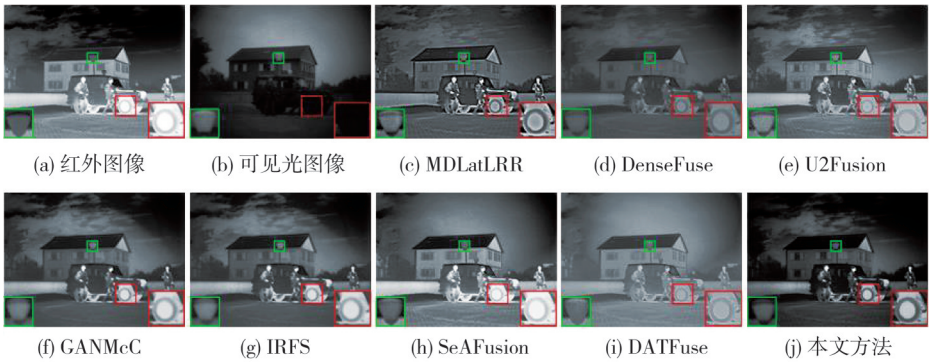
在 TNO[22]和 RoadScene 数据集上对本文方法和现有主流方法进行了结果对比和分析，以验证本文方法的先进性。本文比较了经典图像融合方法中的 MDLatRR[25]、Densefuse[9]、U2Fusion[12]、GANMcC[18]以及近年来的新颖 SOTA 方法 IRFS[26]、SeAFusion[27]、DATFuse[28]。

TNO[22]数据集上的定性和定量结果分别如图 4 和表 1 所示。

表 1　TNO 数据集的定量指标对比

Tab. 1　Comparison of quantitative metrics on the TNO dataset

方法	EN	AG	SCD	SSIM	VIF
MDLatLRR	6.587 4	3.016 9	1.443 0	0.775 0	0.352 3
DenseFuse	6.887 6	3.591 7	1.748 3	0.731 6	0.657 1
U2Fusion	7.071 0	**4.749 6**	1.742 6	0.777 9	0.661 5
GANMcC	6.849 4	2.537 4	1.541 3	0.710 8	0.417 9
IRFS	6.729 4	3.596 4	1.706 4	**0.791 5**	0.674 0
SeAFusion	7.124 2	4.018 5	1.764 3	0.769 4	0.673 3
DATFuse	6.580 3	3.361 2	1.479 2	0.749 4	0.639 4
本文方法	**7.327 2**	4.112 4	**1.874 7**	0.778 3	**0.692 7**



(a) 红外图像　(b) 可见光图像　(c) MDLatLRR　(d) DenseFuse　(e) U2Fusion

(f) GANMcC　(g) IRFS　(h) SeAFusion　(i) DATFuse　(j) 本文方法

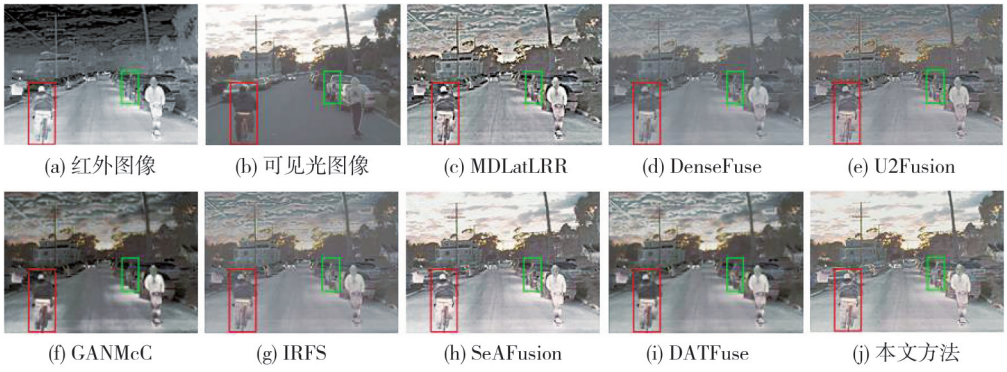
图 4　TNO 数据集的可视化结果对比

Fig. 4　Comparison of visualization results on the TNO dataset

从图 4 的定性结果中可以看到，本文方法相较于其他方法有着更好的视觉感受，在保留重要信息的同时减少了噪声信息的干扰，由红色和绿色框中放大的部分可以看出本文方法能够最大程度地同时保留红外显著目标和可见光纹理细节。表 1 的定量指标对比表明本文方法在信息熵（EN）、差异相关性总和（SCD）以及视觉保真度（VIF）上达到了最高的平均值，在平均梯度（AG）和结构相似性度量（SSIM）上也有较好的表现。

RoadScene 数据集上的定性和定量结果分别如

图5和表2所示。表2中除AG外所有指标都达到最高的平均值，这也验证了本文方法的先进性。



(a) 红外图像 (b) 可见光图像 (c) MDLatLRR (d) DenseFuse (e) U2Fusion

(f) GANMcC (g) IRFS (h) SeAFusion (i) DATFuse (j) 本文方法

图5 RoadScene数据集的可视化结果对比

Fig. 5 Comparison of visualization results on the RoadScene dataset

表2 RoadScene数据集的定量指标对比

Tab. 2 Comparison of quantitative metrics on the RoadScene dataset

方法	EN	AG	SCD	SSIM	VIF
MDLatLRR	6.771 4	3.649 9	1.484 4	0.893 1	0.372 9
DenseFuse	7.058 4	4.542 6	1.786 3	0.920 5	0.547 1
U2Fusion	7.255 5	**6.060 6**	1.732 5	0.915 7	0.509 5
GANMcC	7.329 6	3.660 1	1.759 0	0.823 4	0.360 7
IRFS	6.879 3	5.940 6	1.739 1	0.937 9	0.662 4
SeAFusion	7.275 1	5.519 9	1.797 0	0.934 0	0.661 7
DATFuse	6.730 3	5.283 4	1.511 7	0.914 1	0.627 7
本文方法	**7.476 8**	6.035 2	**1.906 9**	**0.943 8**	**0.680 9**

3.3 消融实验

为了验证模型所提方法以及空间判别特征损失函数L_{df}的有效性，对TNO测试数据集进行相关消融实验的定量分析：1）为验证特征权重分配模块的有效性，仅用自编解码器搭配$l1$-norm规则进行融合实验；2）为验证损失函数L_{df}的有效性，仅去除损失函数L_{df}进行融合实验。选取EN、AG、SCD、SSIM和VIF作为消融实验的客观评价指标，结果如图6和表3所示。

(a) $l1$-norm (b) 去除L_{df} (c) 本文方法

图6 消融实验的可视化结果

Fig. 6 Visualization results of ablation experiment

表3 消融实验的客观评价指标

Tab. 3 Ablation experiment objective evaluation indicators

方法	EN	AG	SCD	SSIM	VIF
$l1$-norm	6.960 2	3.020 4	1.671 2	0.737 4	0.481 9
去除L_{df}	7.324 3	4.047 2	1.819 6	0.766 7	0.664 7
本文方法	7.327 2	4.112 4	1.874 7	0.778 3	0.692 7

由表3可以看出，本文方法在VIF指标上达到了0.692 7，相较于传统$l1$-norm融合规则的0.481 9有较大的提升，而当去除空间判别特征损失函数时，VIF的值仅为0.664 7，这说明本文所提出的融合方法以及空间判别特征损失函数能够提高融合性能，验证了本文方法的有效性。进一步分析可知，大多融合方法不具备灵活选择信息的能力，而本文方法可以突出不同模态源图像中更丰富的有价值信息并抑制噪声信息。精心设计的损失函数则可以驱动主干网络从具有更大激活区域和更丰富特征的类别无关映射中学习更多的判别特征表示。

4 结 论

本文提出了一种基于类激活映射的图像融合方法，根据类激活权值实现对不同源图像中各通道所有特征的可解释重要性评估，将类激活权值的大小作为融合阶段特征信息选择的度量标准，为融合图像中保留更多重要的关键信息。此外，本文方法通过学习更多的空间判别特征，融合了源图像中更丰富的有价值信息并抑制了噪声信息，解决了由于信息选择策略单一导致的图像失真与噪声信息干扰问题。结果表明，与大多数现有的方法相比，本文方法能够更有效地融合互补信息。后续将研究通过优化分类器及激活权重提高图像融合能力。

参考文献：

[1] HA Q, WATANABE K, KARASAWA T, et al. MFNet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes[C]// 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2017: 5108-5115.

[2] TANG L, XIANG X, ZHANG H, et al. DIVFu-

sion：Darkness-free infrared and visible image fusion [J]. Information Fusion，2023，91：477-493.
[3] CHEN J，LI X，LUO L，et al. Infrared and visible image fusion based on target-enhanced multiscale transform decomposition[J]. Information Sciences，2020，508：64-78.
[4] WU M，MA Y，FAN F，et al. Infrared and visible image fusion via joint convolutional sparse representation[J]. Journal of the Optical Society of America A，2020，37(7)：1105-1115.
[5] LI S，ZOU Y，WANG G，et al. Infrared and visible image fusion method based on a principal component analysis network and image pyramid[J]. Remote Sensing，2023，15(3)：685.
[6] CHEN J，WU K，CHENG Z，et al. A saliency-based multiscale approach for infrared and visible image fusion [J]. Signal Processing，2021，182：107936.
[7] LI X，TAN H，ZHOU F，et al. Infrared and visible image fusion based on domain transform filtering and sparse representation[J]. Infrared Physics & Technology，2023，131：104701.
[8] SONG Z，QIN P，ZENG J，et al. EdgeFusion：Infrared and visible image fusion algorithm in low light [C]//Pattern Recognition and Computer Vision . Singapore：Springer Nature Singapore，2023：259-270.
[9] 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.
[10] 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.
[11] XU H，WANG X，MA J. DRF：Disentangled representation for visible and infrared image fusion [J]. IEEE Transactions on Instrumentation and Measurement，2021，70：1-13.
[12] 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.
[13] MA J，TANG L，XU M，et al. STDFusionNet：An infrared and visible image fusion network based on salient target detection [J]. IEEE Transactions on Instrumentation and Measurement，2021，70：1-13.
[14] TANG L，YUAN J，ZHANG H，et al. PIAFusion：A progressive infrared and visible image fusion network based on illumination aware [J]. Information Fusion，2022，83：79-92.
[15] ZHAO W，XIE S，ZHAO F，et al. MetaFusion：Infrared and visible image fusion via meta-feature embedding from object detection[C]//2023 the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR)，2023：13955-13965.
[16] MA J，YU W，LIANG P，et al. FusionGAN：A generative adversarial network for infrared and visible image fusion[J]. Information Fusion，2019，48：11-26.
[17] 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.
[18] 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，2020，70：1-14.
[19] RAO D，XU T，WU X J. TGFuse：An infrared and visible image fusion approach based on transformer and generative adversarial network[J]. IEEE Transactions on Image Processing，2023：3273451.
[20] 余东，蔺素珍，禄晓飞，等. 基于显著性的多波段图像同步融合方法[J]. 红外技术，2022，44(10)：1095-1102.
YU Dong，LIN Suzhen，LU Xiaofei，et al. Saliency-based multiband image synchronization fusion method [J]. Infrared Technology，2022，44(10)：1095-1102.(in Chinese)
[21] ZHOU B，KHOSLA A，LAPEDRIZA A，et al. Learning deep features for discriminative localization [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR)，2016：2921-2929.
[22] TOET A. The TNO multiband image data collection [J]. Data in Brief，2017，15：249-251.
[23] 金海彬，吕志贤，侯木舟，等. 基于特征融合与SVM的内镜图像分类算法研究[J]. 中北大学学报(自然科学版)，2023，44(1)：86-96.
JIN Haibin，LÜ Zhixian，HOU Muzhou，et al. Research on endoscopic image classification algorithm based on feature fusion and SVM[J]. Journal of North University of China(Natural Science Edition)，2023，44(1)：86-96.(in Chinese)
[24] WANG C，XIAO J，HAN Y，et al. Towards learning spatially discriminative feature representations [C]//2021 IEEE/CVF International Conference on Computer Vision(ICCV)，2021：1306-1315.

（下转第610页）