DOI:10.19651/j.cnki.emt.2416031

基于全局与局部注意力的车辆方位场景识别*

翟永杰[1] 刘 璇[1] 王新颖[1] 王乾铭[1] 刘金龙[2]
(1.华北电力大学控制与计算机工程学院 保定 071003; 2.邦邦汽车销售服务(北京)有限公司 北京 100020)

摘 要: 针对当前车辆方位场景识别任务中存在因特征相似而导致的左右类别、前后类别识别混淆等问题,设计了融合全局-局部注意力的车辆方位场景识别方法。首先,引入车辆多方位场景的概念,通过 OSMNet 方位场景匹配网络进行特征提取并进行场景的分类,其次,为使模型在不同方位场景中聚焦关键区域以有效学习车辆的空间方位,设计了全局局部注意力模块。最后,针对部分车辆方位场景之间存在类间距离小于类内距离的问题,设计了全局-局部位置注意力模块。在构建的 8 类场景数据集上进行实验,消融实验显示,本文所提出的 D-CBAM 和 HGLP 模块有效地增强了对特征图的全局和局部信息的捕捉能力,将模型识别准确率提高了 3.54%和 4.22%;对比实验显示,所提模型的准确率达到了 95.49%,与基线模型相比提高了 5.46%,模型总体识别效果优于其他分类模型,其中大部分方位的识别效果优于基线模型。结果表明本文提出的改进分类模型有效的学习了车辆方位信息,为远中近景图像的匹配提供桥梁,同时也为车辆多部件检测和分割等任务奠定基础。

关键词: 车辆方位;场景识别;深度学习;图像分类;全局注意力;位置注意力
中图分类号: TN911.73 **文献标识码**: A **国家标准学科分类代码**: 520.20

Vehicle orientation scene recognition based on global-local attention

Zhai Yongjie[1] Liu Xuan[1] Wang Xinying[1] Wang Qianming[1] Liu Jinlong[2]
(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China;
2. Bangbang Automobile Sales and Service (Beijing) Co., Ltd., Beijing 100020, China)

Abstract: To address issues such as confusion in distinguishing left-right and front-back categories due to similar features in current vehicle orientation scene recognition tasks, we proposed a vehicle orientation scene recognition method that integrates global-local attention. We introduced the concept of multi-view vehicle scenes, utilized OSMNet for feature extraction and scene classification, and developed a global-local attention module to focus on key areas across different orientation scenes for effective spatial orientation learning. Additionally, we designed a global-local positional attention module to address overlapping class distances between certain vehicle orientation scenes. Experiments on an 8-class scene dataset demonstrated that our D-CBAM and HGLP modules effectively enhanced the capture of global and local information in feature maps, improving model recognition accuracy by 3.54% and 4.22%, respectively, in ablation studies. Comparative experiments showed that our model achieved an accuracy of 95.49%, which is 5.46% higher than the baseline model. Overall, our model outperformed other classification models in recognizing most orientations better than the baseline model. These results demonstrate that our improved classification model effectively learns vehicle orientation information, bridging the gap for matching images from distant, intermediate, and near perspectives, and laying a foundation for tasks such as multi-part vehicle detection and segmentation.

Keywords: vehicle orientation; scene recognition; deep learning; image classification; global attention; positional attention

0 引 言

在车辆逐渐增多的当今社会,城市车辆碰撞等事故频发[1-2]。在传统的保险业务中,汽车定损一般要求由保险公司的定损工作人员经过现场勘察,从而判断汽车的损毁状况并记录在案,作为车辆保险理赔的依据。其中车辆现场勘查是汽车保险索赔中的关键部分,但目前一般是由保险公司工作人员到场办理,消耗大量人力物力。传统的汽车交通事故索赔流程中,事故车辆车主不能私自移动车辆,需经过报警、责任认定、保险公司人员到场勘察并拍照等繁琐

收稿日期:2024-05-13
*基金项目:国家自然科学基金资助项目(62373151)、河北省自然科学基金面上项目(F2023502010)、中央高校基本科研业务费专项资金(2023JC006)项目资助

流程后,事故车辆才可以进行移动,增加了城市交通的负担,由此可见传统的汽车理赔过程中存在很多不合理的地方。为了解决这些问题,智能定损系统应运而生[3],通过采用目标检测[4-6]、图像分类[7-8]、图像语义分割[9-11]和实例分割[12]等智能算法,对车辆图片执行智能定损,检测损伤的范围、种类以及受损的部件,从而降低定损成本,精简整个行业流程,推动汽车保险领域的技术进步,加速定损工作的业务流程。

车辆方位场景分类,作为智能定损系统中的重要一环,依托于图像分类技术,能够准确识别来自不同视角和距离的车辆图片中车辆的方位信息。对于同一辆车,定损员会从多个不同的视角拍摄,因此得到的该车部件图片不仅涉及中景,还包括一些远景和近景,由于远景背景信息过多,车辆部分作为主要信息占比较少,导致车辆部件检测的精度受到限制;而近景只含有车辆的局部信息,检测网络虽然可以识别出部件,但是难以根据这些局部信息分辨出部件的左右类别,因此会出现误判的问题;将同一辆车的局部近景、和其对应的中景、远景图片进行匹配,可以帮助检测模型更好的理解近景视角下的部件在整部车辆所处的位置,从而解决左右类别误判问题提高车辆部件检测的准确性[13]。而准确的识别出车辆所处的方位场景,是智能定损系统实现近中远景匹配的基础。

车辆方位场景识别任务通过图像分类技术准确地从不同距离、视角中识别出车辆所处的方位,为各种应用场景提供更准确的图像理解和分析能力,同时也为智能定损系统及理赔过程带来了巨大的助力。

本文提出车辆方位场景的概念,训练并推理了基于改进的 ResNet101 的场景分类模型,旨在更精准地识别车辆图像所属的方位场景,为多部件检测任务提供更准确的先验信息。为了使模型更聚焦关键区域特征,增强对不同场景的感知能力,本文创新性地引入了两个综合考虑全局与局部特征的模块:分布感知注意力模块(distribution-aware convolutional block attention module,D-CBAM)和全局-局部位置注意力模块(hybrid global-local position attention module,HGLP),进一步提升检测模型分类的效果,形成了最终的方位场景匹配网络(orientation scene matching network,OSMNet)。

1 相关工作

与车辆方位场景识别的研究相比,现有的一些工作主要集中于车辆或车型的分类任务上[14-15]。文献[16]提出了一种先分割后检测的深度学习方法对高分辨率遥感图像中车辆进行检测和分类,并在不同数据集上进行微调,利用精确的语义分割图推断车辆的形状和类型,超越了传统的车辆分类方法。文献[17]提出了一种可变形部件模型(deformable part-based models,DPM)来综合考虑车辆的检测和分类,即为每一类车辆训练一个 DPM,然后在交通场景图像中使用这些模型进行车辆检测和提取,接着对提取的车辆图像进行模型对齐和特征提取,最后使用支持向量机进行车辆分类。文献[18]采用了一种基于直方图梯度方向特征的神经网络方法来对车辆的方向进行估计。作者将方向估计视为一个多类别分类问题,将方向分为不同的类别。但是其对尺度变换比较敏感,另外采用的是航拍数据集,目标较小,不适用于车辆定损的场景。还有一些基于深度学习的分类算法,能够有效地识别遥感图像中车辆或建筑物的方位。文献[19]提出了一种定向区域卷积神经网络(oriented region-based convolutional neural network,OR-CNN),在 VGG 模型的基础上,引入了一个新的定向层网络,用于检测建筑物的旋转角度。然后,利用定向矩形来表示建筑物的位置和方向。文献[20]则提出了一种车辆检测框架,首先使用一个基于超特征图的准确车辆提议网络(accurate vehicle proposal network,AVPN),来生成候选的车辆区域,然后,使用一个车辆属性学习网络(vehicle attribute learning network,VALN),来对候选区域进行验证和分类,同时提取车辆的方向和类型等属性。除此以外,很多经典的网络在图像分类任务中取得了巨大的成功。2020 年谷歌团队将 Transformer 架构用于图像分类工作[21],通过自注意力机制和多层感知器来学习图像的全局表示,但是它的数据和计算资源需求量大,且不能充分利用图像的局部信息,不适用于需要充分考虑局部和全局信息的车辆方位场景识别任务。Zhu 等[22]在 2023 年提出了一种基于双层路由注意力的视觉变换器架构 BiFormer 进行图像分类,它可以动态地选择与查询相关的一小部分 token 进行注意力计算,从而节省计算资源,然而 BiFormer 注意力区域的划分是固定的,不能充分发掘全局特征,这在需要更好的全局感知性能的车辆方位场景识别任务中将成为一个局限。

车辆方位场景识别在汽车安全领域和智能定损等各种应用场景中扮演着不可或缺的角色(如图 1 所示)。目前的车辆方位场景分类任务中,仍然面临着一些严峻的挑战。这些挑战主要源自于车辆方位之间存在一些和部件相关的高度相似的特征,现有的分类工作无法充分考虑并利用车辆方位之间的细微差异,难以解决现实场景中由于车辆的左右对称结构引起的部件特征相似性问题。需要更深入的研究来提高分类网络在车辆方位场景中的性能,进而推动车辆部件检测任务中的中近景匹配,改善车辆智能定损系统中近景车辆部件的局部定位工作。

2 研究方法

本文总体网络架构如图 2 所示,首先,将车辆方位场景图像作为输入,通过一个 7×7 的卷积和步长为二的最大池化层生成初步的特征图。为了使网络更好地捕捉空间模式,对组成 ResNet101 的 4 个阶段(stage)进行改进,提出了 E-ResNet101;接着,在每个基本构建块中插入了分布感

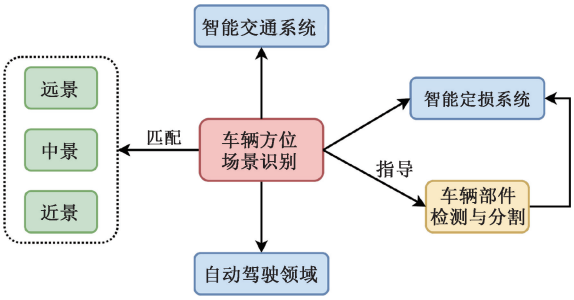


图 1　车辆方位场景识别应用场景

Fig. 1　Application scenarios of vehicle orientation scene recognition

知注意力模块 D-CBAM 以有效捕捉复杂的全局特征；为了聚焦关键区域以学习车辆方位与空间区域之间关系同时充分利用局部细节信息，在第 2 个阶段和第 4 个阶段后引入了全局与局部位置注意力模块 HGLP 形成最终的 OSMNet 方位场景匹配网络。基本特征图经过 OSMNet 提取特征后输出最终特征图；最后，对每个特征图进行平均池化，输出特征向量，通过全连接层，将特征向量映射到类别的概率分布上。

2.1　不同场景下的车辆方位

为了解决多部件检测或分割任务中，因为一些特征高

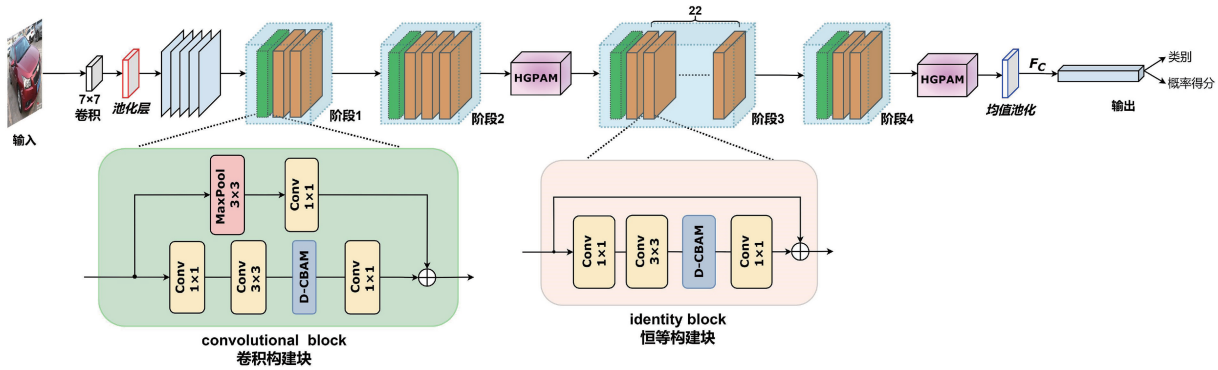


图 2　OSMNet 总体网络结构图

Fig. 2　Overall network architecture of OSMNet

度相似，造成的检测网络对左右类别、前后类别识别率低下的问题，推动检测任务局部近景与中景匹配工作，本文基于风险控制环境下车辆部件的拍摄视角，引入了人为因素，深入研究车辆部件场景化规则，提出了车辆多方位场景的概念。本文运用了改进的 ResNet101 方位场景匹配网络对含有 8 个方位场景的数据集进行训练。其中车辆的各方位场景如图 3 所示。

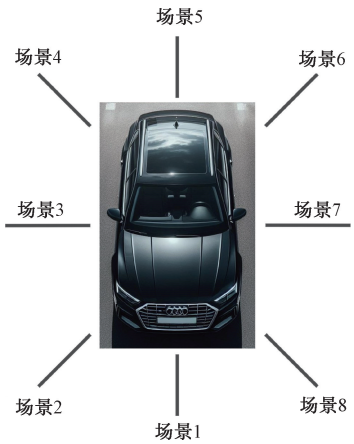


图 3　车辆多方位场景

Fig. 3　Multi-orientation vehicle scenes

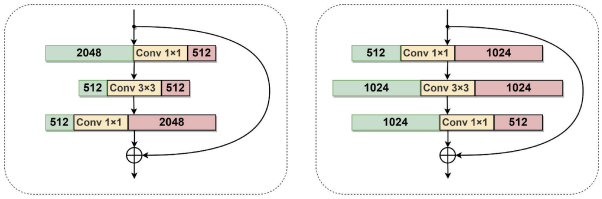
定损员拍摄得到的车辆图片主要是有远景、中景和近景 3 个视角。其中，远景中背景信息过多，车辆部分作为主要信息占比较少，而近景只包含了车辆局部的信息。为了学习车辆的方位关系，用所有的中景图片来训练和测试本文的网络。车辆方位通常分为 4 个主要方向：前、后、左、右，但当车辆方位介于这 4 个主要方向之间时，例如 45°角，将其归类到这 4 个类别变得具有挑战性，因此，将车辆方位分成 8 个场景。场景 1 是从车辆的正前方位进行拍摄的图片，需要包含车辆的两个前车灯；场景 2 是从车辆的右前方位进行拍摄的，需要包含整个右前车灯、车辆右侧局部信息（如右车门）、车辆前侧局部信息（如前保险杠），不包含或只包含小部分左前车灯、右后车灯；场景 3 是从车辆的正右方位进行拍摄的，同时包含车辆右侧前后车灯及其他部件；场景 4 是从车辆的右后方位进行拍摄的，需要包含整个右后尾灯、车辆右侧局部信息、车辆后侧局部信息（如行李箱盖），不包含或仅包含小部分左后尾灯、右前车灯。

2.2　方位场景分类网络架构 E-ResNet

文献[23]中指出传统的 ResNet 结构中的捷径分支是通过简单的恒等映射来实现的，这意味着它直接将输入特征映射到输出，而没有经过任何变换。因此，对于 convolutional block，原始的捷径分支会跳过大部分的特征图激活值，这会导致 1×1 卷积层无法选择有意义的激活值。这种情况下，1×1 卷积层可能无法充分利用特征图的信息，从而限制了网络的表达能力和泛化能力。为了使 ResNet101 架构更加适应本文的方位场景分类任务，减少信息丢失和信号干扰，防止网络过度关注某些局部特征，

而忽略其他特征,导致网络对于平移操作的敏感性增加,本文在捷径分支的 1×1 卷积之前引入一个 3×3 的最大池化(如图 2 所示)。通过在空间上对特征图进行下采样,使得 1×1 卷积可以考虑应该选择哪些激活值。这不仅减少了计算量,而且在一定程度上保留特征图的空间信息,从而使得网络对于平移操作更加鲁棒,提高了整体识别性能。

在所有瓶颈块(bottleneck block)中,主路都是由两个 1×1 卷积和 1 个 3×3 卷积组成的,第一个 1×1 卷积是为了降低通道数进而减少计算量,最后一个的 1×1 卷积是为了特征对齐,所以 3×3 卷积部分被限制了,然而只有它能有效学习特征模式,将其通道数减少虽然提高了计算速度,却降低了网络表达能力。于是本文采用组卷积[24]的思想,将原 ResNet101 中的瓶颈模块(如图 4(a)所示)用改进的瓶颈模块(如图 4(b)所示)代替,将重点放在了 3×3 卷积层上,使其具有更强的学习空间模式的能力。最终将改进基础构建块后的 ResNet101 称为 E-ResNet101(enhanced-resnet101)。



(a) Bottleneck block (b) ResGroup block

图 4 两种不同的构建块架构

Fig. 4 Two different building block architectures

此外,不同场景下的车辆图片包含了全局和局部范围内的一系列特征。以场景 1、场景 3 及场景 5 为例,局部特征关注图像特定关键点或区域的特性,通常能覆盖车辆的某个部分(如图 5 中的蓝色框所示),它主要是一些局部的纹理、边缘、角点和颜色特征,由于其跨度像素并不多,所以很容易对其进行提取。

除局部特征之外,某些特征具有全局跨度性并反映了图像的整体属性:主要体现在大尺度内容结构相似性和一致性(如图 5(a)中的红色矩形所示)、对称性(如图 5(b)中的绿色矩形所示)以及多尺度模式重复性(如图 5(c)中的红色矩形所示)、同一尺度的纹理相似性(如图 5(c)中的绿色矩形所示)。为了有效提取这些具有跨度性的特征,场景分类网络应当具备对全局图像进行理解的能力。

2.3 分布感知注意力模块 D-CBAM

在经典的 CBAM 注意力模块[25]中,通道注意力只使用了通道维度的平均值和最大值,无法有效捕捉到上述复杂的全局特征,本文对 CBAM 注意力中的通道注意力子模块做出改进,提出了一个新颖的分布感知注意力模块 D-CBAM,如图 6 所示,通过引入更多通道维度的信息,如通道维度的方差、斜度(skewness)等,将特征

(a) 场景1 (a) Scene 1 (b) 场景5 (b) Scene 5

(c) 场景3 (c) Scene 3

图 5 不同场景下的全局特征与局部特征示例

Fig. 5 Examples of global and local features in different scenes

图的分布情况考虑进去,进一步提高模型对全局特征的关注。

通道维度的方差衡量了特征图的通道特征值与均值之间的离散程度,即反映了特征图的通道分布范围。如果通道维度的方差较大,说明特征图的通道特征值较为分散,不同通道的特征值的重要性相对均衡,因此需要更加关注全局特征,以充分挖掘不同通道的信息。在车辆场景分类中,不同通道的特征能够反映出不同的视觉属性,通过关注全局特征,可以更好地捕捉到这些视觉属性的共性和变化规律,提高网络的分类性能;通道维度的斜度则反映了特征图在通道维度上的分布偏态情况。斜度描述了特征图的分布形态,如果斜度较大,说明特征图的分布偏斜,即某些通道的特征值的重要性较高,而其他通道的特征值的重要性较低,因此需要更加关注全局特征,以充分挖掘重要通道的信息,通过关注全局特征,有助于更好地捕捉到关键信息以提高分类精度。

通过式(1)来获取通道的方差,通过方差很容易得到标准差,然后,使用均值和标准差计算通道维度的斜度,图 6 中 Σ 表示得到斜度的运算过程,该过程如式(2)所示。

$$\mathrm{Var}(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{x}_i - \overline{\boldsymbol{x}_i})^2 \tag{1}$$

$$\mathrm{Skew}(\boldsymbol{x}) = \frac{1}{N}\sum_{i=1}^{N}\frac{(\boldsymbol{x}_i - \overline{\boldsymbol{x}_i})^3}{\boldsymbol{\sigma}_i^3 + \varepsilon} \tag{2}$$

其中,N 表示通道的数量,i 表示通道的索引,$\boldsymbol{x}_i$ 是第 i 个通道的特征图,$\overline{\boldsymbol{x}_i}$ 是第 i 个通道的平均值,$\boldsymbol{\sigma}_i$ 是第 i 个通道的标准差,ε 是为了防止分母为零而添加的一个小的数值(这里设置为 10^{-7})。

图 6 中的 c 是沿通道维度拼接张量的函数,利用其将通道维度的方差、斜度、均值信息进行组合,然后经过一次全连接与原来的通道注意力特征相加作为 CBAM 空间注意力子模块的输入之一。通过综合考虑均值、方差和斜

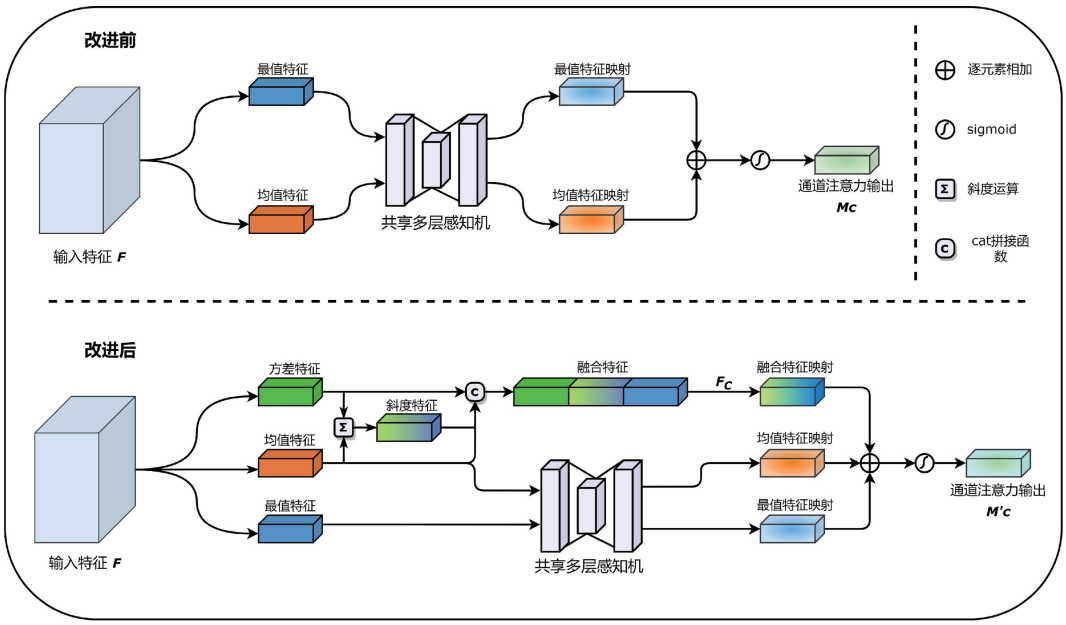


图 6　CBAM 通道注意力子模块改进前后对比

Fig. 6　Comparison of CBAM channel attention submodule before and after improvement

度，可以更好地反映出特征图的分布情况，进而提高对全局特征的关注能力。

2.4　全局与局部位置注意力模块 HGLP

为了进一步提高模型对于空间位置信息的感知能力，使模型可以充分考虑整张输入图像信息，在不同的方位场景中有效地聚焦关键区域以有效学习车辆的空间方位，从而更好的学习车辆方位与空间区域之间的关系，本文引入了全局注意力(global attention module，GAM)模块[26]并对其进行了改进。但是仅考虑空间信息而忽略局部特征，会导致模型对于车辆的边缘、纹理等细节和局部结构的学习不足，无法充分捕捉图像中重要的局部特征信息。例如一些 Transformer 架构，尽管其在处理全局信息方便表现出色[27-28]，但在处理局部信息方面可能存在一些不足，具体表现在数据体量不足且类间相似度较高的方位 3 与方位 7 中(2.5 小节对比实验)。

为了捕捉更充分的局部依赖关系，本文的改进方法是在原始 GAM 中的空间注意力子模块上并联了一个局部位置注意力模块(position attention module，PAM)[29]形成了一个双通道全局-局部位置注意力模块 HGLP，通过在局部特征上建模丰富的上下文依赖关系，使相似的特征相互增益，提高了类内紧凑性。

HGLP 的详细框架如图 7 所示，首先将通道注意力的输出 $\boldsymbol{F}_{\mathrm{C}} \in \mathbf{R}^{C\times H\times W}$ 通过一个卷积层生成 3 个新的特征 $\{\boldsymbol{A},\boldsymbol{B},\boldsymbol{C}\} \in \mathbf{R}^{C\times H\times W}$，接着将它们 reshape 为 $\{\boldsymbol{A}_{\mathrm{r}},\boldsymbol{B}_{\mathrm{r}},\boldsymbol{C}_{\mathrm{r}}\} \in \mathbf{R}^{C\times N}$，其中 $N = H \times W$ 是像素数。然后对前两个矩阵进行转置相乘的矩阵运算后经过 softmax 层得到空间注意力图 $\boldsymbol{E} \in \mathbf{R}^{N\times N}$：

$$\boldsymbol{E}_{ji} = \frac{\exp(\boldsymbol{A}_{ri}^{\mathrm{T}} \cdot \boldsymbol{B}_{rj})}{\sum_{i=1}^{N}\exp(\boldsymbol{A}_{ri}^{\mathrm{T}} \cdot \boldsymbol{B}_{rj})} \tag{3}$$

其中，E_{ji} 表示第 i 个位置对第 j 个位置的影响，衡量了两个位置的相关性，T 表示对矩阵进行转置操作。

最后将 $\boldsymbol{C}_{\mathrm{r}}$ 与空间注意力矩阵 $\boldsymbol{E}_{ji}$ 进行矩阵相乘并将运算后的结果进行 reshape 得到局部位置注意力支路的输出特征 $\boldsymbol{F}_{\mathrm{p}} \in \mathbf{R}^{C\times H\times W}$：

$$\boldsymbol{F}_{\mathrm{p}} = \mathrm{reshape}\left(\sum_{i=1}^{N}(\boldsymbol{C}_{\mathrm{r}} \cdot E_{ji})\right) \tag{4}$$

因为在反向传播过程中，梯度会通过多个层级进行传递，并且在每个层级中都可能会进行逐元素相乘。如果注意力图中的某些元素非常小或非常大，在反向传播过程中就可能会出现梯度消失或梯度爆炸问题；此外，逐元素相乘可能会引入噪声或干扰，这是由于注意力图中的每个元素都会对应地乘以特征图中的相应元素，从而使得特征图中的一些不必要的信息也被放大了。为了缓解这个问题，提高模型的鲁棒性，本文将局部位置注意力子模块输出的注意力图 $\boldsymbol{F}_{\mathrm{p}}$ 与空间注意力子模块输出的注意力图 $\boldsymbol{F}_{\mathrm{s}}$ 分别乘上一个可学习的参数后进行逐元素相加，得到最终的注意力特征图 $\boldsymbol{F}_{2}$：

$$\boldsymbol{F}_{2} = \alpha\boldsymbol{F}_{\mathrm{p}} \oplus (1-\alpha)\boldsymbol{F}_{\mathrm{s}} \tag{5}$$

其中，α 是一个可学习的尺度标量，用于调整局部位置注意力层的输出，其初始化的值设置为 0，使用优化器自动更新该参数，使之最小化损失函数从而实现在学习的过程中自适应地调节局部位置注意力和空间注意力的比重。这样的设计可以使网络首先依赖局部领域信息，并逐渐增加对非局部信息的利用[30]。本文推理过程中所使用的

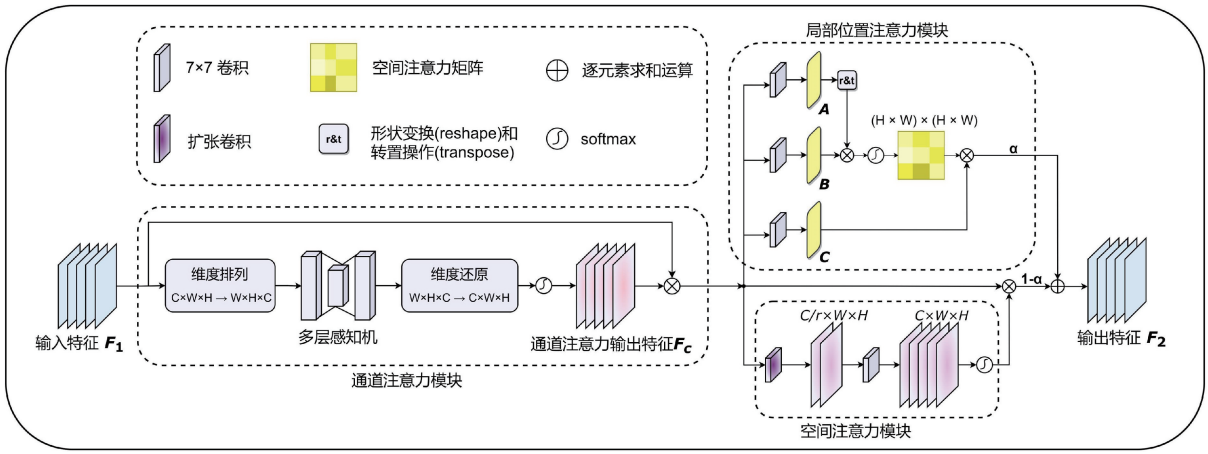


图 7 HGLP 框架图

Fig. 7 HGLP framework diagram

两个 HGLP 中的 α(从左到右)最终的优化结果分别为 0.215、0.152。⊕ 表示逐元素相加运算。

由于车辆方位图片在光照、尺度、视角上具有多样性,而卷积操作带来的局部感受野会使相同方位场景对应的特征出现差异。这些差异导致了类内的不一致性从而影响识别效果。于是将空间注意力子模块中的第一个 7×7 卷积替换成扩张卷积,在特征图被送入空间注意力子模块之前,并对其进行填充(padding)操作以保持输出的特征图的大小保持不变。

PAM 是 HGLP 的关键组成部分,实现了自适应地捕获输入特征图中不同位置之间的依赖关系。通过这种方式,HGLP 显著提高了模型对于空间位置信息的感知能力,增强了模型对图像中局部信息的提取能力,尤其是在处理诸如车辆边界和轮廓等细节方面表现出色。

局部特征对于场景理解至关重要。通过 HGLP,本文在提取全局特征的同时,能够充分利用局部特征,提高了类内紧凑性,从而更好地理解和识别复杂的车辆方位场景,特别是车辆左右镜像对称的情况。这种双通道的位置全局注意力机制有助于综合利用全局和局部信息,提高模型在场景识别任务中的性能。

3 实验结果及分析

3.1 数据集来源与实验环境

本文在一家保险公司提供的车辆图片中选取了 8 类车辆方位场景作为实验对象,参考 ImageNet 格式数据集的构建方法,构建了车辆方位场景数据集,其中训练集与测试集样本图像分别为 25 386 和 6 347 张,数量比例为 4∶1,数据集暂未公开,数据集分布情况如表 1 所示。对文中出现的图像内的车牌等隐私信息进行了马赛克处理。本文所述模型采用 NVIDIA 2080Ti 专业加速卡进行训练与测试;采用的操作系统为 Ubuntu16.04.7 LTS,利用 CUDA11.3 加速训练;使用的计算机语言 Python3.8.12,网络开发框架为 PyTorch,版本为 1.11.0。在训练阶段,将 batchsize 设为 16,使用了 Adam 优化器,初始学习率为 0.000 1,权重衰减为 10^{-4},动量衰减为 0.001。同时,使用了余弦退火学习率调度器(cosine annealing,LR),这种学习率调度方式可以在训练过程中更好地平衡学习速度和模型收敛速度,最小学习率为 10^{-6}。代码迭代训练了 150 个 epoch。本文模型适用于中景单部车辆或是多车辆中主体车辆突出的车辆方位场景数据集,对于主体车辆占比较小的远景和包含多部主次车辆相近的数据集,场景分类的效果提升可能并不明显。

表 1 数据集分布情况

Table 1 Distribution of dataset

方位场景	图像数量/张
场景 1(正前方位)	4 974
场景 2(右前方位)	6 050
场景 3(右方位)	795
场景 4(右后方位)	4 729
场景 5(正后方位)	3 245
场景 6(左后方位)	4 814
场景 7(左方位)	579
场景 8(左前方位)	6 547
总计	31 733

3.2 评价指标

为了验证方位场景分类网络的有效性,本文以 ResNet101 作为基线模型,并使用目前分类任务中常用的评价指标准确率(accuracy,Acc)对模型进行总体的评估,准确率 Acc 表示对测试集进行分类后,在某个类别中被正确分类的比例。准确率的计算方式为:

$$\mathrm{Acc}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}+\mathrm{TN}} \tag{6}$$

表 2 是根据方位场景分类结果建立的混淆矩阵，介绍了上述评价指标中各个字母代表的含义。

表 2　方位场景分类混淆矩阵

Table 2　Confusion matrix for orientation scene classification

真实结果	预测结果	
	属于方位 i	不属于方位 i
属于方位 i	TP	FN
不属于方位 i	FP	TN

为进一步探究改进后的方位场景分类网络的计算复杂度和资源消耗，本文采用了两个关键指标：参数量(Parameters)和浮点运算量(FLOPs)。参数量指模型中所有可训练的权重和偏置的总数量，是衡量模型复杂度的重要指标。浮点运算量(FLOPs)用于评估模型在推理和训练过程中的计算需求。

3.3　消融实验

1) 模块消融实验

为了证明本文模型中每个模块的有效性，本文设计了两组消融实验：(1)证明改进的 E-ResNet101 架构的有效性，表 3 第 1 列为是否采用网络 E-ResNet101 架构；(2)证明 D-CBAM 模块的有效性：在基线模型的基础上，在每个构建块中增加 D-CBAM 模块；(3)证明 HGLP 模块的有效性：在基线模型的基础上，只加入 HGLP 模块进行特征提取学习，经过实验发现在网络的第二个和最后一个阶段后插入 HGLP 块达到最好的效果(详见插入位置消融实验)。最终检测结果如表 3 所示。

表 3　消融实验结果

Table 3　Ablation experiment results

方法	E-ResNet	D-CBAM	HGLP	Acc/%
Baseline				90.03
	√			91.44
		√		93.57
			√	94.25
本文	√	√	√	**95.49**

从表 3 可以看出，改进基本构建块的 E-ResNet101 相较于基线模型的识别效果上有一个整体的提升。单独在网络基本构建块中增加 D-CBAM 模块能显著提高模型的检测准确率，与基线模型相比，识别准确率提高 3.54%，说明 D-CBAM 模块在增强通道维度信息的同时，能够更全面地捕捉特征图的分布情况，从而提高对全局特征的关注能力。单独使用 HGLP 模块对实验结果也有提高，与基线模型相比准确率提高 4.22%，这是由于 HGLP 模块成功捕捉了输入特征图中不同位置之间的依赖关系，特别是在处理车辆边界和轮廓等细节方面表现出色，从而提高了对空间位置信息的感知能力，增强了模型对车辆局部信息的提取能力。而在 E-ResNet101 架构上同时使用 D-CBAM 模块和 HGLP 模块时提升效果最好，与基线模型相比准确率提高 5.46%，这是因为 D-CBAM 模块在全局特征的关注上表现出色，而 HGLP 模块则更专注于空间位置信息的感知，二者在不同方面的优势能够互补，共同作用于网络，在场景识别任务中取得了最佳效果。

为了研究插入位置对识别精度的影响，本文在基线模型上对 HGLP 模块插入位置进行了消融实验，实验结果如表 4 所示。其中第一列为实验组数。由前 4 组实验可以看出，当只加入一个 HGLP 模块时，在最后一个阶段后插入取得的效果最好，这可能是因为最后一个阶段是网络的高级特征提取层，HGLP 模块的引入可以使网络充分利用不同视角和角度下的不变性特征，使得模型更具稳定性和泛化能力。此外在高级特征中更容易对部分和整体的关系进行建模。通过过在最后一个阶段引入 HGLP 模块，能够更好地理解图像中不同部分的语义关系，从而提高对场景的整体理解能力。当把 HGLP 模块插入网络中多个位置时，发现将其在第二个和最后一个阶段后插入后网络表现最佳，这可能是由于在中间层引入了 HGLP 模块后帮助模型更好地整合全局和局部信息，提高了网络对局部的感知能力。第 10 组实验结果可以看出，在第 3 第 4 个阶段之后插入 HGLP 模块后精度下降了，甚至低于只在一个位置引入该模块时的效果，这可能是由于信息在经过多个层传递时会逐渐丢失一些重要的局部信息，此时在网络较后的位置引入多个 HGLP 模块可能引入了过多全局信息，在特征表达较为抽象的情况下导致信息干扰，从而影响了识别精度。综上本文模型最终在第 2 和第 4 个阶段后插入 HGLP 模块。

表 4　插入位置消融实验

Table 4　Insertion position ablation experiment

位置	阶段 1	阶段 2	阶段 3	阶段 4	Acc/%
1	√				92.17
2		√			93.84
3			√		93.52
4				√	93.98
5	√	√			93.91
6	√		√		93.78
7	√			√	94.17
8		√	√		94.04
9		√		√	**94.25**
10			√	√	93.76

2) 学习率超参数的影响

为研究初始学习率对模型性能的影响，本文进行了学

习率消融实验,分别测试了初始学习率为 0.01、0.001 和 0.000 1 时本文模型准确率变化情况,实验结果如图 8 所示。

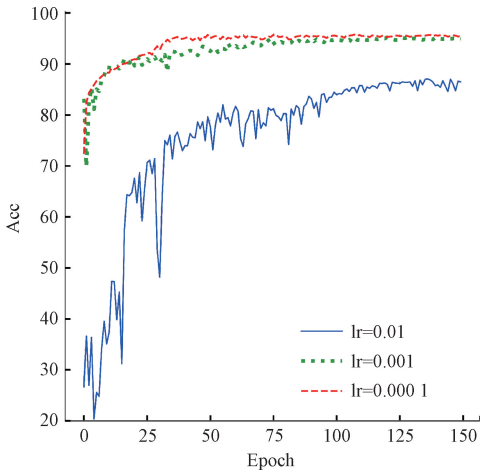


图 8 不同学习率下的 Acc 数据曲线

Fig. 8 Accuracy curves with different learning rates

由图 8 可知,当学习率设为 0.01 时,模型在训练过程中波动较大,最终的准确率相对较低,说明学习率过大容易引起模型训练的不稳定性,可能会陷入局部最优解。当学习率设为 0.000 1 时,模型在训练过程中的收敛速度适中,能够较好地在初期和后期的训练中保持平稳的优化过程,最终达到了相较学习率为 0.001 时更高的准确率,这表明较小的学习率有助于维持训练的稳定性,同时允许模型更精细的调整参数。

3.4 模型复杂度与运算效率分析

为了探究所提出的 OSMNet 网络的复杂度和运算效率,本文设计评估实验,实验结果如表 5 所示。实验统计并比较了本文模型与基准模型在整个数据集上的训练时间以及在参数数量、浮点运算量上的表现。接着随机选择测试集中的 500 张图像进行推理测试,计算得出每张图片的平均推理时间。

表 5 模型复杂度及运算效率评估实验

Table 5 Model complexity and computational efficiency evaluation experiment

方法	Param/M	FLOPs/G	训练时间/h	单张图片推理时间/s
Baseline	42.52	15.63	13.49	0.31
OSMNet(本文)	49.09	17.38	13.92	0.33

综合表 5 的实验结果可以看出,OSMNet 模型的参数数量和浮点运算量略高于基准模型,但其训练时间和单张图片推理时间并未显著增加,这表明本文算法的运算效率能够满足实际应用需求。

3.5 与先进模型对比实验

为了验证本文模型的先进性,本文在车辆多方位场景数据集上实现了多种分类模型并进行对比,包括一些传统的卷积神经网络如 Vgg16[31]、Resnet101[32]、Efficientnetv2[33]、Regnet[34]、ConvNeXt[35] 和一些流行的 Transformer 方法如 Vision Transformer、Swin Transformer[36]、MobileViT[37]、Maxvit[38],测试结果如表 6 所示。

表 6 与先进算法性能对比

Table 6 Performance comparison with State-of-the-Art algorithms %

模型	Acc	场景 1	场景 2	场景 3	场景 4	场景 5	场景 6	场景 7	场景 8
Baseline	90.03	92.86	89.26	86.79	91.23	93.53	87.54	87.93	88.39
Vgg16	90.64	89.15	94.05	67.92	92.28	87.67	94.39	55.17	92.06
Efficientnetv2	91.10	87.74	94.05	73.58	91.97	87.37	95.43	56.03	94.19
Regnet	90.36	91.56	93.64	71.07	90.49	88.14	90.65	68.10	91.52
ConvNeXt	94.45	**93.07**	96.20	86.79	95.24	91.22	95.64	81.03	96.18
Vision_Transformer	86.18	88.44	89.17	67.30	84.99	84.13	88.47	56.90	86.78
Swin_Transformer	91.35	90.85	93.80	77.36	91.75	86.13	95.74	73.28	91.83
MobileViT	88.91	82.61	95.29	57.23	86.15	**93.84**	88.37	69.83	93.28
Maxvit	90.24	88.94	92.07	70.44	90.27	89.52	90.65	63.79	94.35
OSMNet(本文)	**95.49**	90.55	**97.52**	**92.45**	**97.46**	91.99	**96.05**	**89.66**	**98.17**

注:加粗字体为每列最优值。

从表 6 中可以看出,基线模型的平均准确率指标为 90.03%,本文模型的平均准确率指标为 95.49%,与基线模型相比,本文模型的识别效果提高了 5.46%且大部分类别的识别效果都有了显著提高。另外与其他先进模型进行对比,本文模型在绝大多数场景中表现出更好的性能,对场景 2 的识别也接近最高值,一些模型如 ConvNeXt、MobileViT 对场景 1 或场景 5 的识别效果要优于本文的模型,这可能是由于他们能够学习多尺度特征,可以兼顾细节和全局信息,增强了网络的表达能力。尽管如此,它们对数据量小且相似度较高的场景 3 和场景 7 识别效果的改

善并不明显。

表 6 中的测试结果显示，本文模型在场景 3、场景 4、场景 6、场景 7、场景 8 这些场景上是所有模型中识别精度最高的，这表明由于双通道全局与局部位置注意力模块的引入，本文模型能够更准确的捕捉全局和局部信息，并且整合了局部特征和它们的全局依赖关系，提高了类内紧凑性。而场景 3 和场景 7 的识别效果提高也证明了本文模型能够一定程度上应对数据不平衡的问题从而提升了总体的识别效果。升了总体的识别效果。

此外，与基线模型的测试结果相比，本文最终改进后的模型对于绝大多数场景的识别效果都有一定的提升。受到一些主客观因素的影响，本文的数据集中场景 3 和场景 7 数量较少，因此基线算法对这两个类别的识别效果较差，但是本文改进后的模型很好的解决了这个问题，显著提升了这两个场景的识别精度。这是因为本文的 D-CBAM 和 HGLP 模块能够更好地捕捉全局和局部信息，特别是在类别数据较少的情况下，帮助模型更好地理解和识别这些场景，提高了分类的准确性。

3.6　模型可解释性分析与检测结果可视化

为了分析本文模型的检测效果同时更清晰地理解模型的决策过程，使用显著性图（saliency map）[39] 对各个方位场景的部分测试样本进行可解释性分析，可视化结果如图 9(a)～(h)所示。显著性图通过将模型的输出梯度传递回输入图像，可以获取关于每个像素对于模型输出的贡献程度的信息。这些贡献值被用来构建显著性图，其中较高的值表示模型更关注的区域。传统类激活映射（class activation map，CAM）[40] 方法主要关注最终分类层的激活，可能忽略其他层的信息，与 CAM 不同的是，显著性图生成过程依赖于整个网络多通道的梯度信息，能更好地捕捉模型对输入图像整体的关注，因此能够提供更全局的解释。此外，显著性图生成不依赖于具体的网络结构，具有模型无关性，适用于不同结构的神经网络。

(a) 场景1
(a) Scene 1

(b) 场景2
(b) Scene 2

(c) 场景3
(c) Scene 3

(d) 场景4
(d) Scene 4

(e) 场景5
(e) Scene 5

(f) 场景6
(f) Scene 6

(g) 场景7
(g) Scene 7

(h) 场景8
(h) Scene 8

图 9　显著性图可视化结果

Fig. 9　Visualization results of saliency maps

由图 9 可以看出，热图在车辆区域上更加明亮，意味着本文的网络主要依靠车辆前景来进行决策，但是热图不是全部集中在车辆上，仍可以看到一些分散，这是因为网络考虑了车辆主体与其所处环境的相对位置关系，即周围的背景也是预测方位的重要因素。

为了定性分析本文模型的识别效果，图 10 以对称的两组场景为例，显示了本文所述模型与基线模型的输出结果对比。输入图 10(a)、(b)的 4 张图片对应的真值分别为

场景 2、场景 8 以及场景 3、和场景 7。在基线模型的识别结果中,受到结构对称性和特征相似性的影响,场景 3 和场景 7 均未被正确识别出来。并且其他两类识别精度也有待进一步提高。而在本文模型识别结果图 10(b)中,由于局部位置注意力的引入和全局注意力的改进,模型能够精确识别出车辆的方位场景,从图 10(a)和(b)的对比中可以看出,本文模型对于相似性高的场景 3 和场景 7 也能够得到较好的识别效果。

(a) 基线模型输出结果
(a) The output results of baseline model

(b) 本文模型输出结果
(b) The output results of our model

图 10 改进前后方位场景网络输出结果可视化

Fig. 10 Visualization of output results of orientation scene network before and after improvement

4 结 论

针对当前车辆部件检测和分割任务中存在因特征相似而带来的左右类别,前后类别识别混淆等问题,本文提出了一种车辆方位场景识别方法。

首先改进了 ResNet101 网络中的基本构建块提出了 E-ResNet101,使网络更好地捕捉空间模式,提高整体识别性能。然后引入分布感知注意力模块 D-CBAM,将通道维度的方差、斜度等信息考虑在内,使模型在通道维度上更全面地关注特征图的分布情况,从而更有效地学习全局特征。接着引入了全局与局部位置注意力模块 HGLP,进一步整合了局部特征与它们的全局依赖关系,充分利用局部特征,提高了类内紧凑性。本文融入以上两个注意力形成了最终的方位场景匹配网络 OSMNet。实验结果表明,本文模型在车辆方位数据集上的性能明显优于基线模型以及其他先进模型。通过 D-CBAM 和 HGLP 两个模块的引入,模型在绝大部分场景的识别精度都取得了显著提升。尤其是在数据量较少且相似度较高的场景 3 和场景 7 中,本文模型相较于其他模型表现更为出色。本文基于先进的注意力机制充分捕捉局部全局信息聚焦有效区域来改善车辆方位场景识别效果,为该领域提供了新的思路和方法。

对车辆方位场景识别这一任务的研究可以进一步应用于更复杂的车辆多部件检测领域,本文训练出的方位场景识别模型提供了有力的先验知识,为多模型融合及嵌入提供了可靠的基础,为车辆部件检测任务提供了逻辑上的指导和修正。这一研究拓宽了对车辆智能识别的认识并且可以拓展到更为细致和复杂的应用领域如场景分割、车辆定损和自动驾驶,也为构建更智能、更安全、更高效的交通系统提供了有力支持。

下一步的工作可以考虑引入更多先进的视觉特征提取方法,例如学习多尺度特征,提升车辆在远景情况下的识别性能。另外可以考虑将本文模型与车辆部件检测模型进行融合,提升车辆检测或分割性能。对于特定场景下的数据不足问题,可以探索数据增强和迁移学习等方法,以提高模型的泛化能力。

参考文献

[1] 刘智,马社强. 高峰期城市道路交通事故严重程度影响因素分析[J]. 中国人民公安大学学报(自然科学版),2024,30(1):44-50.
LIU ZH, MA SH Q. Analysis of factors influencing the severity of urban road traffic accidents during peak period [J]. Journal of People's Public Security University of China(Science and Technology), 2024, 30(1):44-50.

[2] 胡立伟,雷国庆,赵雪亭,等. 城市拥塞环境下交通事故风险传播及控制研究[J]. 安全与环境学报,2023,23(8):2809-2818.

HU L W, LEI G Q, ZHAO X T, et al. Study on the traffic accident risk propagation and control in urban congestion environment[J]. Journal of Safety and Environment, 2023,23(8):2809-2818.

[3] 谢东升. 基于深度学习的车辆智能定损算法研究[D]. 天津:天津大学, 2019.
XIE D SH. Research on vehicle intelligent damage location algorithm based on deep learning [D]. Tianjin: Tianjin University, 2019.

[4] WU F, JIN G, GAO M, et al. Helmet detection based on improved YOLO V3 deep model[C]. 2019 IEEE 16th International Conference on Networking, Sensing and Control(ICNSC),2019: 363-368.

[5] BO Y, HUAN Q, HUAN X, et al. Helmet detection under the power construction scene based on image analysis[C]. 2019 IEEE 7th international conference on computer science and network technology(ICCSNT), 2019: 67-71.

[6] 卫策, 吕进, 曲晨阳. 改进 YOLOv5s 的复杂交通场景下目标检测算法[J]. 电子测量技术, 2024, 47(2): 121-130.
WEI C, LYU J, QU CH Y. Improved object detection algorithm for complex traffic scenes in YOLOv5s[J]. Electronic Measurement Technology, 2024, 47(2): 121-130.

[7] 朱新源, 任劼, 章为川. 基于注意力机制的双度量小样本图像分类算法[J]. 国外电子测量技术, 2022, 41(8): 34-38.
ZHU X Y, REN J, ZHANG W CH. Attention based bi-metric network for few-shot image classification[J]. Foreign Electronic Measurement Technology, 2022, 41(8): 34-38.

[8] KUHN D, MOREIRA V. BRCARS: A dataset for fine-grained classification of car images[C]. 2021 34th SIBGRAPI Conference on Graphics, Patterns and Images(SIBGRAPI),2021: 231-238.

[9] 李利荣, 丁江, 梅冰, 等. 基于像素注意力特征融合的城市街景语义分割算法研究[J]. 电子测量技术, 2023, 46(20): 184-190.
LI L R, DING J, MEI B, et al. Semantic segmentation method for urban street scenes based on pixel attention feature fusion [J]. Electronic Measurement Technology, 2023, 46(20): 184-190.

[10] 项建弘, 徐昊. 基于深度学习的图像语义分割算法研究[J]. 计算机应用研究, 2020, 37(S2): 316-317,320.
XIANG J H, XU H. Research on image semantic segmentation algorithm based on deep learning[J]. Application Research of Computers, 2020, 37(S2): 316-317,320.

[11] 樊博, 高玮玮, 单明陶, 等. 融合注意力机制与重影特征映射的无人机交通场景目标轻量级语义分割[J]. 电子测量与仪器学报, 2023, 37(3): 21-28.
FAN B, GAO W W, SHAN M T, et al. Lightweight semantic segmentation of UAV traffic scene objects combining attention mechanism and ghost feature mapping[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(3): 21-28.

[12] 伍锡如, 邱涛涛, 王耀南. 改进 Mask R-CNN 的交通场景多目标快速检测与分割[J]. 仪器仪表学报, 2021, 42(7): 242-249.
WU X R, QIU T T, WANG Y N. Multi-object detection and segmentation for traffic scene based on improved Mask R-CNN [J]. Chinese Journal of Scientific Instrument, 2021, 42(7): 242-249.

[13] ZHAI Y, CHEN N, ZHANG Z, et al. SU-VPDN: A scene understanding method for vehicle part detection[J]. Engineering Applications of Artificial Intelligence, 2024, 132: 107956.

[14] 李浩,鲍鸿,詹瑞典.融合多级注意力机制和信息融合的车型识别[J].电子测量技术,2023,46(5):164-171.
LI H, BAO H, ZHAN R D. Research on vehicle type recognition based on multilevel attention mechanism and information fusion[J]. Electronic Measurement Technology, 2023,46(5):164-171.

[15] 杨东, 李丹. 基于 BoTNet 的车辆分类实现[J]. 电子测试, 2021,(24): 57-59.
YANG D, LI D. Implementation of vehicle classification based on BoTNet[J]. Electronic Test, 2021,(24): 57-59.

[16] AUDEBERT N, LE SAUX B, LEFÈVRE S. Segment-before-detect: Vehicle detection and classification through semantic segmentation of aerial images[J]. Remote Sensing, 2017, 9(4): 368.

[17] BAI S, LIU Z, YAO C. Classify vehicles in traffic scene images with deformable part-based models[J]. Machine Vision and Applications, 2018, 29: 393-403.

[18] LIU K, MATTYUS G. Fast multiclass vehicle detection on aerial images[J]. IEEE Geoscience and Remote Sensing Letters, 2015, 12(9): 1938-1942.

[19] CHEN C, GONG W, HU Y, et al. Learning oriented region-based convolutional neural networks for building detection in satellite remote sensing images[J]. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2017, 42: 461-464.

[20] DENG Z, SUN H, ZHOU S, et al. Toward fast and accurate vehicle detection in aerial images using coupled

region-based convolutional neural networks[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(8): 3652-3664.

[21] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale[J]. ArXiv preprint arXiv:2010.11929, 2020.

[22] ZHU L, WANG X, KE Z, et al. BiFormer: Vision transformer with bi-level routing attention[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 10323-10333.

[23] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1492-1500.

[24] DUTA I C, LIU L, ZHU F, et al. Improved residual networks for image and video recognition[C]. 2020 25th International Conference on Pattern Recognition (ICPR), 2021: 9415-9422.

[25] WOO S, PARK J, LEE J, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.

[26] LIU Y, SHAO Z, HOFFMANN N. Global attention mechanism: Retain information to enhance channel-spatial interactions[J]. ArXiv preprint arXiv: 2112.05561, 2021.

[27] WU B, XU C, DAI X, et al. Visual transformers: Token-based image representation and processing for computer vision[J]. ArXiv preprint arXiv: 2006.03677, 2020.

[28] ALMALIK F, YAQUB M, NANDAKUMAR K. Self-ensembling vision transformer(sevit) for robust medical image classification[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2022: 376-386.

[29] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3146-3154.

[30] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks[C]. International Conference on Machine Learning, 2019: 7354-7363.

[31] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. ArXiv preprint arXiv: 1409.1556, 2014.

[32] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[33] TAN M, LE Q. Efficientnetv2: Smaller models and faster training[C]. International Conference on Machine Learning, 2021: 10096-10106.

[34] RADOSAVOVIC I, KOSARAJU R, GIRSHICK R, et al. Designing network design spaces[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10428-10436.

[35] LIU Z, MAO H, WU C, et al. A convnet for the 2020s[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11976-11986.

[36] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 10012-10022.

[37] MEHTA S, RASTEGARI M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer[J]. ArXiv preprint arXiv: 2110.02178, 2021.

[38] TU Z, TALRBI H, ZHANG H, et al. Maxvit: Multi-axis vision transformer[C]. European Conference on Computer Vision, 2022: 459-479.

[39] BROCKI L, CHUNG N. Input bias in rectified gradients and modified saliency maps[C]. 2021 IEEE International Conference on Big Data and Smart Computing(BigComp). IEEE, 2021: 148-151.

[40] ZHOU B, KHOSLA A, LAPEDRIZA A, et al. Learning deep features for discriminative localization[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2921-2929.

作者简介

翟永杰,博士,教授,主要从事电力视觉、模式识别与人工智能、计算机控制系统方面的研究。

E-mail:zhaiyongjie@ncepu.edu.cn

刘璇,硕士研究生,主要从事电力视觉与人工智能方面的研究。

E-mail:liuxuan2@ncepu.edu.cn

王新颖(通信作者),博士在读,讲师,主要从事电力视觉、知识图谱方面的研究。

E-mail:wangxinying@ncepu.edu.cn