

基于八度卷积设计的实时语义分割网络

王鑫, 吴开军*

上海海洋大学信息学院, 上海 201306

摘要 卷积神经网络进行实时图像语义分割时, 处理大片色彩变化较小的像素区域时会存在计算的空间冗余, 同时轻量级网络进行特征提取时的精度较低。针对上述两个问题, 利用改进后 MobileNet v3 和轻量高频采样 (OTCH-L) 模块设计了一种实时语义分割网络。首先利用 hard-swish 激活函数对轻量级网络 MobileNet v3 进行精度补偿, 提出改进后的 MobileNet v3 特征提取网络; 其次基于八度卷积设计 OTCH-L 模块, 解决空间冗余问题, 在保证精度的同时, 降低模型计算量。在 Pascal VOC2012 数据集上进行训练实验, 并在 Pascal VOC2007 数据集上对模型进行验证。实验结果表明, 所提模型分割速度达到 25.94 frame/s, 分割平均交并比 (MIoU) 达 70.34%。相比主流语义分割模型 SegNet、PSPNet、DeepLab v3 plus, 所设计模型在维持分割精度的前提下, 分割速度有显著提升。

关键词 图像处理; 实时语义分割; 八度卷积; 轻量级; MobileNet v3

中图分类号 TP391.4 文献标志码 A

doi: 10.3788/LOP202259.0810015

Real-Time Semantic Segmentation Network Based on Octave Convolution

Wang Xin, Wu Kaijun*

College of Information Technology, Shanghai Ocean University, Shanghai 201306, China

Abstract When a convolutional neural network performs real-time image semantic segmentation, processing large blocks of pixels with small color changes leads to computational spatial redundancy. Also, the accuracy of feature extraction using lightweight networks is low. We designed a real-time semantic segmentation network using the improved MobileNet v3 and lightweight OctConv high-frequency (OTCH-L) module to mitigate both problems. First, the hard-swish activation function was used to compensate for the accuracy of the lightweight network MobileNet v3. Then, we proposed an improved MobileNet v3 feature extraction network. Furthermore, we designed the OTCH-L module based on the octave convolution to solve the problem of spatial redundancy and reduce the computational size of the model while ensuring accuracy. The models were trained and verified on the Pascal VOC2012 and VOC2007 datasets, respectively. The experimental results show that the segmentation speed of the proposed model reaches 25.94 frame/s, and the mean intersection over union (MIoU) reaches 70.34%. Compared with the mainstream semantic segmentation models, such as SegNet, PSPNet, and DeepLab v3 plus, our proposed model significantly enhances the segmentation speed while maintaining segmentation accuracy.

Key words image processing; real-time semantic segmentation; octave convolution; lightweight; MobileNet v3

1 引言

图像的语义分割是计算机视觉领域中最具挑

战的任务。它利用图像的高级语义信息对图像的每一个像素进行分类, 并且分配一个类别标签^[1]。

该项技术在医疗图像、遥感测绘、自动驾驶、室内导

收稿日期: 2021-02-23; 修回日期: 2021-04-16; 录用日期: 2021-05-06

通信作者: *walt@shou.edu.cn

航等领域都有着重要的研究意义。随着深度学习概念的提出,以往的图论和像素聚类方法难以解决的精度和速度问题都得到了有效解决。加之 GPU 行业的迅猛发展,更复杂的卷积神经网络模型得以实现并应用,语义分割技术也在不断改进提升。

Long 等^[2]于 2015 年提出全卷积网络(FCN)。FCN 具有端到端的卷积神经网络体系结构,利用 VGG16 网络,在没有任何全连接层的情况下进行了密集的预测。FCN 首次将卷积神经网络成功地应用在了语义分割任务中,使语义分割工作的精度从图像级上升到了像素级。Badrinarayanan 等^[3]于 2017 年提出 SegNet 模型。SegNet 首次应用编码器-解码器的结构设计。编码器由 13 个卷积层和 5 个最大池化层组成,并且不使用全连接层,在最大池化时存储对应最大池化的索引;而解码器则利用最大池化层的索引进行上采样。SegNet 中编码器结构为 VGG16 网络,虽然增加深度有利于神经网络提取特征,但是网络深度达到一定程度时,深层网络总体表现不如浅层网络。Zhao 等^[4]于 2017 年提出 PSPNet。该网络采用金字塔池化层融合来解决全局信息缺失的问题,增大了感受野,很好地处理了图像场景中的上下文关系,丰富了特征信息。传统 PSPNet 模型经 ResNet 提取特征后,进行金字塔池化融合操作。该操作通过不同尺度的感受野融合图像区域中局部的上下文信息与全局的上下文信息,使图像的特征信息更加丰富。然而传统的 PSPNet 进行特征提取时采用 ResNet,因此在实际应用中分割速度较慢,难以满足实时分割的需求。Chen 等^[5]在 2018 年提出了 DeepLab 网络。DeepLab 以全卷积神经网络为基础,在末端添加全连接条件随机场(CRF)。DeepLab 模型首先使用双线性插值法对全卷积网络的输出结果进行上采样,得到粗略的分割结果,然后对分割结果中的每个像素点构造 CRF 模型并且求解,以提升分割精度。蔡雨等^[6]在 2020 年提出了一个基于特征融合的实时语义分割网络。该网络输出深层高级别语境信息,并引入注意力优化机制,对两路输出特征图进行多尺度融合。虽然该网络利用注意力优化机制代替上采样,优化网络的输出,但在速度上仍有提升空间。

目前卷积神经网络在处理图像语义分割问题时仍有两个方面待提升:一方面在进行卷积操作时,处理大块较为稳定的像素区块信息时会存在过多空间冗余;另一方面为了使模型更加轻量化,仅

利用轻量级提取网络,得到的特征图精度不足,难以满足分割要求。本文针对卷积计算中的空间冗余问题,提出了一种基于八度卷积^[7]设计的轻量级高频采样(OTCH-L)模块;针对轻型特征提取网络精度不足问题,利用 hard-swish 激活函数^[8]替换 MobileNet v3 网络^[9]最后一层激活函数进行精度补偿,并提出一种改进的 MoblieNet v3 网络。所提网络首先利用改进后 MobileNet v3 进行特征提取,然后利用 OTCH-L 模块提取高频特征图,最后对高频特征图与最大池化后的结果和通道调整 1×1 卷积后的结果进行通道叠加,得到最终分割结果。整体网络达到较高的分割精度和分割速度。

2 所提网络内容

2.1 八度卷积

图像信号可分为高频通道和低频通道,高频通道包含细节信息,低频通道体现图像大体信息。利用卷积进行计算时,低频通道会造成空间冗余。八度卷积可以有效解决卷积计算中的空间冗余问题,因此利用八度卷积代替传统卷积神经网络,将特征图分为高频通道和低频通道^[10],如图 1 所示。

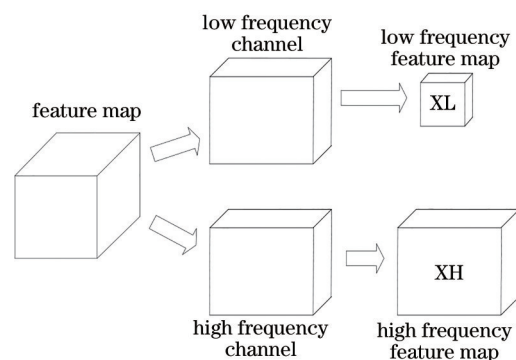


图 1 特征图高频部分和低频部分示意图

Fig. 1 High frequency and low frequency of feature map

OctConv^[7]中间层的计算如图 2 所示,将特征图(X)分成高频分量(XH)和低频分量(XL)。对 XH 进行卷积和平均池化操作,得到 YH,对 XL 进行上采样和卷积操作,得到 YL,最终输出特征图 YH 和 YL。

OctConv 的优势为计算量少,设置参数 α 为低频通道占比。假设使用普通卷积计算得到的输入尺寸为 $C_{in} \times W \times H$,卷积尺寸为 $C_{out} \times K \times K$,则输出尺寸为 $C_{out} \times W \times H$ 。考虑到输出特征图中每个数据都需要进行 $C_{in} \times K \times K$ 次计算,总计算量可表示为

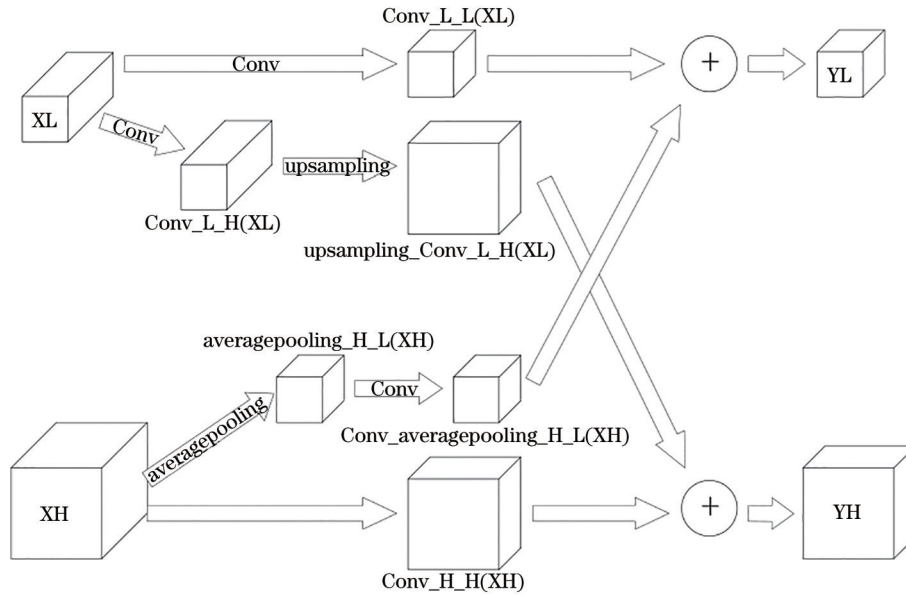


图 2 OctConv 操作示意图

Fig. 2 Schematic of OctConv operation

$$S_{Conv} = (C_{out} \times W \times H) \times (C_{in} \times K \times K) \quad (1)$$

使用 OctConv 时,共有 4 个卷积操作,它们的计算量之和表示为

$$S_{OctConv} = \frac{\alpha^2 + 2\alpha(1 - \alpha) + 4(1 - \alpha)^2}{4} \times S_{Conv} \\ = \frac{3(\alpha - 1)^2 + 1}{4} \times S_{Conv} \quad (2)$$

理论上,当 $\alpha = 1$ 时,计算量为普通卷积的 1/4,此时模型计算量最少。但在实际应用中,为了保证模型在一定分割精度下同时获得较快的分割速度,应选取适当的 α 值。当 OctConv 提取低频分量占比 $\alpha = 0.3$ 时效率最高。

2.2 swish 函数和 hard-swish 函数

语义分割任务中使用轻量级网络 MobileNet v3 进行特征提取时,网络分割速度大幅增加,但精度会不可避免地下降。因此防止精度下降过于严重,在 MobileNet v3 深层引入 hard-swish 函数进行精度补偿。

hard-swish 函数的出现是受到 swish 激活函数的启发,swish 定义为

$$swish(x) = x \cdot \text{sigmoid}(x), \quad (3)$$

其中,

$$\text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (4)$$

谷歌团队在 MNIST 数据集上进行测试,改变带有 swish 和 ReLU 两种激活函数的全连接网络的层数,观察两种激活函数的性能变化^[11],结果如图 3

所示。在 42 层以内的情况下,两种激活函数性能基本一致,而在 40~50 层中,swish 函数的性能比 ReLU 函数好。所以在深层网络中,swish 函数能提供更好的精度。

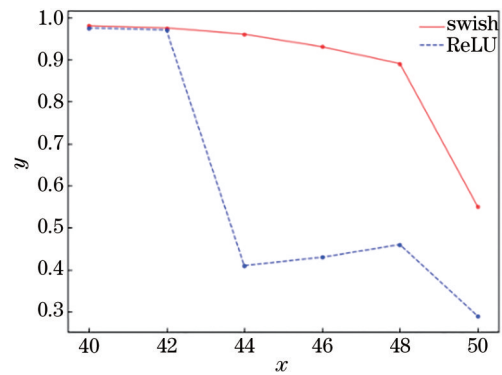


图 3 swish 与 ReLU 对比

Fig. 3 Comparison between swish and ReLU

在轻量级模型中,计算 sigmoid 函数的代价较大。为了减小计算成本,使用与 swish 性能相当的 hard-swish 函数,表达式为

$$\text{hard-swish}(x) = x \frac{\text{ReLU}_6(x + 3)}{6} \quad (5)$$

ReLU_6 表示为

$$\text{ReLU}_6(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x \leq 6 \\ 6, & x > 6 \end{cases} \quad (6)$$

ReLU_6 是普通的 ReLU 函数,限制最大输出值为 6。该函数在移动端设备上的运算精度为 float16

时,也能有很好的数值分辨率^[12]。图 4 为 swish 和 hard-swish 函数对比曲线。

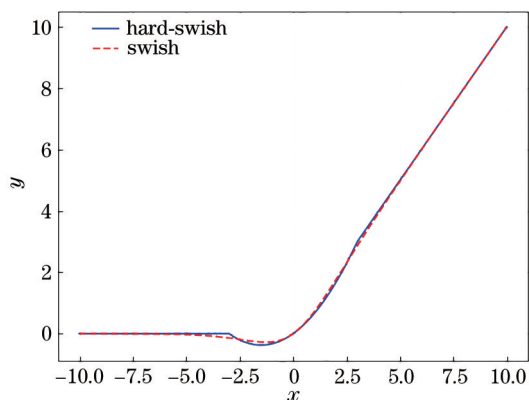


图 4 swish 函数和 hard-swish 函数对比曲线图

Fig. 4 Comparison between swish function and hard-swish function

由图 4 可看出, swish 和 hard-swish 函数基本一致,但 swish 中包含 sigmoid 计算,会消耗较多计算资源。hard-swish 函数中只有 ReLU₀ 函数计算,所利用计算资源较少。所以所提模型利用 hard-swish 函数对 MobileNet v3 最后一层激活函数进行修改,利用改进后的 MobileNet v3 作为特征提取网络。

2.3 所提模型

2.3.1 OTCH-L 模块

受到深度可分离卷积和残差网络的启发,设计了一个加入瓶颈块的高频八度卷积模块来减少网络参数量和计算量。加入瓶颈块的目的是降低维度,增加通道间信息的可交互性。图 5 为采样块结

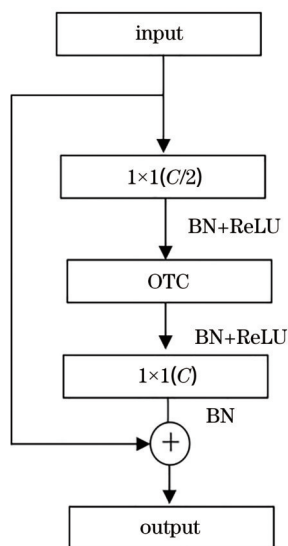


图 5 OTCH-L 模块

Fig. 5 OTCH-L module

构。采样块首先利用 1×1 的瓶颈块把特征图通道数降为原来的 $1/2$; 然后通过 OTC 卷积层进行八度卷积操作, 获取 $\alpha=0.3$ 的高频特征图; 最后通过 1×1 的瓶颈块升维, 并与 input 进行通道融合, 达到采样高频特征的目的。图 5 中 OTC 表示 OctConv 模块, 1×1 代表瓶颈块, BN 表示批归一化操作, ReLU 为激活函数, C 为通道数。

2.3.2 所提网络设计

所提网络利用改进后的 MobileNet v3 进行特征提取后, 得到特征图。在 DeepLab v3 的启发下, 将后续网络模型设计为三路进行并行传播^[13]: 第一路为池化层, 利用最小工作量和参数量来获取特征图中的高阶特征; 第二路为 OTCH-L 模块, 利用特征图中高频部分得到的特征来丰富结果中的细节; 第三路通过 1×1 卷积产生 logits。对经过池化和高频采样处理后的结果进行双线性插值, 得到上采样处理想要的分辨率。最后对所有得到的结果进行通道叠加 (concat), 得到了输出结果。整体网络结构如图 6 所示。

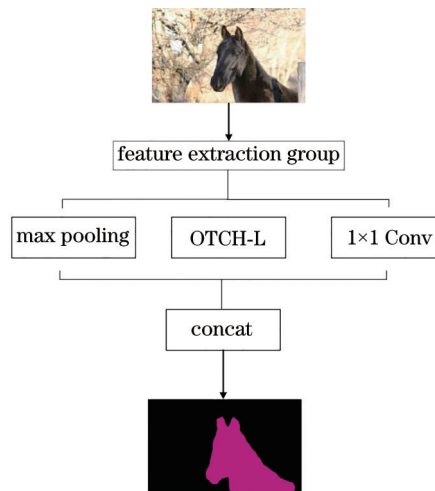


图 6 所提模型

Fig. 6 Proposed model

3 实验与分析

3.1 实验环境

基于 Python 编程环境, 在 PyTorch 框架下搭建实验模型。算法实现使用 Windows 10 系统、Python 版本 3.6、PyTorch 版本 1.2.0、CUDA 版本 10.0、CUDNN 版本 7.0。硬件参数为 i5-9600KF 处理器、GPU 为 RTX2070 Super、显存为 8 GB、运行内存为 32 GB。

3.2 实验数据集

主要在 Pascal VOC2012 数据集^[14]上进行训练和验证, VOC2012 数据集中有 20 个分类, 包含背景共 21 个类别, 如表 1 所示。

表 1 数据集包含类别

Category	Content
Person	person
Animal	bird, cat, cow, dog, horse, sheep
Vehicle	plane, bicycle, boat, bus, car, motorbike, train
Indoor	bottle, chair, dining table, potted plant, sofa, tv /monitor

VOC2012 数据集共包含 17125 张图片, 但并不会全部使用, 抽取适用于分割任务的 2913 张图片组成训练集和验证集, 并打乱顺序, 以 80% 训练集和 20% 验证集来训练模型。

3.3 模型训练

训练模型主要是基于 PyTorch 框架和 OpenCV 视觉库的。PyTorch 是动态框架, 在序列计算过程中, 可以很方便地计算参与的变量对于目标函数的梯度, 易于实现网络的反向传播, 相比于其他框架, 更适用于所提模型。并且 PyTorch 能很简单地运行在 GPU 上, 使用 .cuda() 方法即可利用 GPU 进行训练计算。

在实验过程中, 最终输入模型的超参数如表 2 所示。

表 2 模型超参数

Parameter	Value
Batch size	8
Input image size	473×473
Epoch	100
Learning rate	1×10 ⁻⁶

训练模型时, 先将所有训练集输入网络模型中, 规定 batch size 为 8, 随机 8 张图片为一组, 对随机的 8 张图片按照随机的尺度 (0.5~1.5 范围) 进行放缩; 然后再按照 1/2 的概率对全部图片进行左右翻转变换的处理; 最后将所有训练图片统一裁剪成 473×473 的大小输入到网络中。放缩、旋转与裁剪处理使训练样本有更多尺寸和形状, 更加具有随机性, 避免模型过拟合现象。

3.4 评价标准

采用的语义分割评价标准有两种: 类别均像素精度 (MPA) 和平均交并比 (MIoU)。

类别均像素精度 (MPA): 假设数据集中有 $K+1$ 个类别, 分别计算每个类被正确分类的像素数占总像素数的比例, 并累加求平均, 公式为

$$P_{MPA} = \frac{1}{K+1} \sum_{i=0}^K \frac{T_{P_i}}{T_{P_i} + F_{P_i}}, \quad (7)$$

式中: T_{P_i} 为分类正确的像素点个数; F_{P_i} 为分类错误的像素点个数。

平均交并比 (MIoU): 模型对每一类预测的结果和真实值的交集与并集的比值, 求和再平均的结果, 表达式为

$$P_{MIoU} = \frac{1}{K+1} \sum_{i=0}^K \frac{N_{TP}}{N_{FN} + N_{FP} + N_{TP}}, \quad (8)$$

式中: N_{TP} 代表预测正确, 并且预测结果是正类, 真实结果是正类; N_{FP} 代表预测错误, 并且预测结果是正类, 真实结果是负类; N_{FN} 代表预测错误, 并且预测结果是负类, 真实结果是正类^[15]。MIoU 关系如图 7 所示, 其中 N_{TN} 代表预测正确, 并且预测结果是负类, 真实结果是负类。

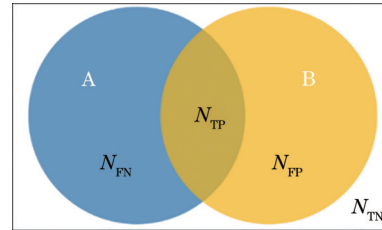


图 7 MIoU 示意图

Fig. 7 Schematic of MIoU

3.5 实验结果

3.5.1 低频占比 α 对整体性能影响探究

OctConv 卷积的核心思想是减少邻域像素之间的空间冗余, 利用特征空间中的频率高低来划分高频区和低频区, 并且按照一定比例进行采样输出。在 OctConv 中只有一个全局的超参数 α , 即特征图低频部分的占比。在同样的训练设置下进行, 利用 ImageNet 对在替换 OctConv 后的 ResNet 和 MobileNet v3 网络的性能指标进行对比。实验使用不同 α 值的 OctConv 对网络中卷积进行替换, 对比修改后的两个网络的精度和速度, 得到的结果如表 3 所示。其中 R+OctConv 代表 ResNet-50+OctConv, M+OctConv 代表 MobileNet v3+OctConv。

图 8 为实验结果可视化展现。从图 8 可以看

表 3 不同 α 值对模型性能的影响

Table 3 Influence of different α values on model performance

Network	α	Accuracy / %	Speed / (frame · s ⁻¹)
ResNet-50		74.5	8.1
R+OctConv	0	74.7	8.9
R+OctConv	0.3	76.2	11.3
R+OctConv	0.6	75.5	14.1
R+OctConv	0.9	73.1	17.5
MobileNet v3		66.1	15.3
M+OctConv	0	66.8	15.6
M+OctConv	0.3	68.3	19.2
M+OctConv	0.6	67.1	22.3
M+OctConv	0.9	65.7	25.4

出, 当将 ResNet-50 和 MobileNet v3 中的普通卷积替换为 OctConv 后, 在 $\alpha=0.3$ 时, 模型的准确率趋于最高, 并且速度也相对全部为高频通道时的输出有所提高。由于所提模型解决的是图像分割问题, 处理此类问题需要兼顾分割准确率和分割速度, 所以在设计的 OTCH-L 模块中, OctConv 的低频通道占比 α 取值为 0.3。

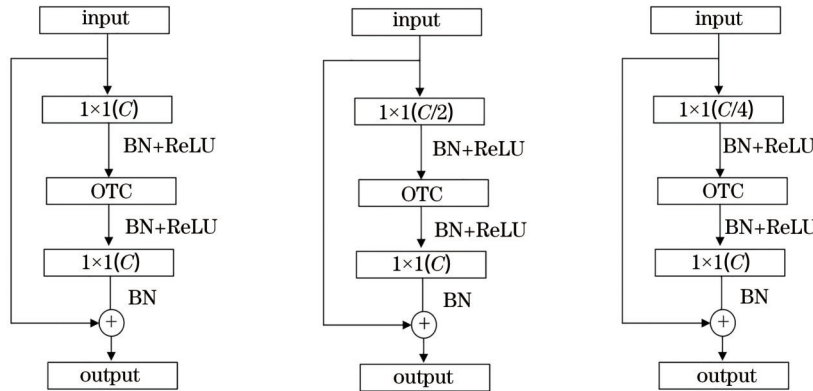


图 9 不同降维通道数的残差模块示意图

Fig. 9 Schematic of residual modules with different dimensionality reduction channels

表 4 为不同通道降维数瓶颈块对 OTCH-L 模块性能影响的对比实验结果。未进行降维操作的采样模块的 MIoU 为 71.54%, 参数量为 756×10^3 , 速度为 22.86 frame/s。相比未进行降维操作的采样模块, 降维到原通道数 1/2 的采样模块的 MIoU 降低 1.20 个百分点, 参数量降低 24.5%, 速度提升

3.5.2 采样模块通道降维探究

为了提升轻型高频采样模块的效率, 对采样模块内部进行通道降维, 但不同降维数会对模块性能产生影响。实验分别测试模块中不同通道数的瓶颈块对分割准确率、参数量以及速度的影响。图 9 为实验所用三种通道数不同的模块结构示意图。

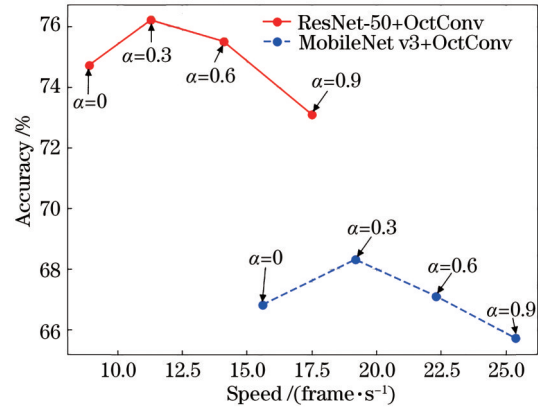


图 8 α 对网络性能的影响

表 4 不同通道降维数瓶颈块间性能的对比

Table 4 Performance comparison between neck blocks with different dimensionality reduction channels

Module	MIoU / %	Parameter / 10^3	Speed / (frame · s ⁻¹)
OTCH-L(C)	71.54	756	22.86
OTCH-L(C/2)	70.34	571	25.94
OTCH-L(C/4)	67.01	477	26.42

割精度和速率的平衡,因此对采样模块采取通道数降维 1/2 的设计方式。

3.5.3 经典网络分割结果对比

为充分验证所提网络的有效性,将所提网络与经典语义分割网络 SegNet、PSPNet、DeepLab v3 plus 进行对比。性能对比结果如表 5 所示,其中 This Paper 代表所提网络。

表 5 不同网络间性能的对比如

Table 5 Performance comparison between different networks

Network	MIoU / %	MPA / %	Speed / (frame·s ⁻¹)
SegNet	59.86	72.45	10.57
PSPNet	65.28	81.51	16.68
DeepLab v3 plus	66.33	82.04	19.56
This Paper	70.34	83.65	25.94

由表 5 可以看出,在分割性能方面,所提网络均高于三种经典网络,MIoU 分别提升 17.51%、7.75%、6.05%,MPA 分别提升 15.46%、2.63%、1.96%,速度分别提升 145.41%、55.52%、32.62%。虽然与 DeepLab v3 plus 相比,所提网络分割精度提升较小,但分割速度得到了显著提升。

对所提网络与文献[6]中提出的基于特征融合的实时语义分割算法进行对比,结果如表 6 所示,其中 Other paper 代表文献[6]提出的方法,This Paper 代表所提网络。

表 6 所提网络与近年算法的性能对比

Table 6 Performance comparison between the proposed network and a recent algorithm

Network	MIoU / %	Speed / (frame·s ⁻¹)
Other paper	68.43	14.14
This Paper	70.34	25.94

由表 6 可以看出,相对文献[6]中的模型,所提模型在 MIoU 上提升 2.79%,速度上有 83.45% 的提升,较为显著。

3.5.4 不同特征提取网络对比

由于所提模型中的特征提取网络为改进后的 MobileNet v3,尝试将特征提取网络更换为 ResNet50、Xception^[16]以及经典的 MobileNet v3 分别进行实验,检验改进后的 MobileNet v3 性能是否可以达到标准。实验将 batch size 设置为 4,其余参数按照所提模型训练时超参数的设置进行训练,实验结果如表 7 所示。

由表 7 可以看出:与 ResNet50 网络进行对比,改进的 MobileNet v3 的 MIoU 和 MPA 分别降低

表 7 不同特征提取网络的性能对比

Table 7 Performance comparison of different feature extraction networks

Module	MIoU / %	MPA / %	Speed / (frame·s ⁻¹)
ResNet50	72.07	85.06	11.68
Xception	69.33	83.74	17.56
MobileNet v3	67.92	81.39	26.60
Improved MobileNet v3	70.34	83.65	25.94

2.40% 和 1.66%,但速度提升 122.09%;与 Xception 进行对比,改进的 MobileNet v3 的 MIoU 提升 1.46%,MPA 降低 0.11%,速度提升 47.72%;与未进行优化的 MobileNet v3 进行对比,改进的 MobileNet v3 的 MIoU 和 MPA 分别提升 3.56% 和 2.78%,速度降低 2.48%。由此可以看出改进后 MobileNet v3 达到精度补偿效果,并且速度超过 25 frame/s,符合实时分割标准。

3.5.5 分割结果展示

为了直观展示所提模型的分割效果,在相同的参数设置情况下,分别训练三种经典分割模型和所提模型进行对比实验。在 PASCAL VOC2007 数据集上对经过训练后的四个模型进行测试,对比结果如图 10 所示。由图 10 可以看出:对于第一行图片中飞机的分割结果,飞机尾翼在三种经典分割网络中都没有体现,而所提模型对飞机尾翼的处理明显优于另外三种网络;对于第二行图片中人物的分割结果,DeepLab v3+ 在分割中忽略了人物中间的空隙,并且存在错误分类的像素点,SegNet 和 PSPNet 对人物空隙的处理也明显不如所提模型;对于第三行图片中电视的分割结果,三种经典网络均存在多余像素点,所提模型对电视轮廓细节的处理更加优秀;对于第四行图片中鸟类的分割结果,SegNet 对翅膀的处理明显存在缺失,PSPNet 对翅膀的处理存在明显的多余像素点,虽然 DeepLab v3+ 对翅膀的处理相对更好,但是存在多余的像素点,所提模型输出的语义分割结果明显优于三个经典模型;对于第五行图片中人物和马的分割结果,SegNet 的分割结果中存在错误分类,马蹄处的不同颜色像素点被处理为 person 标签,并且对人物腿部处理明显不足,PSPNet 只对人物的腿部细节处理不足,DeepLab v3+ 和所提模型表现更优秀;对于第六行图片中猫的分割,所提模型对耳部轮廓的处理结果更加清晰且更贴近原图。

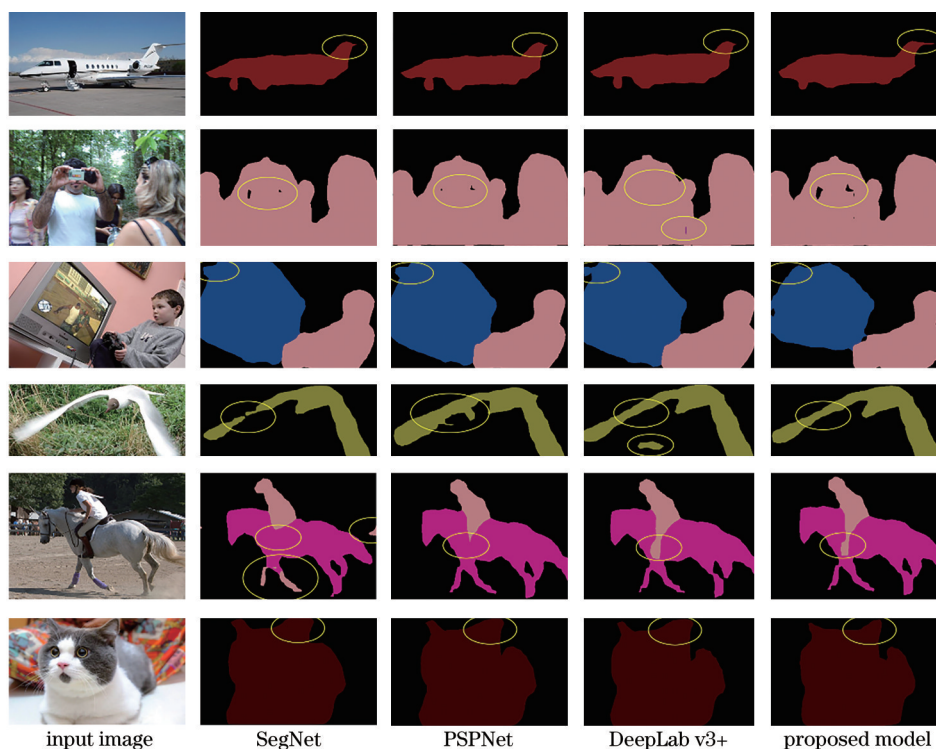


图 10 实验结果对比

Fig. 10 Comparison of experimental results

4 结 论

设计了基于八度卷积的实时语义分割网络。通过降低卷积计算中的空间冗余提升模型的分割速度,并通过 hard-swish 激活函数对 MobileNet v3 进行精度补偿,得到改进后的 MobileNet v3 网络。首先通过改进的 MobileNet v3 进行第一步特征提取;然后利用基于八度卷积和残差思想设计的 OTCH-L 模块获取到高频特征信息,并与最大池化的结果、 1×1 卷积通道调整后的结果并行传播;最后进行通道叠加得到最终分割结果。实验结果表明,相比经典语义分割网络,所提网络的 MIoU、MPA 均有提高,在速度上有显著提升,达到实时语义分割标准(25 frame/s)。在 Pascal VOC2007 数据集测试中,所提网络的速度达到 25.94 frame/s, MIoU 达到 70.34%。

由于这次训练和测试均在 Pascal VOC 数据集上进行,其中大多数物体均是在光线较好情况下进行标注的。但模型在实际应用时会面对夜间或室内弱光源环境。未来研究中可以利用更多弱光源数据集或能见度较低的模糊数据集对模型进行训练和参数调整,使其可以更好适应不同的实际环境。

参 考 文 献

- [1] Jiang F, Gu Q, Hao H Z, et al. Survey on content-based image segmentation methods[J]. Journal of Software, 2017, 28(1): 160-183.
姜枫, 顾庆, 郝慧珍, 等. 基于内容的图像分割方法综述[J]. 软件学报, 2017, 28(1): 160-183.
- [2] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 3431-3440.
- [3] 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.
- [4] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 6230-6239.
- [5] 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.
- [6] Cai Y, Huang X G, Zhang Z A, et al. Real-time semantic segmentation algorithm based on feature fusion technology[J]. *Laser & Optoelectronics Progress*, 2020, 57(2): 021011.
蔡雨, 黄学功, 张志安, 等. 基于特征融合的实时语义分割算法[J]. *激光与光电子学进展*, 2020, 57(2): 021011.
- [7] Chen Y P, Fan H Q, Xu B, et al. Drop an octave: reducing spatial redundancy in convolutional neural networks with octave convolution[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 3434-3443.
- [8] Avenash R, Viswanath P. Semantic segmentation of satellite images using a modified CNN with Hard-Swish activation function[C]//14th International Conference on Computer Vision Theory and Applications, October 30-November 3, 2017, Hammamet, Tunisia. Arab: Arab Computer Society, 2017.
- [9] Mei S P. Public garbage classification system based on MobileNetV3[D]. Wuhan: Wuhan Textile University, 2020.
梅书枰. 基于 MobileNetV3 公共垃圾分类系统[D]. 武汉: 武汉纺织大学, 2020.
- [10] Niu X M, Lü X Q, Gu Y, et al. Breast cancer histopathological image classification based on improved ResNeXt[J]. *Laser & Optoelectronics Progress*, 2020, 57(22): 221021.
牛学猛, 吕晓琪, 谷宇, 等. 基于改进 ResNeXt 的乳腺癌组织病理学图像分类[J]. *激光与光电子学进展*, 2020, 57(22): 221021.
- [11] Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation[M]//Navab N, Hornegger J, Wells W M, et al. *Medical image computing and computer-assisted intervention-MICCAI 2015*. Lecture notes in computer science. Cham: Springer, 2015, 9351: 234-241.
- [12] Szegedy C, Liu W, Jia Y Q, et al. Going deeper with convolutions[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, 2015, Boston, MA. New York: IEEE Press, 2015: 15523970.
- [13] Zhang H, Cisse M, Dauphin Y N, et al. Mixup: beyond empirical risk minimization[EB/OL]. (2017-10-25)[2019-10-05]. <https://arxiv.org/abs/1710.09412>.
- [14] Everingham M, Gool L, Williams C K I, et al. The pascal visual object classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [15] Ma Z H, Gao H J, Lei T. Semantic segmentation algorithm based on enhanced feature fusion decoder [J]. *Computer Engineering*, 2020, 46(5): 254-258, 266.
马振环, 高洪举, 雷涛. 基于增强特征融合编码器的语义分割算法[J]. *计算机工程*, 2020, 46(5): 254-258, 266.
- [16] Costa D, Madeira H, Carreira J, et al. Xception™: a software implemented fault injection tool[M]//Benso A, Prinetto P. *Fault injection techniques and tools for embedded systems reliability evaluation*. Frontiers in electronic testing. Boston: Springer, 2003, 23: 125-139.