引用格式：AI Qinglin, ZHANG Junrui, WU Feiqing. AF-ICNet Semantic Segmentation Method for Unstructured Scenes Based on Small Target Category Attention Mechanism and Feature Fusion[J]. Acta Photonica Sinica, 2023, 52(1):0110001
艾青林,张俊瑞,吴飞青.基于小目标类别注意力机制与特征融合的AF-ICNet非结构化场景语义分割方法[J].光子学报,2023,52(1):0110001

基于小目标类别注意力机制与特征融合的AF-ICNet非结构化场景语义分割方法

艾青林[1],张俊瑞[1],吴飞青[2]
(1 浙江工业大学 特种装备制造与先进加工技术教育部/浙江省重点实验室,杭州 310023)
(2 浙大宁波理工学院 信息科学与工程学院,宁波 315100)

摘　要:针对非结构化道路分割难度大、小目标检测精度较低等问题,构建基于小目标类别注意力机制与特征融合的AF-ICNet轻量级实时语义分割网络。采用空洞空间卷积池化金字塔融合不同尺度特征感受野以增强网络的全局感知能力。嵌入CA注意力机制,建立通道信息和空间位置信息以增强网络对非结构化道路小目标类别语义特征的提取能力。针对类别分布不均衡问题,改进权重交叉熵损失函数。利用AF-ICNet模型对Cityscapes与IDD数据集进行训练,在Cityscapes测试图像中分割的MIoU达到了71.5%,在IDD测试图像中分割的MIoU达到了62.5%。搭建实验测试系统进行实景测试,测试结果表明,AF-ICNet有效提升了非结构化道路及小目标类别的分割精度,并满足测试的实时性要求。

关键词:小目标类别语义分割;AF-ICNet;CA注意力机制;空洞空间卷积池化金字塔;损失函数
中图分类号:TP391　　**文献标识码**:A　　**doi**:10.3788/gzxb20235201.0110001

0　引言

非结构化道路可行驶区域检测技术是计算机视觉与自动驾驶研究领域的热点之一。在实际道路驾驶、大型工程作业以及机器人野外工作场景中,存在大量非结构化道路场景。非结构化道路相比于结构化道路,具有道路与周边颜色差异较小,道路目标物种类较多、信息复杂等特点。使用传统方法检测非结构化道路区域,存在检测精度低、实时性差、小目标检测效果差等问题[1]。由于小目标障碍物和行人会对道路可行驶区域检测造成严重干扰,而小目标在图像中分辨率低、携带的信息较少而导致其特征表征能力较差,大型类别的主导也容易导致小目标类别被忽视[2],因此如何提升非结构化道路小目标检测能力是目前亟待解决的技术问题。

近年来,基于神经网络的道路图像分割方法得到了快速发展,比如FCN[3]、SegNet[4]、U-Net[5]、DeepLab[6]等经典网络。其中,张凯航等提出基于SegNet的非结构道路可行驶区域语义分割[7],龚志力等提出基于改进DeepLabV3+的非结构化道路分割方法[8],这些分割方法虽然可以有效分割非结构化道路区域,但实时性较差,其中部分网络的实际分割速度FPS小于1帧/s,无法部署在实时性要求较高的系统中。而轻量型网络ENet[9]、ICNet[10]等虽然实时性较好,但对小目标物体的分割效果较差。为了兼顾网络的实时性与非结构化道路中小目标物体分割精度,本文提出基于小目标类别注意力机制与特征融合的AF-ICNet(Attention and Feature fusion ICNet)非结构化道路场景语义分割方法,该方法主要内容有1)修改网络特征融合,使用空洞卷积特征金字塔替代池化以减少池化对网络特征提取的影响,并使用不同尺度的特征融合以扩大网络感受野,提高整体网络对图像的分割精度。2)基于小目标类别建立CA注意力机制(Coordinate attention)模块,形成通道和空间注意力信息的长期依赖关系,以提高网络对不同复杂道路及小目标物体语

基金项目:国家自然科学基金(No.52075488),浙江省自然科学基金(No.LY20E050023)
第一作者(通讯作者):艾青林,aqlaql@163.com
收稿日期:2022-06-29;**录用日期**:2022-08-18

义分割精度。3)建立融合权重的损失函数,针对非结构化道路的样本类别不均衡问题,对图像样本类别赋予不同的权重,以进一步提升网络对图像中小目标类别的分割精度。

1 基于AF-ICNet的非结构化道路语义分割方法

1.1 基于小目标注意力与特征融合的AF结构

1.1.1 扩大感受野的空洞空间卷积特征融合

在ICNet中,考虑到此前的基于FCN的语义分割网络无法有效的融合全局特征信息,网络中使用了空间金字塔池化(Pyramid Pooling Module, PPM)模块,用于聚合不同区域的特征信息进而获取全局的特征信息。PPM模块结构如图1所示。PPM通过融合上下层不同大小的特征信息,实现了多尺度的特征融合。PPM虽然增强了特征表征能力,但是其中大量使用的GAP全局池化操作,在增大感受野的同时也降低了图像分辨率,为保证网络可以正常训练,需要统一图像尺寸,这会导致上采样恢复至输入大小时,池化造成的细节信息的丢失无法恢复。

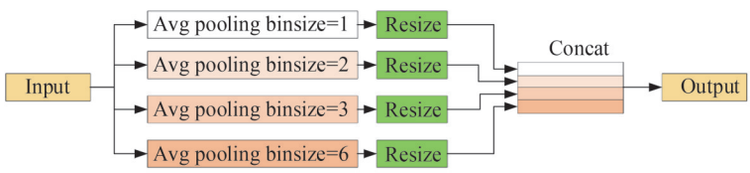

图1 金字塔池化PPM结构

Fig.1 Pyramid pooling module structure

针对此问题,本文参考DeeplabV2[11]模型,修改原网络中的特征融合模块,为减少大量池化的影响,替换网络中的PPM,使用空洞空间卷积池化金字塔(Atrous Spatial Pyramid Pooling, ASPP)。改进后特征融合如图2(a)所示,网络使用空洞卷积替代了大多数的池化。采样率的不同会对网络分割产生影响,经过实验分析,本文使用了效果最好的四种不同采样率的卷积,分别为一个1×1卷积和三个3×3的采样率为6、12、18的空洞卷积,以有效的获取多尺度信息。三个带padding的空洞卷积,使得卷积的感受野分别扩大到了23×23、47×47、71×71,如图2(b)所示,不同的感受野卷积结果进行Concat操作,小感受野便于网络获取精细的特征信息,大感受野便于获取目标在图像中的整体空间信息,通过大小不同感知野的特征融合,提升了网络对非结构化道路的整体分割效果。

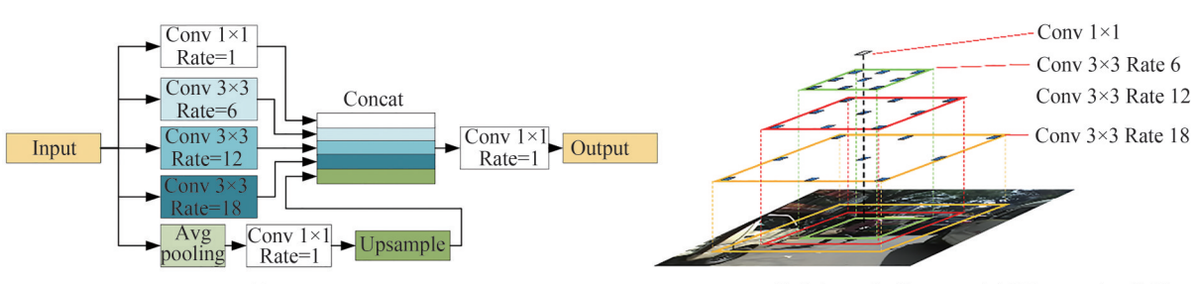

图2 空洞空间卷积池化金字塔ASPP结构

Fig.2 Atrous spatial pyramid pooling structure

1.1.2 增强小目标分割的CA注意力机制

在非结构化道路环境中,注意到非结构化道路中的小目标类别数量较多,为进一步提升网络对小目标物体信息的获取能力,在网络中引入注意力机制。通道注意力经典网络SENet[12]仅仅考虑通道之间的关系来衡量每个通道的重要性,往往会忽略目标的位置信息。增加了空间注意力机制的CBAM[13]中通道注意力信息和空间注意力信息是相互独立的,相较于SENet提升有限。为此,本文研究一种融合空间与通道信息的新型注意力方法——CA注意力机制(Coordinate Attention)。将其与ICNet网络语义分割结合,在保证实

时性的基础上，进一步实现小目标类别分割精度的提升[14]。

将全局池化进行两个维度的分解，即使用$(H,1)$或$(1,W)$的池化核，使其分别沿着水平坐标与垂直坐标方向，对每个通道进行编码操作，即可将上述全局池化编码公式分解为

$$z_c^h(h)=\frac{1}{W}\sum_{0\leqslant i\leqslant W}x_c(h,i) \tag{1}$$

$$z_c^h(w)=\frac{1}{H}\sum_{0\leqslant j\leqslant H}x_c(j,w) \tag{2}$$

基于以上生成的两个特征，进一步将两个特征图进行结合操作，然后使用1×1的卷积，对其进行变换操作

$$f=\delta(F_1[z^h,z^w]) \tag{3}$$

式中，F_1为1×1的卷积变换函数，方括号表示沿空间维度的结合操作，δ为非线性激活函数h-Swish。将中间特征映射f分解成两个单独的张量$f^h\in R^{C/r\times H}$和$f^w\in R^{C/r\times W}$，r为模块大小缩减率。分别将f^h与f^w变换为具有相同通道数的张量，并经过sigmoid激活，得到的g^h与g^w为

$$g^h=\sigma(F_h(f^h)) \tag{4}$$

$$g^w=\sigma(F_w(f^w)) \tag{5}$$

最后得到的注意力模块输出为

$$y_c(i,j)=x_c(i,j)\times g_c^h(i)\times g_c^w(i) \tag{6}$$

在本文网络中，CA注意力机制模块如图3所示。将输入分别沿X和Y方向进行平均池化，经过维度转换后，将输入的$C\times H\times W$的向量转换为了$C\times H\times 1$和$C\times W\times 1$两个向量。池化后将两个方向的信息沿同一维度进行合并操作。使用r减少通道数以降低模型复杂度。经过批归一化(Batch normalization)和激活函数h-Swish后，将网络再次沿先前维度分离、卷积以恢复到最初的大小。经sigmoid激活后，与原输出合并进行reweight操作。两个方向的单独分离处理可以使网络得到沿着两个空间方向的长期依赖关系，同时保留了精确的位置信息，有利于网络更加精确地定位到所需位置，提升网络对于非结构化道路环境中小目标物体类别分界处边缘的分割效果。

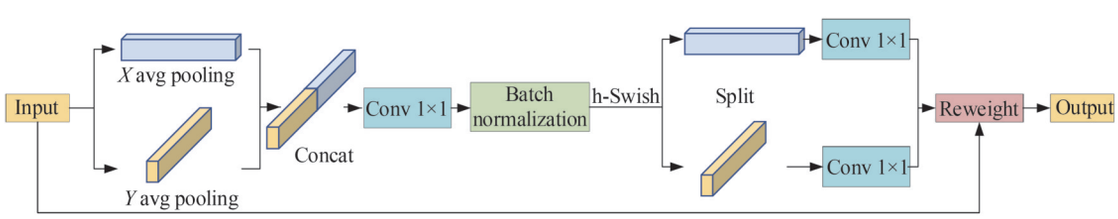

图3　坐标注意力机制结构

Fig.3　Coordinate attention structure

1.1.3　注意力机制与特征融合组合模块(AF)

为了提升非结构化道路小目标类别的分割精度，本文基于ASPP特征融合与CA注意力机制，提出改进的AF(Attention and Feature fusion)模块结构，如图4所示。模块以ASPP为基础骨干，在每个支路嵌入CA

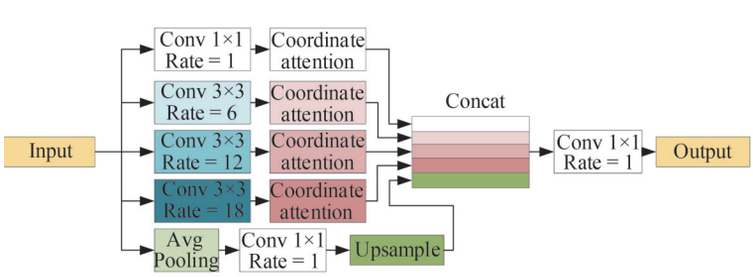

图4　AF模块结构

Fig.4　Attention and feature fusion structure

注意力机制，使网络同时关注到不同感知野的注意力信息。ASPP有效获取了不同采样率获取的图像特征，可以有效提升网络整体的分割精度，但针对小目标类别提升有限。因此通过在不同尺度特征信息后增添CA注意力，弥补了ASPP的不足。修改嵌入CA注意力相比网络整体，参数不在一个量级，因此即使在每条支路均添加了CA注意力，网络仍然处于轻量级，可以保证分割的实时性。

1.2 缓解类别不均衡的双权重交叉熵损失函数

道路类别像素数量统计如图5所示。道路图像部分类别，如车辆、绿化带、道路等样本出现频次较高，而部分小目标类别出现频次则远低于平均值，数据集存在严重的类别分布不均衡的问题。出现频率较小的类别在交叉熵损失函数计算中权重占比较小，训练过程中占比较大的类别会很快达到较高的精度，而较小的类别精度很难再提升。ICNet选用的是标准交叉熵损失函数。分别在三条支路上进行损失函数计算，总损失函数值由三条支路函数值叠加求得，标准的交叉熵损失函数并不能缓解类别平衡问题。

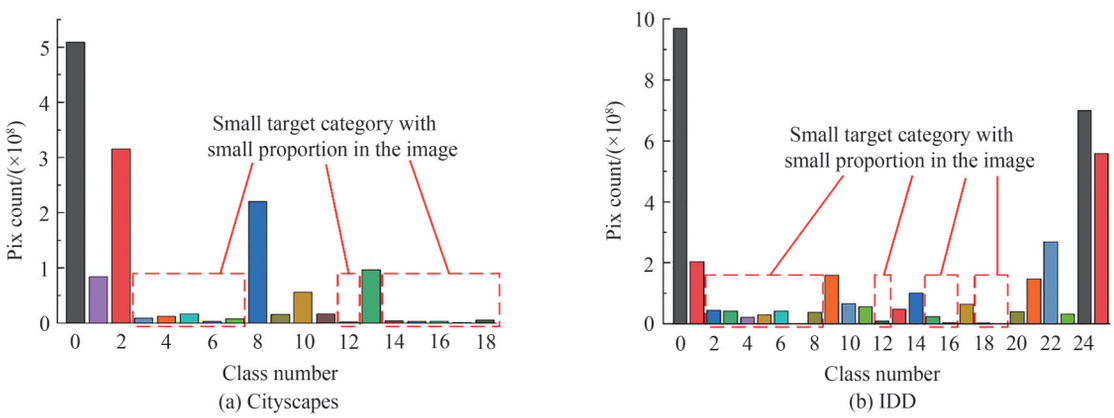

图5 Cityscapes数据集与IDD数据集的样本分布

Fig.5 Sample distribution of Cityscapes dataset and IDD dataset

为此本文修改损失函数，设计带权重的交叉熵损失函数，以提升网络对占比较小类别的关注。带类别权重的交叉熵损失函数表达式为

$$L=-\sum_{i=1}^{c}\omega_i y_i \log(P_i) \tag{7}$$

式中，参数ω_i用于平衡样本权重，针对像素较少类别，参数ω_i应较大，以提升对其的关注。本文参考ENet类别平衡方法[15]，定义权重ω_i为

$$\omega_i=\frac{1}{\ln(c+p_i)} \tag{8}$$

式中，p_i表示为相应类别i在图像中的像素占比。c为一设定超参数，用于限制权重间的间隔，c值越大，样本权重间的差值就越小。类别在图像中的像素占比越小，权重参数ω_i就越大，也越有利于网络对这些类别的信息提取能力。

基于以上权重损失函数，改进后的网络总损失函数表达式为

$$L=-\sum_{t=1}^{T}\lambda_t\sum_{i=1}^{c}\omega_i y_i \log\left(\frac{1}{\mathrm{softmax}(q)}\right) \tag{9}$$

式中，T表示支路数，在本文中T取3，λ_t表示支路权重，根据支路的不同分辨率图像，给定不同的数值，取值范围[0,1]。ω_i表示类别权重，根据不同的道路复杂程度，c的值有所不同。

通过计算分析，c取值在[1,1.1]区间内，可以保证类别权重差值在合理范围内。c较小时，类别权重差异过大，网络对占比较大类别的关注度过低，会影响整体的分割精度。c较大时类别权重差异较小，会退化到标准交叉熵损失函数。本文在取值范围内对c与λ进行了参数敏感性分析，如图6所示。

从图6可知，对于参数λ，当λ在0.4～0.6区间取值时，不同数据集的分割精度MIoU均取得了较高值。对于参数c，Cityscapes数据集在c=1.1时MIoU取得最大值，IDD数据集则是在c=1.02时MIoU取得最大

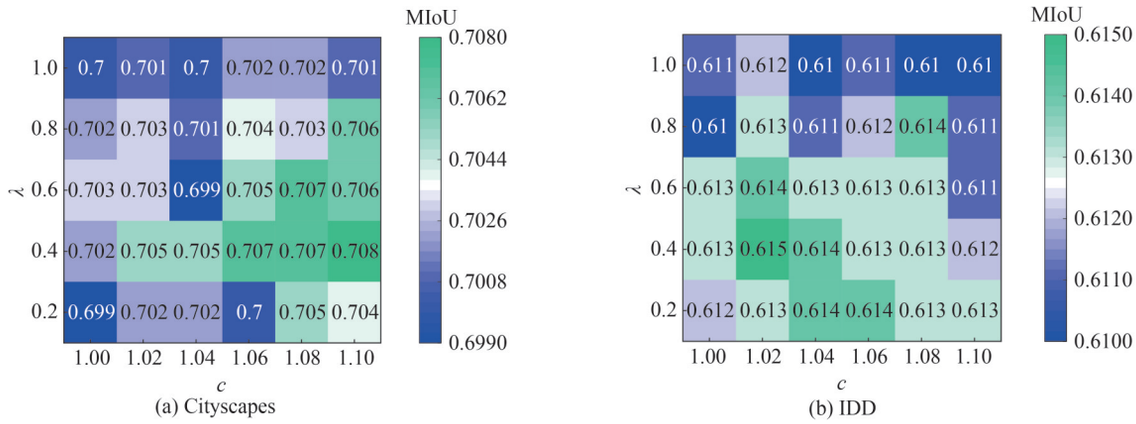

图6　λ与c的参数敏感性分布

Fig.6　Parameter sensitivity distribution of λ and c

值。分析认为，对于复杂程度较高的非结构化道路，c的取值影响类别权重差值，而较大的权重差值会提升网络的分割能力，因此c的取值非常重要。改进的损失函数通过对不同类别赋予不同的权重，增强了网络对像素占比较少样本的特征提取能力，有效地提升了网络对小目标类别的分割精度。

1.3　AF-ICNet网络整体结构

本文提出的AF-ICNet网络整体结构如图7所示，网络经由三个支路，分别对不同分辨率的图像进行特征提取。输入图像的分辨率基于Base size，分别压缩与扩张为0.5×Base size与2×Base size，分辨率从小到大分别输入到三个支路中，记为Sub1、Sub2、Sub4。Sub1采用较复杂的PSPNet网络骨干，大量的道路语义

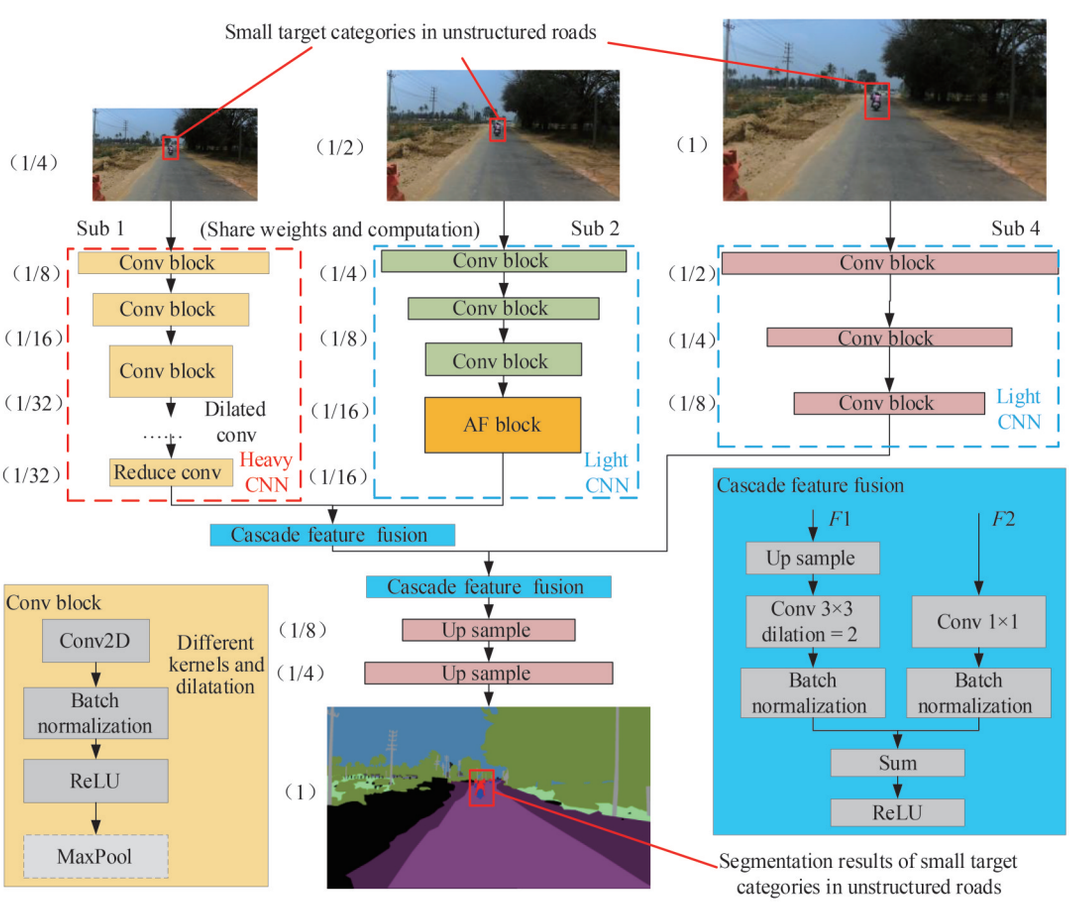

图7　AF-ICNet网络整体结构

Fig.7　The structure of AF-ICNet

信息均在此支路中获取，但是由于图像分辨率较低，网络对小目标的语义信息获取能力较差。Sub2分支与Sub1分支在骨干网络第一步的三层间共享部分卷积参数，通过低分辨率与中分辨率卷积参数的共享，加快了网络的分割速度。本文删除了Sub1中的PPM模块，修改Sub2分支，将ASPP与CA注意力机制相结合，建立AF模块。改进的Sub2支路兼顾网络实时性同时，进一步提升网络对图中的小目标类别的特征提取能力。Sub4输入分辨率较大，因此采用更少的网络层数。每个卷积后均连接批归一化（Batch normalization）及ReLU激活函数。为融合不同支路间的特征信息，在层级间使用级联特征融合（Cascade Feature Fusion，CCF），输入双通道信息F_1、F_2，F_1先进行双线性插值并使用大小为3×3，空洞率为2的卷积核来精修上采样特征。F_2使用1×1的卷积使其特征数量与F_1保持一致。F_1、F_2均进行批归一化操作，然后直接特征图相加并经过激活函数ReLU。经由三条支路以不同分辨率获取特征信息，网络有效地获取了不同分辨率的非结构化道路的语义信息。

2 数据集测试与分析

2.1 不同道路环境数据集选取

为满足多路况复杂道路环境检测的需求，本文使用了标准化道路数据集Cityscapes及非结构化道路数据集IDD（Indian Driving Dataset）进行训练与测试。

Cityscapes是一个关于城市街道场景的语义理解图像数据集，包括2 975张训练集图像，500张验证集图像和1 525张测试集图像。原数据集包含19个密集像素标注。由图8可知，结构化道路图像中道路边缘较为明确，道路比较平坦，纹理基本一致。车辆、道路标志信号等较为明确。AF-ICNet网络可以进一步提升对样本中小目标物体特征提取能力。

图8 Cityscapes数据集训练样本图片

Fig.8 Cityscapes dataset training sample image

IDD数据集由VARMA G等于2019年提出，这是一个专用于非结构化环境中道路场景感知的数据集。它由一位于车上的前置摄像头在印度的海得拉巴、班加罗尔及其郊区道路上进行拍摄获得[16]。

本文根据实际道路与数据集标注情况，选取其中的9 947张精细标注图像，将数据集分割为6 993张训练集图像，925张验证集图像和2 029张测试集图像。数据集包含多级标签标注。考虑非结构化道路实际分割需求，选取3级标签数据，将图像分割为26个类别。图9为IDD数据集的样本图与标签图。由图可知，非结构化道路图像中道路边缘较不明显，道路纹理大都不一致，且车辆、行人等小目标物体较多，无明显的道路标志，因此提取非结构化道路特征信息难度较大。

图9 IDD数据集训练样本图片

Fig.9 IDD dataset training sample image

2.2 数据集训练环境配置

实验在一台CPU为Intel i5-9400F、GPU为GTX2070、内存16 GB的计算机上运行。实验环境操作系统Ubuntu 18.04，语言环境为Python 3.6，编译环境为Pytorch 1.10.1，CUDA版本11.3。

为了使网络可以正常训练数据集，在训练时设置Cityscapes与IDD数据集的Base size为1 024，网络中Sub1与Sub4尺寸大小将被分别resize为0.5×Base size与2×Base size。Crop size设置为960。两个数据集的Batch size均设置为4。优化器采用随机梯度下降（Stochastic gradient descent，SGD）算法更新参数。为加速优化，采用动量梯度下降（Momentum SGD），在SGD基础上引入一阶动量，减少震荡过程，在加速收敛的同时保证梯度下降的平稳性，其优化公式为

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta - \gamma v_{t-1}) \tag{10}$$

$$\theta = \theta - v_t \tag{11}$$

相比起传统的SGD，Momentum SGD引入动量超参数γ，取值满足$0 \leqslant \gamma < 1$，本文中选取$\gamma=0.9$。使用学习率衰减调度器，在提供更快的收敛速度的同时保证算法收敛。学习率η初始选取0.01，衰减率选取0.000 1。训练轮数选取200轮，每一轮训练后均进行验证集验证。

2.3 数据集实验评价指标

网络分割精度评估指标为平均交并比（Mean Intersection over Union，MIoU）与像素精确率（Pixel Accuracy，PixAcc），实时性评估指标为实验测试速度FPS（帧/s）。

2.3.1 平均交并比MIoU

采用混淆矩阵统计模型分类的结果，分类结果混淆矩阵如表1所示。

表1 TP、FP、FN、FP含义

Table 1 The meaning of TP, FP, FN, FP

Real value	Predictive value	
	Positive	Negative
Positive	TP（True positive）	FN（False negative）
Negative	FP（False positive）	TN（True negative）

交并比IoU在语义分割中表示真实值和预测值两个集合的交集与并集之比。根据混淆矩阵，IoU可以改写成式（12），MIoU为对所有类别的IoU求平均值。

$$\mathrm{IoU} = \frac{\mathrm{TP}}{\mathrm{FP} + \mathrm{FN} + \mathrm{TP}} \tag{12}$$

2.3.2 像素精确率PixAcc

像素精确度PixAcc表示图像中正确分类的像素所占百分比。根据混淆矩阵，PixAcc可以改写为

$$\mathrm{PixAcc} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}} \tag{13}$$

2.3.3 每秒传输帧数FPS

FPS表示画面每秒传输帧数，用于判断分割图像的速度，计算公式为

$$\mathrm{FPS} = \frac{B}{\Delta t} \tag{14}$$

为测试图像分割帧数，在每次预测前后分别对当前时间进行记录，以获得时间差Δt，B表示Batch size。FPS越高，网络的实时性越好。

2.4 数据集训练测试与结果分析

2.4.1 Cityscapes数据集训练与测试

基于AF-ICNet进行训练和验证实验，在训练过程中，模型每完成一个epoch，进行验证集验证，记录验证集的损失函数值、MIoU以及PixAcc。通过以上操作，可以及时掌握模型的训练情况。

由图10可知，改进损失函数的函数值在Cityscapes数据集上随着迭代轮数的增加而降低，MIoU整体呈现随轮数增加而上升的趋势，并最后达到了较高数值，证明改进的损失函数与训练策略在Cityscapes数据集上的有效性。

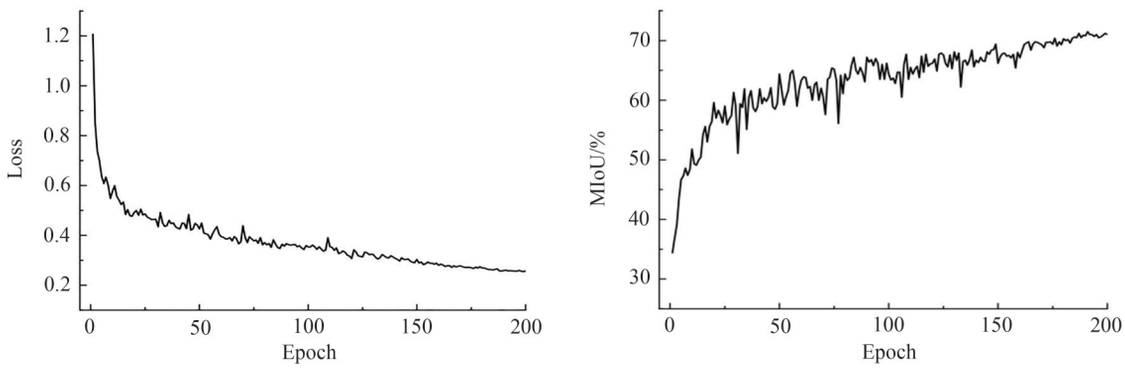

图 10 Cityscapes 数据集训练过程曲线
Fig.10 Cityscapes dataset training process curve

经过 200 轮训练后，模型在验证集上的结果与现有模型结果对比如表 2 所示。从表 2 中信息可以看出，本文网络算法（AF-ICNet）在 Cityscapes 数据集上有良好的表现。AF-ICNet 在验证集上的 MIoU 达到了 71.5%，高于 SegNet 与 ENet 网络，略低于 DeepLabV3 Plus-MobileNet。本文网络的 PixAcc 达到了 81.3%。

表 2 AF-ICNet 在 Cityscapes 数据集上的测试结果
Table 2 Test results of AF-ICNet on Cityscapes dataset

Model	MIoU/%	PixAcc/%	FPS	Param/($\times10^6$)
SegNet[17]	57.9	—	13.3	29.5
ENet[18]	58.3	—	90.9	0.4
ICNet	69.7	80.9	45.5	26.5
DeepLabV3Plus-MobileNet	72.1	83.8	10.4	41.3
Proposed AF-ICNet	**71.5**	**81.3**	**43.5**	**27.9**

针对小目标类别，AF-ICNet 同样取得了更高的结果。Cityscapes 中占比较小的行人、交通灯、交通标志类别的 IoU 分别从 73.4%、55.5%、67.3% 提升到了 76.3%、61.2%、72.2%，证明了 AF-ICNet 有效的提升了小目标类别的分割精度。FPS 达到了 43.5 帧/s，虽低于 ENet 与 ICNet 网络，但远高于 SegNet 与 DeepLabV3Plus-MobileNet，满足实时性要求。本文网络的 MIoU 达到了与高精度网络 DeepLabV3Plus-MobileNet 相近的精度，但是参数量却远低于该网络，FPS 也远高于该网络。证明了改进 AF-ICNet 能在保证实时性的情况下，实现高精度的道路语义分割。

图 11 为 ICNet 与 AF-ICNet 对验证集的分割测试结果，图中虚线框为边界混淆区域，实线框为小目标类

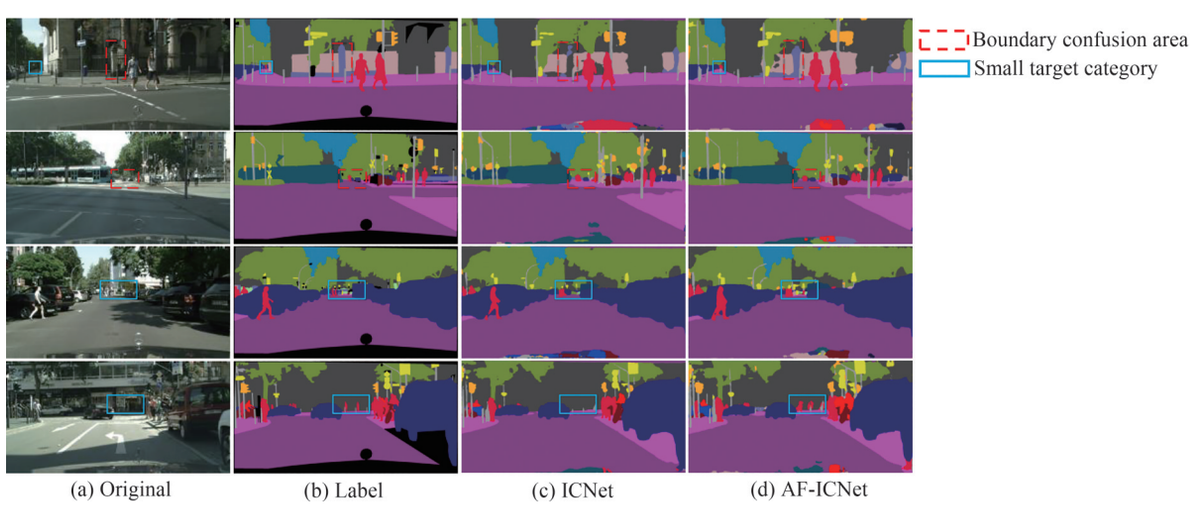

图 11 Cityscapes 数据集分割测试结果
Fig.11 Test result of Cityscapes dataset

别区域。对于整体分割准确率，AF-ICNet略优于ICNet，如图11虚线框所示。AF-ICNet分割边缘更贴近标签图像，对于相似类别的区分能力较强，例如道路与步行道类别。

AF-ICNet对小目标类别分割精度明显优于ICNet，如图11实线框所示。ICNet对于远景的小目标提取能力较差，往往会将这类小目标类别归于远景中的大目标类别。AF-ICNet则能够有效地将远景小目标类别提取出来，并且当图像中存在多种小目标类别混杂情况时，AF-ICNet能够将小目标区分开来。

2.4.2 IDD数据集训练与测试

基于AF-ICNet进行训练和验证实验，模型每完成一个epoch，进行验证集验证，记录验证集的损失函数值、MIoU以及PixAcc。

由图12可知，改进损失函数的函数值在IDD数据集上走势与Cityscapes数据集类似，MIoU最后同样达到了较高数值，证明改进的损失函数与训练策略在IDD数据集上的有效性。经过200轮训练后，模型在验证集上的结果与现有模型结果对比如表3所示。

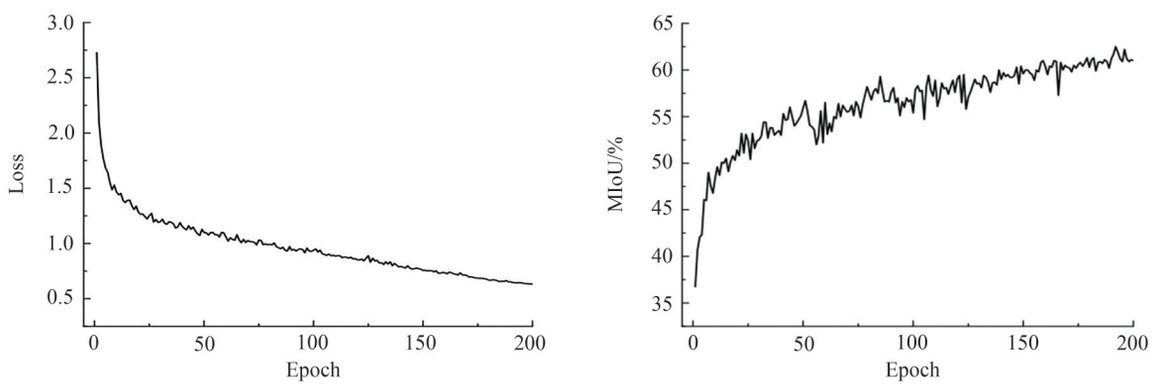

图12 IDD数据集训练过程曲线

Fig.12 IDD dataset training process curve

表3 AF-ICNet在IDD数据集上的测试结果

Table 3 Test results of AF-ICNet on IDD dataset

Model	MIoU/%	PixAcc/%	FPS	Param/($\times10^6$)
SegNet	31.9	—	13.0	29.5
ENet	48.3	—	83.3	0.4
ICNet	60.9	89.4	47.6	26.5
DeepLabV3Plus-MobileNet	61.2	89.6	11.2	41.3
Proposed AF-ICNet	**62.5**	**89.8**	**41.7**	**27.9**

从表3中信息可以看出，本文网络算法（AF-ICNet）在IDD数据集上同样有良好的表现。AF-ICNet在验证集上的MIoU达到了62.5%，高于其他网络。PixAcc达到了89.8%。针对小目标类别，AF-ICNet同样取得了更高的结果。非结构化道路中占比较小较难分割的的的行人、骑者、交通信号类别的IoU分别从58.8%、70.0%、26.3%提升到了60.4%、72.0%、34.3%，证明了AF-ICNet有效的提升了小目标类别的分割精度。FPS达到了41.7帧/s，虽低于ENet与ICNet网络，但仍远高于SegNet与DeepLabV3Plus-MobileNet，满足实时性要求。AF-ICNet在IDD数据集上的MIoU超出DeepLabV3Plus-MobileNet 1.3%，证明了本文改进AF-ICNet对非结构化道路的分割精度提升很大。

由于非结构化道路复杂程度更大，ICNet与AF-ICNet对验证集分割精度差距进一步加大，如图13所示，图中虚线框为边界混淆区域，实线框为小目标类别区域。对于整体分割准确率，ICNet分割图像中出现了更多的相似类别的混淆情况，与标签图像相比错误率显著增大，AF-ICNet的整体分割精度则高于ICNet，如图13中的虚线框所示。

AF-ICNet对远景图像中的小目标分割能力比ICNet更强，如图13中的实线框所示。即使在图像中占比极小的远景小目标物体，AF-ICNet仍然有很强的能力将其分割出来，而ICNet的小目标分割效果则明显

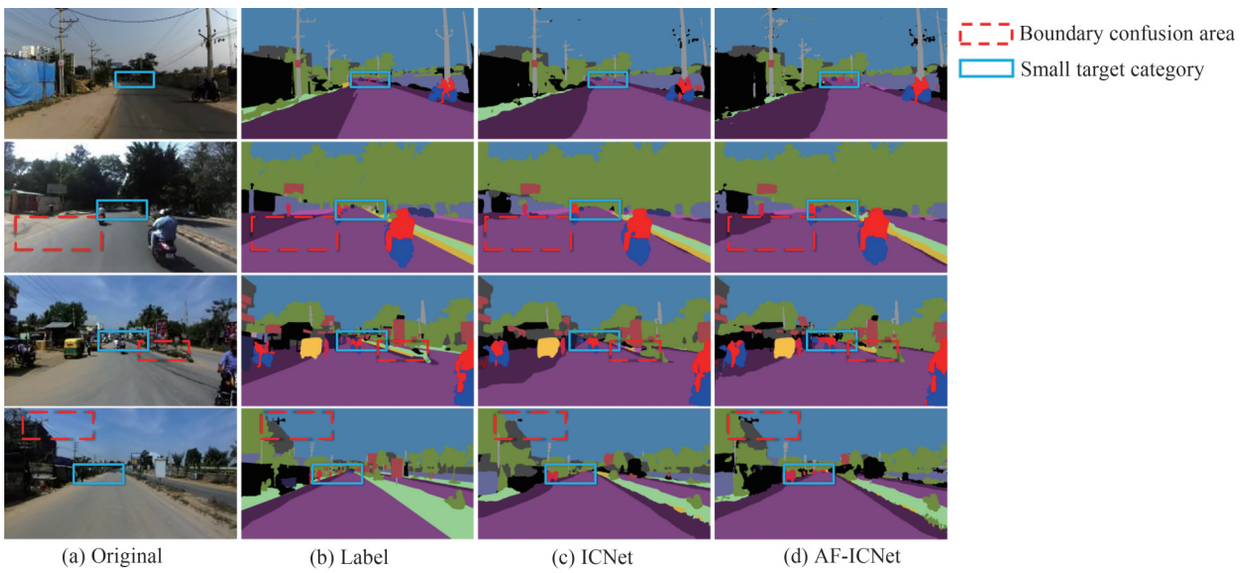

图 13 IDD 数据集非结构化道路分割测试结果

Fig.13 Test result of unstructured road segmentation of IDD dataset

劣于 AF-ICNet。

2.5 消融实验

消融实验类似于控制变量的思想，在机器学习领域常使用消融实验来分析不同的因素对神经网络实验产生的影响[19-20]。为进一步分析各改进模块对原始模型 ICNet 的影响，将本文的方法分别裁剪成 4 组进行训练。其中，第 1 组为原始的 ICNet 模型，第 2 组增加使用权重交叉熵损失函数，第 3 组在第 2 组基础上增加改进 ASPP，第 4 组在第 3 组基础上增加 CA 注意力机制，即为本文的 AF-ICNet 方法。消融实验结果如表 5 所示，其中“√”表示实验中包括该结构，“×”表示实验中未包括该结构。

表 4 消融实验

Table 4 Ablation experiment

Group	W-CE	ASPP	CA	Cityscapes			IDD		
				MIoU/%	PixAcc/%	FPS	MIoU/%	PixAcc/%	FPS
Exp 1	×	×	×	69.7	80.9	**45.5**	60.9	89.4	**47.6**
Exp 2	√	×	×	70.8	81.1	45.2	61.5	89.0	45.4
Exp 3	√	√	×	71.0	81.1	44.4	62.1	89.0	43.4
Exp 4	√	√	√	**71.5**	**81.3**	43.5	**62.5**	**89.8**	41.7

分析消融实验，各改进模块对于网络分割精度均有明显提升。对于 Cityscapes 数据集，使用权重交叉熵损失函数，MIoU 提升了 1.1%。增加 ASPP 特征融合，MIoU 进一步提升了 0.2%。在 ASPP 基础上增加 CA 模块，形成 AF-ICNet 网络，MIoU 再次提升 0.5%，达到 71.5%，而 PixAcc 为 81.3%。对于 IDD 数据集，使用权重交叉熵损失函数，MIoU 提升了 0.6%。增加 ASPP 特征融合，MIoU 进一步提升 0.6%，证明了 ASPP 有效提升了非结构化道路分割精度。在 ASPP 基础上增加 CA 模块，形成 AF-ICNet 网络，MIoU 再次提升 0.4%，达到 62.5%，而 PixAcc 为 89.8%。AF-ICNet 网络的 FPS 虽略低于原网络，但是仍满足实时性要求。综上，AF-ICNet 与原 ICNet 相比，能在保证实时性要求的同时，大幅提升网络的分割精度。

3 实景测试实验

3.1 实验测试系统

为验证 AF-ICNet 网络模型在实际环境中的测试效果，本文进行了非结构化道路的实景测试实验。搭建了实验测试系统，如图 14 所示。实验测试系统由 TurtleBot 机器人底盘、笔记本电脑、Microsoft KinectV1

相机等组成。使用笔记本电脑控制TurtleBot机器人的移动，并采集相机图像。笔记本电脑配置如下：CPU为Intel i7-9750H、GPU为GTX1660Ti、内存16 GB。电脑操作系统为Windows11，语言环境为Python 3.8，编译环境为Pytorch 1.10.1，CUDA版本11.3。使用实验测试系统对非结构化道路进行实景测试，如图15所示。实验中相机拍摄RGB图像分辨率为1 280×960，拍摄速度FPS为12帧/s，满足实时性需求。实验选取非结构化程度较高的场景，其具有道路边线不明确，纹理不一致等问题。同时为验证本文方法对小目标的分割效果，人为增加小目标障碍物。

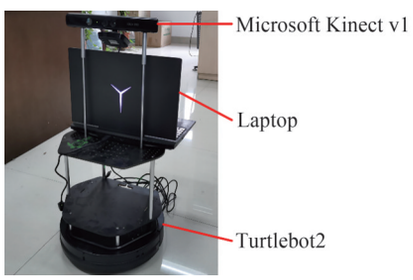

图14　实验测试系统

Fig.14　The experimental testing system

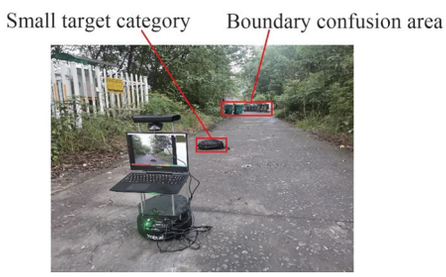

图15　非结构化道路现场测试图

Fig.15　The image of field work on unstructured roads

3.2　现场测试实验分析

利用搭建的实验测试系统在现场拍摄图像，选取了非结构化程度较高的道路，并按照非结构化数据集训练的模型结果，使用AF-ICNet网络进行语义分割，结果如下图16所示。

图16　现场测试非结构化道路分割结果

Fig.16　Segmentation results of unstructured roads tested in the field

从图16中可知，AF-ICNet网络能有效分割出非结构化道路类别，其中关于小目标类别与边界混淆区域的分割效果比ICNet网络更好。图16实线框为小目标障碍物区域。对比分割图像可以看出，AF-ICNet在小目标分割中取得了更好的效果，小目标障碍物的边缘更加平滑，与原图像的匹配度更高。图16虚线框为边界混淆区域，从图中可知，AF-ICNet能将道路中的边界混淆区域准确地分割出来，从而可以有效地获取非结构化道路的语义信息。在移动测试平台上，AF-ICNet网络单帧1 280×960高分辨率图像的分割速度为17帧/s，这证明本文网络在性能配置一般的计算机上对高分辨率图像的处理速度同样较快。当分辨率为

1 280×720时，分割速度达到了31帧/s，完全满足道路实时性检测要求。实景测试实验表明，AF-ICNet在测试场景应用中取得良好的检测效果，具有很强的实际应用价值。

4 结论

本文针对非结构化道路环境语义分割存在误检率较高、边界容易混淆、小目标检测能力较差等问题，提出基于小目标注意力机制与特征融合的AF-ICNet复杂道路语义分割方法。采用空洞空间卷积池化金字塔融合不同尺度特征感受野以增强网络的全局感知能力。利用CA注意力机制模块提升了网络对图像中占比较小类别的信息提取能力。修改交叉熵损失函数提升了网络对出现频次较少类别的关注。数据集测试实验表明，AF-ICNet对Cityscapes数据集分割精度MIoU与PixAcc分别达到了71.5%与81.3%，对IDD数据集分割精度MIoU与PixAcc分别达到了62.5%与89.8%。搭建实验系统进行了实景测试。实景测试证明，AF-ICNet网络对小目标类别与非结构化道路整体分割能力提升显著，与ICNet网络相比，在保证测试实时性的情况下，实现了更高精度的语义分割。未来研究工作中将增加道路环境的复杂度，如增加雨、雪、雾等复杂天气条件，在进一步优化网络结构基础上，实现更加复杂非结构化场景的语义信息提取和分割。

参考文献

[1] GUO Lei, WANG Qiulong, XUE Wei. et al. A small object detection algorithm based on improved YOLOv5[J]. Journal of University of Electronic Science and Technology of China, 2022, 51(2): 251-258.
郭磊, 王邱龙, 薛伟, 等. 基于改进YOLOv5的小目标检测算法[J]. 电子科技大学学报, 2022, 51(2): 251-258.

[2] CHEN Haoran, PENG Li. Detection algorithm of small target in receptive field block[J]. Journal of Frontiers of Computer Science & Technology, 2021, 15(2): 346-353.
陈灏然, 彭力. 感受野下的小目标检测算法[J]. 计算机科学与探索, 2021, 15(2): 346-353.

[3] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(4): 640-651.

[4] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 12(39): 2481-2495.

[5] RONNEBERGER O, FISCHER P, BROX T. U-Net: convolutional networks for biomedical image segmentation[C]. International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015: 234-241.

[6] CHEN L C, PAPANDEROU G, KOKKINOS I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.

[7] ZHANG Kaihang, JI Jie, JIANG Luo, et al. The semantic segmentation of driving regions on unstructured road based on segnet architecture[J]. Journal of Chongqing University(Natural Science Edition), 2020, 43(3): 79-87.
张凯航, 冀杰, 蒋骆, 等. 基于SegNet的非结构道路可行驶区域语义分割[J]. 重庆大学学报, 2020, 43(3): 79-87.

[8] GONG Zhili, GU Yuhai, ZHU Tengteng, et al. Unstructured road recognition based on attention mechanism and lightweight DeepLabv3+[J]. Microelectronics & Computer, 2022, 39(2): 26-33.
龚志力, 谷玉海, 朱腾腾, 等. 融合注意力机制与轻量化DeepLabv3+的非结构化道路识别[J]. 微电子学与计算机, 2022, 39(2): 26-33.

[9] LIU Bin, LIU Hongzhe. Lane detection algorithm based on improved enet network[J]. Computer Science, 2020, 47(4): 142-149.
刘彬, 刘宏哲. 基于改进Enet网络的车道线检测算法[J]. 计算机科学, 2020, 47(4): 142-149.

[10] HENGSHUANG Z, XIAOJUAN Q, XIAOYONG S, et al. ICNet for real-time semantic segmentation on high-resolution images[J]. arXiv preprint, arXiv:1704.08545, 2018.

[11] LIANG-CHIEH C, YUKUN Z, GEORGE P, et.al. Encoder-decoder with atrous separable convolution for semantic image segmentation[J]. arXiv preprint, arXiv:1802.02611, 2018.

[12] JIE H, LI S, GANG S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 8(42):2011-2023.

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

[14] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[J]. arXiv preprint, arXiv: 2103.02907,2021.

[15] PASZKE A, CHAURASIA A, KIM S, et al. ENet: A deep neural network architecture for real-time semantic segmentation[J]. arXiv preprint, arXiv:1606.02147, 2016.

[16] VARMA G, SUBRAMANIAN A, NAMBOODIRI A, et al. IDD: a dataset for exploring problems of autonomous

navigation in unconstrained environments[J]. arXiv preprint, arXiv:1811.10200,2018.
[17] BERMAN M, TRIKI A R, BLASCHKO M B. The lovasz-softmax loss: a tractable surrogate for the optimization of the intersection-over-union measure in neural networks[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 4413-4421.
[18] WANG K, YANG J, YUAN S. et al. A lightweight network with attention decoder for real-time semantic segmentation [J]. The Visual Computer, 2022, 38(5): 2329-2339.
[19] HUANG Ziliang, FANG Chenhao, OUYANG Xiaoping, et al. Research on the sensing system of lower limb exoskeleton robot based on multi-information fusion[J]. Chinese Journal of Engineering Design, 2018, 25(2): 159-166.
黄梓亮, 方晨昊, 欧阳小平, 等. 基于多信息融合的下肢外骨骼机器人感知系统研究[J]. 工程设计学报, 2018, 25(2): 159-166.
[20] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(6): 1137-1149.

AF-ICNet Semantic Segmentation Method for Unstructured Scenes Based on Small Target Category Attention Mechanism and Feature Fusion

AI Qinglin[1], ZHANG Junrui[1], WU Feiqing[2]
(1 *Key Laboratory of Special Purpose Equipment and Advanced Manufacturing Technology, Ministry of Education & Zhejiang Province, Zhejiang University of Technology, Hangzhou*, 310023, *China*)
(2 *School of Information Science and Engineering, Ningbotech University, Ningbo* 315100, *China*)

Abstract: There are a lot of unstructured road scenes in the actual road driving, large-scale engineering work and field work of robots. Compared with the structured road, the unstructured road has the characteristics of small color difference between the road and its surroundings, large variety of road target species and complex information. Traditional methods for detecting unstructured road areas have problems such as low detection accuracy, poor real-time performance and poor detection effect for small targets. Small target obstacles and pedestrians will seriously interfere with the detection of viable road areas. Small target has low resolution and little information in the image, which leads to poor feature characterization capability. The dominance of large categories also easily leads to the neglect of small target categories. To solve the above problems, we construct a lightweight real-time semantics segmentation network of AF-ICNet based on small target category attention mechanism and feature fusion. Firstly, the pyramid pooling module structure in ICNet is replaced by atrous spatial pyramid pooling, which combines feature receptive fields of different scales to reduce the pooling effect, and finally enhances the network's ability to perceive the global image. On this basis, we embedded coordinate attention mechanism in improved ICNet model. We establish channel information and spatial location information to enhance the network's ability to extract the small target category semantics features of unstructured roads. This method of fusing channels and spatial attention is different from both SE-Net and CBAM. Finally, in view of the imbalance of category distribution in unstructured road scenes, we design a Weighted Cross-Entropy loss function to improve the network's attention to small target categories. The weight of the branch can effectively improve the attention of the network to images of different resolutions. The weight of the category can effectively improve the network's attention to small target categories. In order to verify the validity of the super parameters, we carried out the parameter sensitivity analysis experiment, and the value range of the parameters is determined. Based on the above improvements, we designed AF-ICNet semantic segmentation network. In order to verify the improvement of the model, we use the AF-ICNet model to train Cityscapes and IDD datasets. After training, 19 categories are segmented in the Cityscapes dataset, the final MIoU of AF-ICNet reaches 71.5%, and the final PixAcc reaches 81.3%. 26 categories are segmented in the IDD dataset and the final MIoU of AF-ICNet reaches 62.5%, and the final PixAcc reaches 89.8%. To verify the effectiveness of each improvement point, we perform the ablation experiments. We divided into four groups for the experiment. The four groups of experiments are the original network, the network with Weighted Cross-Entropy, the network with ASPP and Weighted

Cross-Entropy, and AF-ICNet with CA attention mechanism. The experimental results show that each improvement point of AF-ICNet can effectively improve the network segmentation accuracy on the basis of guaranteeing the real-time performance of the network. In order to further verify the effectiveness of the improved method in practical application, we establish an experimental testing system for a field test, and use training model of the IDD dataset to test. In the real scene test experiment, AF-ICNet effectively segmented the road area and the surrounding objects, and for the manually placed small target objects, the segmentation edge of AF-ICNet is more accurate. In terms of test speed, AF-ICNet achieves a segmentation speed of 17FPS on a 1 280×960 image, and a segmentation speed of 31FPS on a 1 280×720 image, which fully meets the real-time requirements of road segmentation. The test results show that AF-ICNet effectively improves the network segmentation effect. In the case of real-time performance, the segmentation accuracy of small target categories is improved.

Key words: Semantic segmentation of small target category; AF-ICNet; CA attention mechanism; Atrous spatial pyramid pooling; Loss function

OCIS Codes: 100.3008; 110.2970; 150.1135; 200.4260

Foundation item: National Natural Science Foundation of China (No.52075488), Zhejiang Provincial Natural Science Foundation of China (No.LY20E050023)