基于MFF-STDC网络的室外复杂环境小目标语义分割方法

艾青林，刘元宵，杨佳豪

（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）

摘 要：针对轻量化网络在复杂环境中小目标类别物体分割效果较差的问题，本文搭建了基于多层级特征融合的MFF-STDC网络模型。首先，通过多次叠加基于分组卷积的特征提取模块，使网络特征提取能力提升。其次，通过分层权重注意力优化模块与通道注意力（CA）机制，提升多尺度特征信息的融合能力。最后，建立基于自适应复制算法的A-Cityscapes数据集、A-IDD数据集以及Field数据集，增加数据集中小目标类别的数量，并完成训练与测试。MFF-STDC网络与STDC对比，mIoU分别提升了4.01%、3.65%、2.94%，并且对复杂环境中小目标类别的分割效果远好于其他网络。搭建实景测试实验平台，测试结果表明，MFF-STDC网络有效提升了小目标类别的语义分割精度与分类能力，并且满足实时性要求。

关键词：计算机应用；小目标类别检测；多层级特征融合；CA机制；自适应复制算法

中图分类号：TP391 **文献标志码**：A **文章编号**：1671-5497(2025)08-2681-12

DOI：10.13229/j.cnki.jdxbgxb.20231299

Small target swmantic segmentation method based MFF-STDC network in complex outdoor environments

AI Qing-lin，LIU Yuan-xiao，YANG Jia-hao

（*Key Laboratory of Special Purpose Equipment and Advanced Manufacturing Technology，Ministry of Education and Zhejiang Province，Zhejiang University of Technology，Hangzhou* 310023，*China*）

Abstract：The MFF-STDC network model based on multi-level feature fusion is built in this paper to solve the problem that lightweight networks have weak segmentation effect for small target category objects in complex environment. Firstly，by superimposing the feature extraction module based on group convolution many times，the feature extraction capability of the network is improved. Secondly，the combination ability of multi-scale feature information is improved by hierarchical attention module and CA mechanism. Lastly，A-Cityscapes dataset A-IDD dataset and Field dataset were built based on adaptive replication algorithm，the number of small target categories in the dataset was increased，and training and testing were completed. The MFF-STDC network improves the mIoU by 4.01%，3.65%，and 2.94% respectively comparing with the STDC，and segmentation of the small target categories in the complex environment is much better than

收稿日期：2023-11-24.

基金项目：国家自然科学基金项目（52075488）；浙江省自然科学基金项目（LY20E050023）.

通信作者：艾青林（1976-），男，教授，博士，研究方向：智能机器人技术和基于机器学习的复杂环境语义分割方法. E-mail：aqlaql@163.com

that of other networks. A real-world testing experimental platform is built, and the test results show that the MFF-STDC network effectively improves the semantic segmentation accuracy and classification ability of small target categories, and meets the real-time requirements.

Key words: computer applications; small target category detection; multi-level feature fusion; coordinate attention mechanism; adaptive replication algorithm

0 引 言

复杂环境下小目标类别的语义分割是计算机视觉领域的一个重要研究方向[1],目前研究者多采用上下文语义信息推理、多尺度特征信息提取、边缘信息提取能力增强、损失函数增加、数据增强等方法提高对小目标类别物体的分割效果[2,3]。

Takikawa等[2]提出的Gated-SCNN网络增加了新的支路,用于提取边缘信息,达到提升小目标类别分割效果的目的。Chen等[3]通过空洞空间金字塔池化(Atrous spatial pyramid pooling, ASPP)结构改进的DeepLab v3+模型,有较强的多尺度特征信息提取能力,在小目标分割任务上效果显著。但这些网络都有着较大的参数量和计算量,难以部署到室外机器人上,并且难以满足实时性要求[4]。

Fan等[5]在BiSeNet[6]上提出了STDC网络模型,STDC网络模型是以其特征提取模块短期密集连接(Short-term dense concatenate, STDC)进行命名的轻量化网络模型。其在BiSeNet的基础上改进了原有的边缘提取支路,减少了参数量,且分割效果显著。但STDC网络上下文语义路径对小目标类别的特征提取能力有限,在应用到小目标类别的分割任务时,难以取得较好的分割效果。

本文在STDC网络基础上提出了短期密集连接的多层级特征融合(Muti-layer feature fusion, MFF),即MFF-STDC轻量化网络模型,在保证实时性的情况下进一步加强算法在室外复杂环境下对小目标的分割效果。其主要创新点如下:首先通过多次叠加基于分组卷积的特征提取模块,提升网络特征信息提取能力。其次通过分层权重注意力优化模块与多感受野的信息结合模块,增强网络多层级间信息的提取能力,在特征融合模块(Feature fusion module, FFM)中加入通道注意力(Cooralinate attention, CA)机制[7],加强上下文语义信息的融合。最后建立基于自适应复制算法的A-Cityscapes数据集、A-IDD数据集以及Field数据集,增加小目标类别的样本数量,使得经过训练与测试的网络模型在小目标分割任务上具有更好的通用性与分割效果。

1 基于多层级特征融合的MFF-STDC语义分割方法

1.1 网络整体结构

MFF-STDC网络的整体结构如图1所示,分为空间细节路径(spatial path)和上下文语义路径(context path)。空间细节路径从分割图的真实值开始,通过对分割图进行边缘提取并将提取结果输入到细节检测器(detail head),完成对Stage3的指导,得到空间细节信息,再通过多感受野的信息结合(Multi-scale information combination, MIC)模块进一步强化空间细节信息。在上下文语义路径中首先利用Stage3~Stage5中的DW-STDC(Depth-wise STDC)模块提取特征信息。其次利用分层权重的注意力优化模块(Hierarchical attention module, HAM)处理2个不同层次特征图的信息,通过层与层之间有选择的按比例融合,使网络学习针对性变强。最后通过改进后的通道注意力特征融合模块(Coordinate attention-feature fusion module, CA-FFM)将空间细节信息与上下文语义信息进行特征融合,得到最终的输出结果。

1.2 基于分组卷积改进的特征提取模块

考虑到在判断小目标类别物体与背景噪声时,网络需要加强不同层级之间的信息融合能力与更深层次的特征信息提取能力,会加大模型的运算量,故需要尽可能地降低网络其他部分的运算量。原有STDC网络中的特征提取模块,采用常规卷积提取特征,本文对STDC模块进行改进,利用分组卷积代替该模块中的常规卷积,以降低该模块的计算量与参数量[8]。同时,利用节省下来的计算资源,可以叠加更多层的改进模块,如图1中Stage3~Stage5所示,这种改进结构可使网络获取更深层次的小目标特征信息,实现

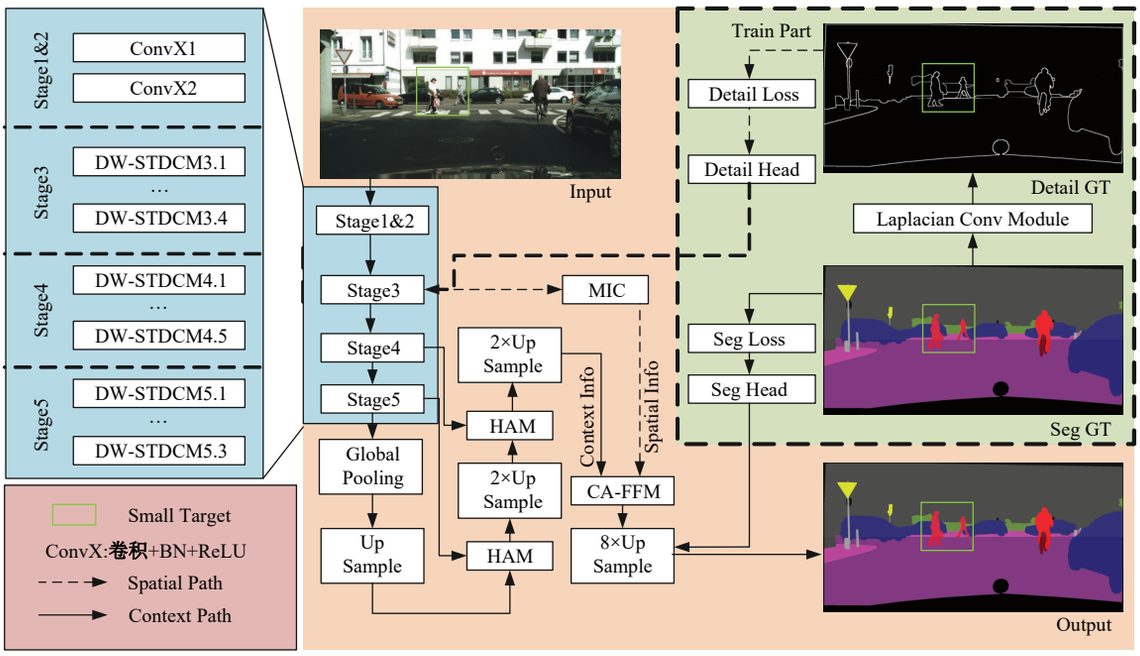


图 1　MFF-STDC 网络整体结构

Fig. 1　Overall structure of the MFF-STDC newwork

大感受野内各个像素信息的充分融合。

改进后的特征提取模块为 DW-STDC，对于 DW-STDC 模块的第一个卷积层，仍采用常规卷积，以使上一层流通下来的信息可以充分交互，对于后续的几层卷积层，采用分组数为输出通道数的分组卷积，以降低计算量。对于最后一层卷积层，其输入通道数与输出通道数相等，即该卷积为深度可分离卷积。在完成最后不同特征层次的通道维度融合后，使用通道重排打乱不同维度间的排序，以达到通道间信息交互的目的。将 DW-STDC 模块在 Stage3～Stage5 中多次叠加，提取网络更深层次的特征信息。

1.3　不同层次的特征信息结合与提取

1.3.1　基于分层权重的注意力优化模块

STDC 网络上下文语义路径中不同尺度的信息融合仅通过一个注意力融合模块（Attention refine module，ARM）进行简单相加[6]。该模块获取信息的能力不够强，且不同层级的信息融合方式也较为简单。为了进一步提升网络中多尺度信息的融合能力，本文通过双层感知注意力融合模块（Double layer perception refine module，DLPRM）提高对通道信息的提取能力，再利用权重交互使不同尺度的信息有了倾向性的融合，构建了分层权重的注意力优化模块 HAM。DLPRM 结构如图 2 所示，首先利用全局平均池化和最大池化分别对输入进行不同方式的处理，其次通过一个沙漏形的双层感知机完成通道间的信息交互，最后将得到的注意力特征用于指导输入特征。

HAM 的结构如图 3 所示，2 个输入分别为不同层级的 2 个特征图，通过 DLPRM 的信息提取以及带有权重的相加操作，网络模型在训练时的反向传播中更容易筛选有效信息，使得不同层级中的网络特征信息有选择性地被保留下来。

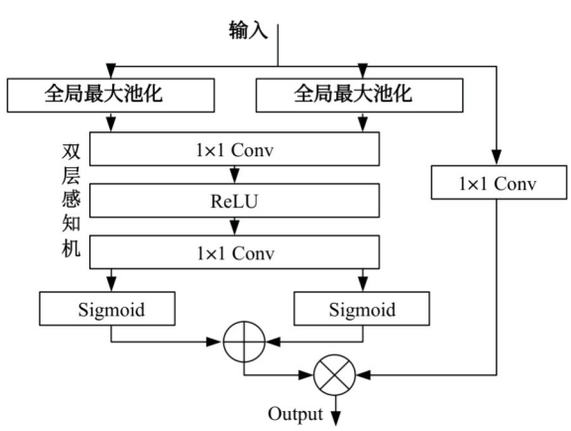


图 2　DLPRM 的结构

Fig. 2　Structure of DLPRM

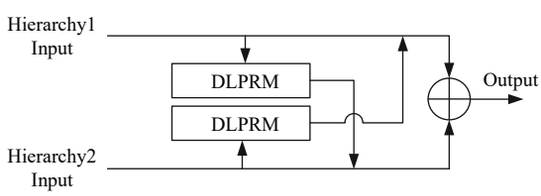


图 3　HAM 的结构

Fig. 3　Structure of HAM

1.3.2 基于多感受野的信息结合模块

在原STDC网络的空间细节路径中,Stage3的输出在训练时直接与边界信息图利用损失函数进行反向传播[6]。而Stage1～Stage3较为简单,能获取的空间信息层次单一,不利于不同尺寸、不同范围的细节信息的学习。而单一使用空洞卷积又容易导致特征图连续性丧失,同时造成小目标物体的信息学习困难。为此,本文采用了一种多感受野的信息结合模块对空间细节路径提取的信息进行强化,并将其命名为MIC模块,该模块的结构如图4所示。

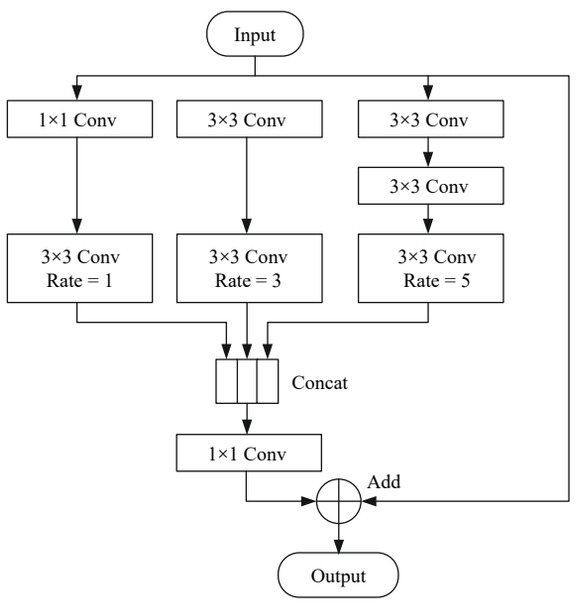


图4 MIC模块的结构

Fig. 4 Structure of MIC module

从图4中可以看出,将输入特征图复制3份,分别经过若干个卷积层后,再经过不同扩张率的空洞卷积层,获取不同尺寸的感受野。将3种感受野的特征通道叠加,并经过一个1×1的卷积层调整通道维度后与原特征图相结合,用于指导原特征图。

1.3.3 基于CA机制的FFM

在原STDC网络FFM中,仅用一个全连接操作进行语义信息与空间信息的融合[6],对此增加CA机制[7],加大小目标类别区域的权重,同时也能起到抑制无关噪声的作用[9]。本文对原有的FFM作了改进,并将其命名为CA-FFM。其结构如图5所示。

在CA-FFM中,获取注意力的方式依旧为在常规的网络运算外以额外支路对原特征图进行指导。CA-FFM通过精确的位置信息对通道关系

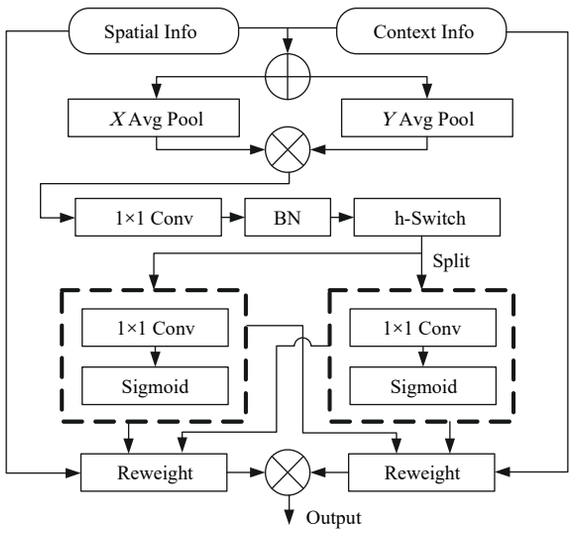


图5 改进后的CA-FFM

Fig. 5 Improved CA-FFM

和长期依赖性进行编码,增强输入的特征图在方向和位置上的敏感。将全局池化操作分解为X、Y 2个方向上的平均池化操作,其自身构成坐标关系的2个方向本身就蕴含了充足的空间信息,对单个方向维度的压缩更是让空间信息的权重能参与到模型学习中,精确定位到感兴趣的区域。整个过程将通道和空间的权重融合转移到2个不同方向的权重依赖上,节省了计算量,使网络性能更高,有效扩大了模型的感受野,从而提升了对小目标类别物体的分割效果。

1.4 不同室外环境数据集建立

测试网络分割精度时,考虑到已有的Cityscapes数据集[10]和IDD(india driving dataset)数据集[11]内小目标物体数量较少,本文通过自适应复制算法构建了专门针对小目标样本的数据集,然后进行分割精度对比实验。本次实验采用了多个数据集进行测试,包括A-Cityscapes(Augment Cityscapes)数据集、A-IDD(Augment IDD)数据集和Field数据集(自建野外数据集)。因为A-Cityscapes数据集和A-IDD数据集是以Cityscapes数据集和IDD数据集为基础进行数据增强得到的数据集,其中数据集划分与原数据集划分方式一致。A-Cityscapes数据集含5 000张图像,被分为2 975张训练集图像、500张验证集图像和1 525张测试集图像。A-IDD数据集含10 003张图像,被分为6 993张训练集图像、981张验证集图像和2 029张测试集图像。

1.4.1 自适应复制数据增强算法

利用自适应复制算法建立A-Cityscapes数据

集和A-IDD数据集，增加其中的小目标样本数量。自适应复制算法主要分为以下4个步骤。

第一步，找出图像中的灭点（vanishing point），将图像平面映射到球形极坐标系，并以1°为间隔建立90×360的网格空间。将初始网格权值 $W_0(\varphi, \lambda)$ 设为0，利用直线段检测算法（Line segment detector，LSD）获取图像中所有直线段，对图像中已提取出来的任意2条直线段 l_1 与 l_2，计算其交点 (φ, λ) 的网格权值 $W_i(\varphi, \lambda)$：

$$W_i(\varphi,\lambda)=W_i-1(\varphi,\lambda)+\|l_1\|\cdot\|l_2\|\cdot\sin(2\theta),\quad i=1,2,\cdots,n,\ W_0(\varphi,\lambda)=0 \tag{1}$$

式中：n 为坐标位置 (φ, λ) 下的相交线段组数，图像中同一交点可能存在多组相交线段，相交线段组数越多的交点 (φ, λ) 对应网格权值 $W_i(\varphi, \lambda)$ 越大；θ 为2条直线段小于90°的夹角，通过式(1)可以得出，2条直线段越接近45°，对应交点所被赋予的网格权值更大。当第一个灭点的选择迭代次数达到105时，置信度就可达到0.999 9[12]。通过灭点形成的性质可以得到105×360个3灭点组合，将3个灭点代入评估式(1)计算权值总和，选择权值总和最高的一组灭点方案作为最终的灭点方案。

第二步，对复制的小目标对象选择合理的缩放比例，借助灭点在相机成像中的性质辅助尺寸判断。

以图6为例，AB 是垂直于水平面的一条直线段，现将 AB 平移到 EF 处，可得 AB、EF 是相互平行且等长的直线段，且 AB、EF 垂直于水平面，虚线 AE 与 BF 相互平行。点 V 是平行线 AB、EF 投影到相机成像面 P 上的灭点，即 $A'B'$ 延长线与 $E'F'$ 延长线的交点。点 O 是平行线 AE、BF 投影到相机成像面 P 上的灭点，即 $A'E'$ 延长线与 $B'F'$ 延长线的交点。设 A 点的坐标为 (x, y)，A' 点的坐标为 (x', y')，V 点的坐标为 (x_v, y_v)，设 $\angle A'VE'$ 为 α，$\angle F'E'N$ 为 β，可得图像移动后的大小缩放比例 r_{scale} 为：

$$r_{\text{scale}}=\frac{\sqrt{(x'+x_v)^2+(y'+y_v)^2}}{\sqrt{(x+x_v)^2+(y+y_v)^2}}\cdot\frac{\sin\alpha}{\sin(\pi-\alpha-\beta)} \tag{2}$$

第三步，对选定的目标进行仿射变换。对于任意目标内部的任意一点 $K(k_x, k_y)$，其移动后的坐标为 (k_x', k_y')，用 $\vec{k}$ 表示其在原向量空间中的向量，用 $\vec{k'}$ 表示该点经仿射变换后在另一个向量空间的向量，则尺寸自适应变化的仿射变换过程可表示为：

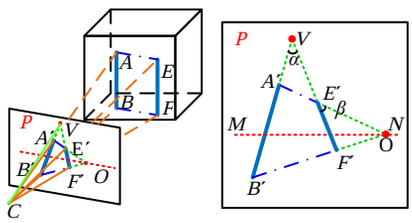


图6　空间目标在平面上投影示意图

Fig. 6　Schematic diagram of space object projection on a plane

$$\vec{k'}=SR\vec{k}+\vec{t} \tag{3}$$

式中：S 为缩放矩阵；R 为旋转矩阵；$\vec{t}$ 为平移向量。该式展开合并后使用齐次坐标表示为：

$$\begin{bmatrix}k_x'\\k_y'\\1\end{bmatrix}=\begin{bmatrix}r_{\text{scale}}\cos\theta & -r_{\text{scale}}\sin\theta & t_x\\ r_{\text{scale}}\sin\theta & r_{\text{scale}}\cos\theta & t_y\\ 0 & 0 & 1\end{bmatrix}\begin{bmatrix}k_x\\k_y\\1\end{bmatrix} \tag{4}$$

第四步，选取放置复制后对象的坐标位置，对合适区域中的任意坐标点，定义被选中的坐标点作为映射落点的评估权值为 W_{p}，则该值的获取为：

$$W_{\text{p}}=W_{\text{area}}\cdot W_{\text{over}} \tag{5}$$

式中：W_{area} 为面积评估值函数。函数 W_{area} 表达式为：

$$W_{\text{area}}=\begin{cases}\dfrac{1}{r_{\text{scale}}\tau}, & r_{\text{scale}}\tau>T\\ \dfrac{1}{T}+\mu\left(T-\dfrac{1}{r_{\text{scale}}\tau}\right), & r_{\text{scale}}\tau\leqslant T\end{cases} \tag{6}$$

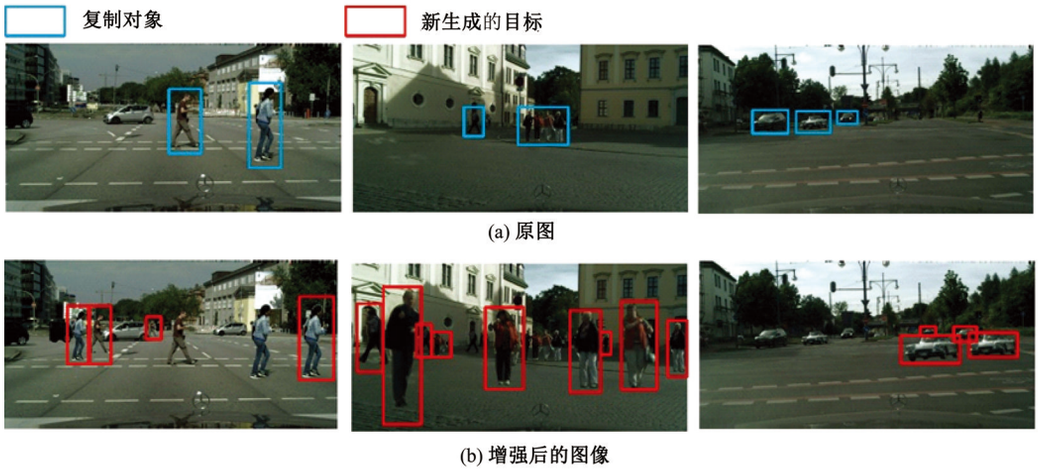
式中：τ 为对象自身所占像素数量与整张图像像素数量的比值，当 $\tau<0.01$ 时，认为该目标为一个小目标；T 为界定小目标对象的阈值；μ 为达到阈值后概率的上升系数。

式(5)中 W_{over} 为覆盖重叠补正函数。函数 W_{over} 表达式为：

$$W_{\text{over}}=\prod_{i=1}^{n}\left(1-\frac{S_{\text{over}}}{S_i}\right)\left(1-\lambda\frac{S_{\text{over}}}{S_{\text{new}}}\right) \tag{7}$$

式中：S_i 为图像中其他存在的可复制目标中第 i 个的面积；S_{over} 为其与新生成对象重叠的面积；S_{new} 为新生成对象的面积；λ 为自身重叠补正系数。

通过落点评估函数可以在同一个数据集中对同一张图片中的目标进行复制，如图7所示，也可以在同一个数据集的不同图片中跨图像选取需要复制的目标对象进行数据增强，如图8所示。

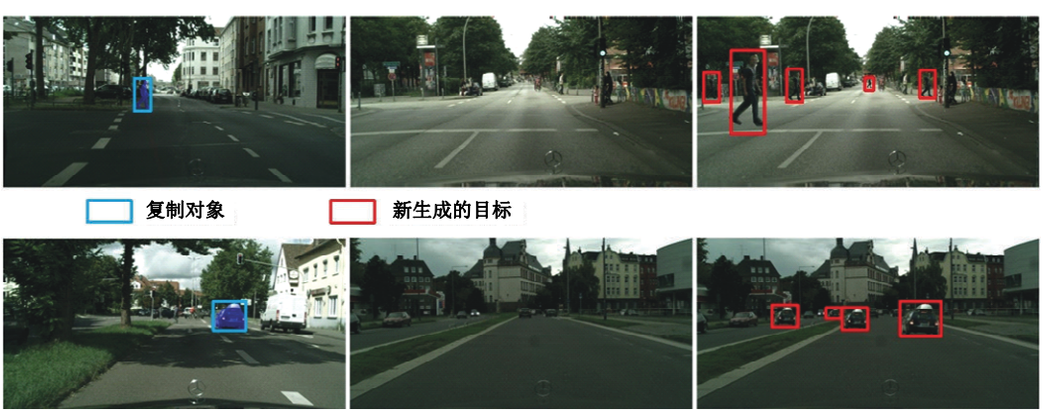


(a) 原图

(b) 增强后的图像

图7 同一图像内的复制数据增强效果

Fig. 7 Effect of intra-image copy data augmentation



(a)含复制对象的图像 (b)待粘贴的图像 (c)增强后的图像

图8 跨图像复制数据增强效果

Fig. 8 Effect of cross-image copy data augmentation

1.4.2 建立Field数据集

本文还建立了信息混杂环境中(如河流中的碎石堆、草地上的土堆、平原上远处的牛群等)的小目标数据集(Field数据集),并采用本文提出的网络算法测试了Field数据集中小目标物体的分割效果。本文选取网络上部分合适的野外环境图片,结合实地拍摄,建立了一套野外地形数据集,并使用Labelme软件进行了标注。该数据集包含2 142张图像,其中包含了山地、林地、草原、海滩、野外公路等场景地形,图像可能包含草地、石地、沙地、泥地、水面等多种地形环境,且大部分单张图像包含复数种类的地形,并含有树木、石块、灌木、动物以及少量人造建筑等阻挡物。这2 142张图像被分为1 714张训练集图像、428张验证集图像和测试集图像。图9为野外地形数据集中的原图与标注图。由于野外环境基本无人造痕迹,故较难获取灭点与消隐线,因此在寻找灭点时对



图9 野外地形数据集部分图片及其标注

Fig. 9 Some pictures of field topography dataset and their annotations

构成灭点的直线设定数量阈值,若没有达到阈值的灭点则对该图像仅使用随机复制增强。

2 实验测试与分析

2.1 实验环境与参数配置

在本实验中,网络模型的硬件配置如下:Intel i5-9400CPU,英伟达RTX2070显卡,8 GB显存,16 GB内存。软件环境如下:Ubuntu18.04操作系统,python环境为python3.10.4,使用py-

torch深度学习框架，版本为1.11.0，CudaToolKit版本为11.3.1。各数据集训练时的输入尺寸统一为1 024×512，训练时的batch size设置为8，按照轮次训练，换算后统一迭代200个epoch。在进行训练时，使用随机梯度下降（Stochastic gradient descent，SGD）优化器。在本实验中，初始学习率设置为0.01，冲量超参数λ设置为0.9，权重衰减系数设置为0.000 5。

2.2　实验结果评估方法

2.2.1　平均交并比mIoU

对于图中像素是否为一个类别的判断，基于像素自身原本的类别，存在4种情况，并存在4种参数TP、FN、FP和TN，分别对应4种情况，如表1所示。

表1　分类结果混淆矩阵及其参数

Table 1　Classification result confusion matrix and its parameters

实际情况	预测结果	
	正例	反例
正例	TP（真正例）	FN（假反例）
反例	FP（假正例）	TN（真反例）

语义分割中交并比IoU是指该类的预测范围和实际范围的交集与其并集之比，可表示为：

$$\mathrm{IoU}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}} \tag{8}$$

而平均交并比mIoU则是所有类的交并比的平均值，可表示为：

$$\mathrm{mIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ij}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ij}-p_{ii}} \tag{9}$$

式中：p_{ij}为将第i类判别为第j类。mIoU用混淆矩阵参数可等效表示为：

$$\mathrm{mIoU}=\frac{1}{k+1}\sum_{i=0}^{k}\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}+\mathrm{FP}} \tag{10}$$

2.2.2　计算量FLOPs与参数量Params

实验中，网络模型的计算量通过浮点运算数（Floating point operations，FLOPs）进行考量，其中浮点运算数为每秒浮点运算次数。网络模型的大小通过参数量（Parameters，Params）进行评估，其中参数量是指网络模型中需要训练的参数总数。

2.3　网络推理精度对比分析实验

表2为不同网络对本文建立的3个数据集进行测试得到的分割精度对比实验数据。将本文中的3个数据集，按比例随机分成训练集与验证集，分别使用本文改进后的网络（MFF-STDC）、SegNet[13]、ENet[14]、BiSeNet、DeepLabV3+（MV2）、Segformer[15]、STDC网络进行训练，并利用验证集进行验证，同时计算各网络mIoU。

表2　不同网络在不同数据集上测试的mIoU

Table 2　mIoU of different networks tested on different datasets

模型	mIoU/%		
	A-Cityscapes	A-IDD	Field
SegNet	44.36	30.72	29.49
ENet	57.93	43.12	38.64
BiSeNet	69.23	54.31	44.91
DeepLabV3+（MV2）	72.14	57.71	47.79
Segformer	71.32	56.13	47.12
STDC	71.85	56.34	46.84
本文网络（MFF-STDC）	75.86	59.99	49.78

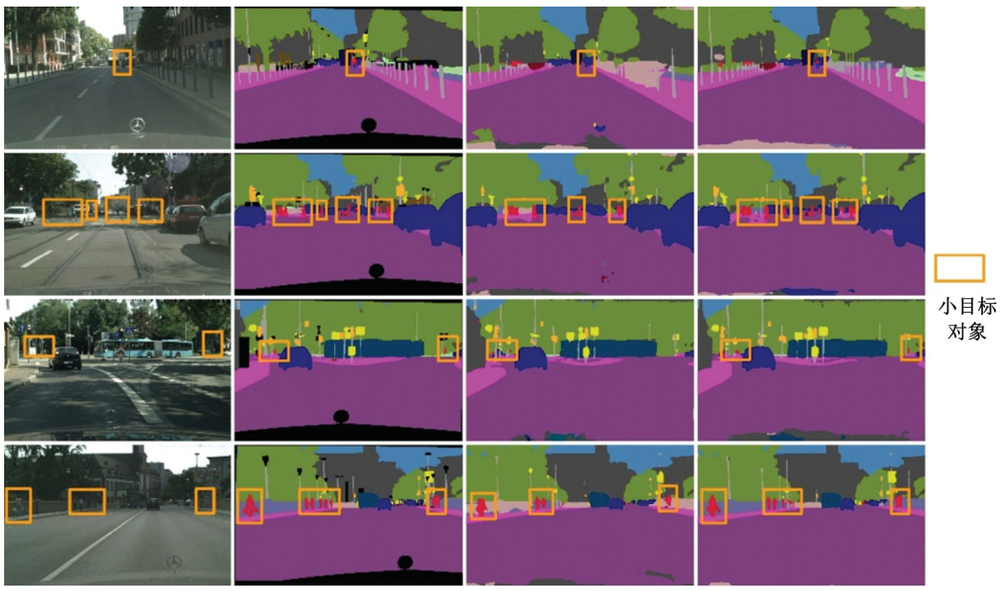
从表2的信息中可以看出，本文网络（MFF-STDC）在3个数据集上都有着良好表现，在A-Cityscapes数据集、A-IDD数据集和Field数据集上的MIoU都高于其他测试网络，分别达到了75.86%、59.99%和49.78%。对比STDC网络，本文网络分割精度有了极大进步，分别提升了4.01个百分点、3.65个百分点和2.94个百分点。

对于小目标类别的物体，MFF-STDC同样取得了良好的结果。在A-Cityscapes数据集中，目标较小的自行车、行人、交通信号灯、交通标识类别的IoU与STDC相比分别从72.32%、75.59%、60.35%、71.44%提升到了76.01%、82.88%、65.33%、73.51%；在A-IDD数据集中，目标较小的行人、骑手、摩托车类别的IoU与STDC相比分别从57.36%、69.37%、68.28%提升到了63.52%、74.75%、73.79%；在Field数据集中，目标较小的土堆、石头类别的IoU与STDC相比分别从63.98%、56.57%提升到了66.33%、62.14%。

图10～12为MFF-STDC与STDC的分割效果，可以明显看出MFF-STDC对小目标类别的分割结果更接近原始标注图，且分割出的小目标物体数量更多，尤其是在室外环境最复杂的Field数据集中，分割效果最明显，因此进一步证明了MFF-STDC对小目标类别的分割精度有极大的提升。

2.4　网络模型大小对比分析实验

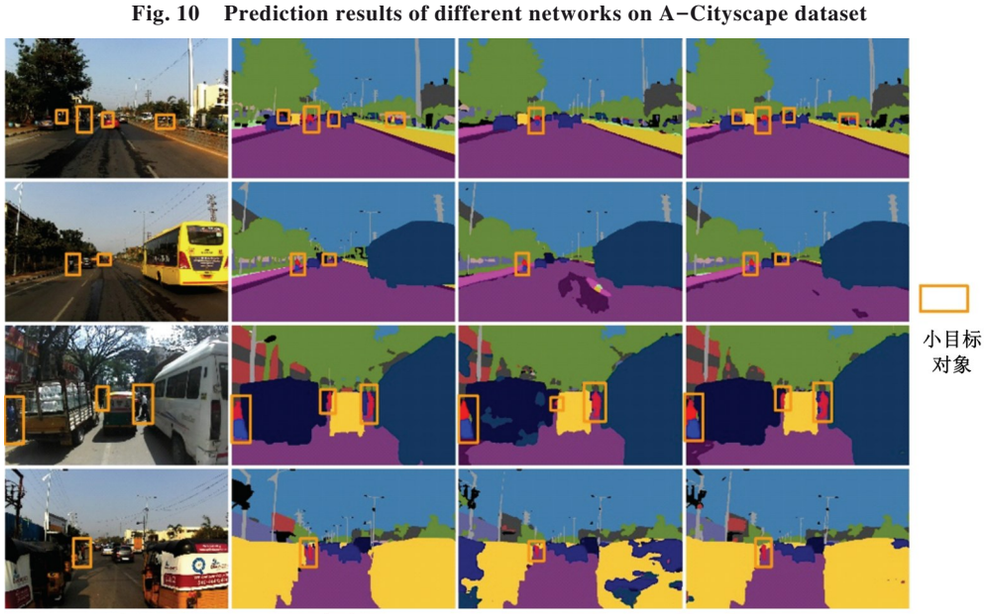
为比较本文模型与其他模型的精度与模型大



(a) 原图 (b) 标注 (c) STDC (d) MFF-STDC

图10 不同网络在A-Cityscapes数据集上的预测结果

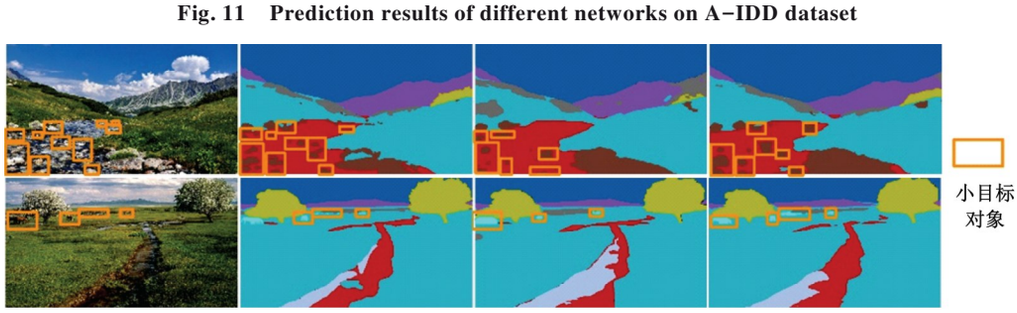
Fig. 10 Prediction results of different networks on A-Cityscape dataset



(a) 原图 (b) 标注 (c) STDC (d) MFF-STDC

图11 不同网络在A-IDD数据集上的预测结果

Fig. 11 Prediction results of different networks on A-IDD dataset



(a) 原图 (b) 标注 (c) STDC (d) MFF-STDC

图12 不同网络在Field数据集上的预测结果

Fig. 12 Prediction results of different networks on Field dataset

小，根据前文已训练完成的模型，对测试集进行推理，得到了不同网络的平均交并比mIoU、像素精度picAcc、参数量Params。由于模型大小不受使用的数据集影响，因此本实验选用A-Cityscapes数据集为代表。

本文网络（MFF-STDC）、SegNet、ENet、BiSeNet、DeepLabV3+（MV2）、Segformer、STDC网络在A-Cityscapes数据集上的精度及模型大小如表3所示。

表3　不同网络的精度和模型大小

Table 3　Acuracy and model sizes of different network

模型	mIoU/%	picAcc/%	Params/M
SegNet	44.36	59.32	29.5
ENet	57.93	66.73	0.4
BiSeNet	69.23	78.65	59.24
DeepLabV3+（MV2）	71.64	83.82	45.57
Segformer	70.32	81.14	3.72
STDC	71.85	81.32	7.08
本文网络（MFF-STDC）	75.86	83.87	5.43

从表3中可以看出，在精度方面，无论是平均交并比mIoU还是像素精度picAcc，本文网络（MFF-STDC）都取得了最高的精度数值，即75.86%与83.87%。而在模型大小方面，虽然ENet参数量很少，但MFF-STDC的分割精度远高于ENet。而使用了Transformer机制的Segformer网络参数量略小于MFF-STDC，但分割精度比MFF-STDC低。与其他网络相比，MFF-STDC网络参数量都有一定程度的减少，但分割精度均有较大幅度提高，这点从前文精度实验中也可证明。因此，本文方法（MFF-STDC）相较于其他方法在精度和参数量上进行综合对比评估有明显的优越性。

2.5　消融实验

为验证各改进模块的具体效果，以消融实验的方式测试各改进模块的实际影响。将本文方法分成6组进行训练，改进的模块包括DW-STDC模块、特征提取模块CA-FFM、分层权重的注意力优化模块HAM、多感受野的信息结合模块MIC。其中，第1组为原始的STDC模型实验，第2～5组分别将每一种改进模块与STDC结合进行效果实验，第6组将所有改进模块与STDC结合进行实验，即本文的MFF-STDC方法实验。实验效果以A-Cityscapes数据集为例，不同改进模块对预测精度、参数量以及推理速度的影响如表4所示。

表4　消融实验

Table 4　Ablation experiment

实验组别	DW-STDC	CA-FFM	HAM	MIC	mIoU/%	Params/M	FLOPs/G
Exp 1	×	×	×	×	71.85	8.57	16.95
Exp 2	√	×	×	×	72.67	5.36	13.72
Exp 3	×	√	×	×	72.31	8.576	16.953
Exp 4	×	×	√	×	72.09	8.561	16.95
Exp 5	×	×	×	√	72.35	8.661	17.702
Exp 6	√	√	√	√	75.86	5.43	14.48

从表4中结果可以看出，各改进模块对于最终的精度都有一定程度的提升。其中，DW-STDC选用了精度提升较高的一种方案（1.2提到的将Stage3～Stage5阶段中模块数确定为4、5、3），与STDC模型对比，mIoU提升了0.82个百分点。CA-FFM和HAM并没有增加太多网络的计算量，但精度却分别提升了0.46个百分点和0.24个百分点，可以看出这2个模块有着较高的性价比。MIC模块为新添加的模块，一定程度上增加了计算量的负担，但对于小目标类别的信息特征提取，MIC模块有着重要的作用，且也换取了0.5个百分点的精度提升。可以看出，每个改进模块都对网络分割精度提升起到了较大的作用。将所有改进模块与STDC结合得到MFF-STDC，对其进行测试，可以看出，MFF-STDC与STDC相比，mIoU有了极大的提升，且参数量与计算量都有一定程度的降低。

对DW-STDC模块在Stage3～Stage5阶段中的模块数设定，进行对照实验。输出通道通常为4的倍数，所以设置$C_{out}/16$、$C_{out}/4$、C_{out} 3种分组数进行实验。对于模块层数的设定，以小样本数据量进行训练，递增或递减逐级调整各阶段的模块数并观察网络模型的精度变化，最后选取精度变化最明显的模块层数进行正常样本量训练，并比对训练结果。表5为不同卷积分组数与模块层数的模型参数量与精度结果，由表5可知，将Stage3～Stage5阶段中模块数确定为4、5、3，卷积分组数为$C_{out}/16$，可使得参数量（Params）比原始模块减少3.21 M，计算量FLOPS减少3.23 G，分

表5 不同分组数与层数的模型参数量与精度

Table 5 Model parameter quantity and precision under different grouping numbers and layers

分组数	各阶段模块层数			参数量	计算量	精度
	Stage3	Stage4	Stage5	Params/M	Flops/G	mIoU/%
1	2	2	2	8.57	16.95	71.85
$C_{out}/16$	2	2	2	4.34	12.03	71.33
$C_{out}/16$	4	5	3	5.36	13.72	72.67
$C_{out}/16$	7	9	5	7.08	16.26	73.02
$C_{out}/4$	1	2	2	4.43	12.20	70.89
$C_{out}/4$	2	4	3	5.51	14.08	71.98
$C_{out}/4$	2	8	6	7.35	16.90	72.42
C_{out}	1	2	2	4.34	12.03	69.83
C_{out}	2	4	3	5.36	13.72	71.04
C_{out}	2	8	6	7.08	16.26	71.78

割精度提高0.82个百分点。

3 实际环境实验

3.1 实景测试系统

为测试本文方法在实际应用中的有效性,本文搭建了实景测试系统,用于测试本文方法在真实环境下的效果。

如图13(a)所示,测试系统由TurtleBot移动机器人底盘、笔记本电脑、Microsoft Kinect V1相机等组成。基于笔记本电脑的程序驱动,TurtleBot移动机器人可在室外场景移动,并使用相机以最高30张/秒的速度拍摄图像。图13(b)为该测试系统工作现场。



图13 实景测试系统与测试系统工作现场

Fig. 13 Actual scenario testing system and testing system worksite

3.2 实景实验结果

本文使用该测试系统对周边室外环境进行实景测试,并使用本文网络模型(MFF-STDC)与STDC网络模型对结果进行测试,结果如图14所示。

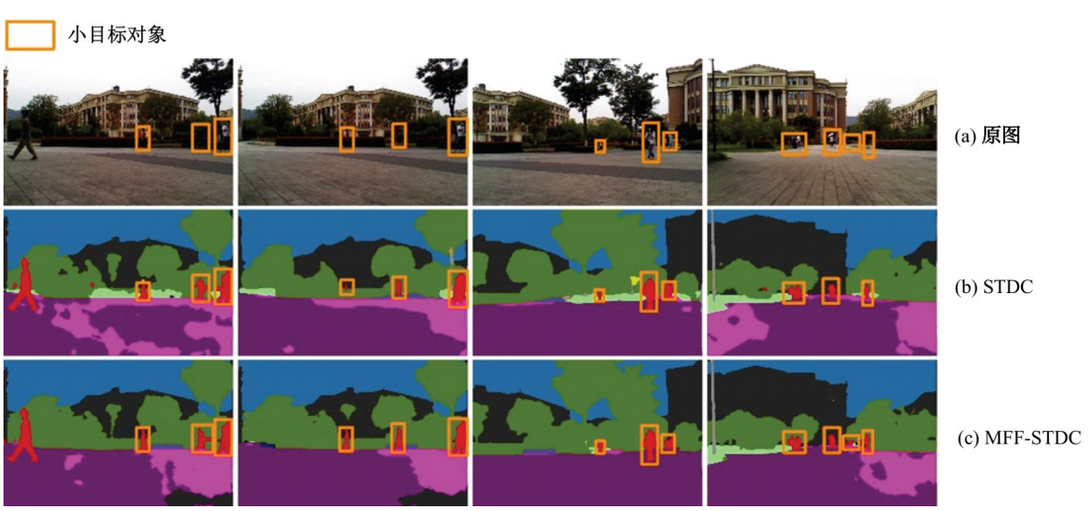


图14 实际环境测试结果

Fig. 14 Actual environment test results

从实际环境测试结果可以看出，传统轻量化语义分割网络STDC的分割结果仍较为粗糙，对于图像中的树木、建筑、道路等较大的背景类别，有较为分明的分割结果，但仍然出现了将建筑区域识别为树木、同类道路分类混淆的情况。对于小目标对象，如图14中的行人，只能部分识别，且出现了2个行人被标记为一团的情况，稍近处的行人轮廓也不清晰，远处的行人甚至有遗漏未标记的情况，在实际应用中可能会导致事故发生。而本文网络（MFF-STDC）的预测结果有着更好的表现，背景类别分割更加精细且更少误判，对于每个小目标对象都做到了成功检测，且有着更为精细的轮廓。可以看出，MFF-STDC在应用了多尺度感受野与多层次注意力机制后，对于大尺度的背景类别与小尺度的行人对象等类别的识别都能有效提升检测效果，而针对小目标与占比较少类别的数据增强方法更是让学习后的网络对于小目标有着更敏锐的感知能力，更不易造成漏判，且能得到更加精细的分割效果。

4　结束语

本文针对复杂环境中分割不精确、小目标物体识别能力差、分割边缘不准确等问题，提出基于多层级特征融合的MFF-STDC。在特征提取中用分组卷积替换常规卷积，并叠加更多特征提取模块，使得在不增加运算量的情况下提升网络的特征提取能力。利用多层级之间的交叉融合，提升小目标物体的信息获取能力。增加MIC模块，强化对特征细节的提取。利用CA机制改善语义特征与细节特征的融合结果。对使用了自适应复制算法进行数据增强的A-Cityscapes数据集、A-IDD数据集和自建的Field数据集进行训练与测试，实验表明，MFF-STDC对3个数据集测试的mIoU分别达到了75.86%、59.99%和49.78%，与其他网络分割精度相比有明显提高。通过消融实验测试了各改进模块对MFF-STDC网络模型在分割精度上有一定程度的提升。搭建实景实验平台，通过实景测试证明了MFF-STDC在小目标类别识别效果和网络整体分割精度上都比STDC有着显著的提升，实现了在保证实时性的情况下，达到更好的目标检测效果。未来将进一步扩充野外小目标样本数据集，并加强在野外环境与恶劣天气中的小目标语义分割与识别效果。

参考文献：

[1] 张艳，张明路，吕晓玲，等. 深度学习小目标检测算法研究综述[J]. 计算机工程与应用，2022，58(15)：1-17.
Zhang Yan，Zhang Ming-lu，Lyu Xiao-ling，et al. Review of research on small target detection based on deep learning[J]. Computer Engineering and Applications，2022，58(15)：1-17.

[2] Takikawa T，Acuna D，Jampani V，et al. Gated-SCNN：Gated shape CNN for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway，NJ：IEEE，2019：5229-5238.

[3] Chen L C，Papandreou 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.

[4] 杨玉敏，廖育荣，林存宝，等. 轻量化卷积神经网络目标检测算法综述[J]. 舰船电子工程，2021，41(4)：31-36.
Yang Yu-min，Liao Yu-rong，Lin Cun-bao，et al. A survey of object detection algorithms for lightweight convolutional neural networks[J]. Ship Electronic Engineering，2021，41(4)：31-36.

[6] Yu C，Wang J，Peng C，et al. BiSeNet：bilateral segmentation network for real-time semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich：IEEE，2018：334-349.

[5] Fan M，Lai S，Huang J，et al. Rethinking bisenet for real-time semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway，NJ：IEEE，2021：9716-9725.

[7] Hou Q，Zhou D，Feng J. Coordinate attention for efficient mobile network design[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway，NJ：IEEE，2021：13708-13717.

[8] Ioannou Y，Robertson D，Cipolla R，et al. Deep roots：improving CNN efficiency with hierarchical filter groups[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway，NJ：IEEE，2017：5977-5986.

[9] 霍光，林大为，刘元宁，等. 基于多尺度特征和注意力机制的轻量级虹膜分割模型[J]. 吉林大学学报：工学

版, 2023, 53(9): 2591-2600.
Huo Guang, Lin Da-wei, Liu Yuan-ning, et al. Lightweight iris segmentation model based on multi-scale feature and attention mechanism[J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(9): 2591-2600.

[10] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, N J: IEEE, 2016: 3213-3223.

[11] Varma G, Subramanian A, Namboodiri A, et al. IDD: A dataset for exploring problems of autonomous navigation in unconstrained environments[C] // Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV). Piscataway, N J: IEEE, 2019: 1743-1751.

[12] Shi Q, Liu M, Li S, et al. A deeply supervised attention metric-based network and an open aerial image dataset for remote sensing change detection[J]. IEEE Transactions on Geoscience and Remote Sensing, 2021, 60: 1-16.

[13] 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, 39 (12): 2481-2495.

[14] Paszke A, Chaurasia A, Kim S, et al. ENet: a deep neural network architecture for real-time semantic segmentation[J]. Arxiv Preprint, 2016, 6: No. 160602147.

[15] Xie E, Wang W, Yu Z, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.