

文章编号: 2095-2163(2022)10-0079-07

中图分类号: TP301.6

文献标志码: A

基于多尺度特征融合的道路场景语义分割

高建瓴, 陈楠, 喻明毫

(贵州大学 大数据与信息工程学院, 贵阳 550025)

摘要: 针对传统的语义分割方法实时性较低,同时感受野也较少,计算量比较大并且需要很长的处理时间等问题,将基于快速卷积神经网络模型(Fast-SCNN)改进后的CFast-SCNN应用于道路场景语义分割,首先在原网络编码部分进行模型剪枝,压缩模型的体积,提高该网络的推理速度;然后在其主干网络部分进行多尺度特征融合操作,扩大感受野,提高大目标边界准确率,保证分割的精度。使用道路场景分割中常用的Cityscapes数据集进行试验,并且使用语义分割领域中常用的性能评价指标验证了方法的有效性,实验结果表明改进后的网络相较于原网络各性能都有一定提升。

关键词: 语义分割; 实时性; 模型压缩; 多尺度

Semantic segmentation of road scenes based on multi-scale improvement

GAO Jianling, CHEN Nan, YU Minghao

(College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China)

[Abstract] Semantic segmentation in traffic scenes is not only the key technology in unmanned driving, but also one of the key technologies of computer vision. The traditional semantic segmentation methods have low real-time performance, less receptive fields, large amount of calculation and need a long processing time. To solve this problem, the improved CFast-SCNN based on fast convolution neural network model (Fast-SCNN) is applied to road scenes semantic segmentation. Firstly, the model is pruned in the coding part of the original network to compress the volume of the model and improve the reasoning speed of the network; Then the multi-scale feature fusion operation is carried out in the backbone network to expand the receptive field, improving the accuracy of large target boundary and ensuring the accuracy of segmentation. The Cityscapes dataset commonly used in road scenes segmentation is selected to test, and the performance evaluation indexes commonly used in the field of semantic segmentation verify the effectiveness of the method. The experimental results show that the performance of the method is improved compared with the original network.

[Key words] semantic segmentation; real-time performance; model compression; multi-scale

0 引言

当前,伴随着人工智能研究热潮的全球兴起,深度学习方法也随即得到了高速发展,在计算机视觉^[1]领域的应用迄至目前已然臻至成熟,同时也还是现如今进行的人工智能研究中至关重要的理论组成部分。深度学习中的图像语义分割主要是为了对输入图像信息进行图像语义划分、划分为像素级别的分类,以更好地实现对目标对象的精确识别定位,从而完整地呈现其轮廓特征。有关研究已经逐渐成为计算机视觉领域的一个热点,且另有研究指出,语义分割任务也是计算机视觉领域的一个难点。

图像语义分割是一类利用图像中的高层语义特征对单个图像像素进行语义预测识别和语义分类的

重要技术。随着当代人工智能技术的不断深入发展,在各种计算机视觉前沿理论研究体系中,图像语义分割已经日渐成为居于核心的基础部分,在诸多领域中皆有着广泛的应用价值,例如在医疗影像领域^[2],图像语义分割能够分割出细节信息,使医生能够从图像中更清晰地了解到病人的身体信息,为医生诊断病情提供帮助。在人们拍照时,语义分割方法能够精准地识别脸部轮廓信息,辅助人脸美化算法,释放人类爱美的天性。在自动驾驶中^[3],图像语义分割^[4]能够帮助汽车识别场景中的道路、障碍物等信息并进行精确定位,使汽车能够安全地行驶。随着技术革新和生活中需求的大幅提升,就越发体现出对语义分割技术进行创新研究的重要性,且如前文所述,语义分割任务也是计算机视觉领域的一个难点。主要表现在:

基金项目: 国家自然科学基金(62062021,61872034);贵州省科学技术基金资助项目(黔科合基础[2020]1Y254);贵州省自然科学基金资助项目(黔科合基础[2019]1064)。

作者简介: 高建瓴(1969-),女,副教授,硕士生导师,主要研究方向:数据库系统、数据挖掘;陈楠(1997-),女,硕士研究生,主要研究方向:智能算法、图像处理;喻明毫(1997-),男,硕士研究生,主要研究方向:智能算法、图像处理。

收稿日期: 2022-03-01

(1)在物体层次上,由于光照、距离等拍摄条件的不同,相同的物体在图像中的表现可能差距很大,而且各物体之间经常存在遮挡、割裂等现象,因此正确地标记出各物体的语义信息并不容易。

(2)在类别层次上,由于图像内容复杂多变,相同类别的物体在图像中可能千差万别,而不同类别的物体又可能高度相似。

(3)在背景层次上,现实场景中的图像往往存在繁杂凌乱的背景,这也明显增大了图像语义分割任务的难度。

在交通道路场景中,判断汽车行车路况从而躲避障碍进行安全驾驶,对场景的环境进行感知是一个棘手、且重要的问题。道路中的场景复杂多变,传统方法无法实现对道路场景中各个物体的检测。如今众多的科研人员^[5-7]将卷积神经网络应用到无人驾驶的环境感知领域,而当下的无人驾驶技术中大多利用的是价格昂贵的激光雷达来进行环境感知,另有研究指出视觉感知成本低廉,也有助于产品化,同时卷积神经网络还是近来的研究热点,因此基于卷积神经网络对道路场景进行语义分割获取环境信息技术进行研究很有必要。

在当前基于计算机视觉的图像理解分析这一类别的应用中,图像语义分割技术是最重要的组成部分之一,不仅已被广泛使用在工业界中,即便是在当下的学术界,也是研究攻关的热门方向之一。作为图像信息处理当中最重要、最基础、而且也是最复杂的技术问题之一,图像语义分割相关技术的完善将能够带动其他领域的发展,在当今社会具有广阔的市场前景和实用价值,因而也有着重要的研究意义。

当下已问世的大多数语义分割模型都是以编码器-解码器作为网络的核心,编码器是将信息进行编码,输入到特征空间中,以此获取图像的特性信息;解码器则将输入的信息编码映射到空间的分类器中以执行分割。虽然语义分割常常被应用于实时分割的任务中,但是,目前的大部分卷积神经网络层次较深、实时性不够高,计算量比较大并且需要很长的处理时间,本文则基于快速卷积神经网络模型(Fast-SCNN)^[2],在其主干网络模块中,结合模型剪枝、多尺度卷积优化等操作,提出新的网络模型CRFast-SCNN。研究中,先在编码部分进行模型剪枝,压缩模型的体积,提高该网络的推理速度,接着就在其主干网络部分进行多尺度卷积优化操作,扩大感受野,提高大目标边界准确率,保证分割的精度。最后通过对模型进行训练。验证了改进后的

CRFast-SCNN 相较于传统语义分割网络的准确率更高。

1 语义分割相关模型

2014年,提出和定义了一个全卷积网络(Fully Convolutional Networks, FCN)^[3]。目前看来,该类网络技术主要是应用于图像语义信息的分割,改编自预先训练好的分类网络VGG-Net结构(Very Deep Convolutional Networks, VGG-Net)。该网络中的所有全连接层全部被替换成可直接输入任意大小尺寸图像的卷积层。不仅如此,还进一步提出了像素精度、平均像素精度和MIou等语义分割模型的评价标准。FCN在网络计算中也首次真正地实现了深度卷积神经网络从端到端的语义分割。在该项研究之后,利用深度卷积神经网络进行图像语义分割进入了一个快速发展时期,于是学界经常使用全卷积神经网络来解决图像的语义分割问题。

Ronneberger等人^[4]在2015年提出的U-Net通过简单地拼接解码器获得的特征映射和解码器在每个对应阶段的上采样特征映射,构建了一个梯形网络模型结构。通过跳跃连接(Skip Connection, SC)架构,该方法允许网络解码器在网络编码结构的任何阶段学习编码器池化操作中丢失的图像特征。这种网络结构在低样本数据集的医学图像分割方面取得了可观成果,也进行了大量的实验研究。

2016年,剑桥大学团队基于FCN设计了一种卷积神经网络结构SegNet^[5],具有堆成编码解码结构。一部分编码结构是采用了基于VGG-16网络的前13层卷积网络,都对应着一个特定类型的解码器,下采样方式使用最大池化,在上采样阶段,使用下采样中记录的边界信息精确映射到原始位置,通过逐层恢复细节来提高分割精度,用于道路场景的分割。尽管多层最大池化与上采样可以获得更多的平移不变特征,但特征的空间分辨率仍将丢失,精度仍有提高的空间。

2016年,PSPNet^[6]基于FCN的不足已然显现。虽然FCN是当时语义分割效果最好的模型,但是当FCN进行语义分割预测时,不考虑全局信息,也就导致不能快速正确地判断像素的上下文信息。为了更好地解决上述所有问题,PSPNet模型中选用了金字塔池化模块,该模块可以更有效地聚合不同区域间的上下文信息,进而获取全局信息的处理能力也得到了极大的提升,文章通过实验验证了该结构的有效性。

深度残差网络 (Dilated Residual Networks, DRN)^[7] 是 2017 年 CVPR 收录的论文 Dilated Residual Networks 中提出的模型, 该文主要提出了在语义分割中使用膨胀卷积, 并且还解决了使用膨胀卷积时产生的问题。

2019 年, DANet^[8] 模型则被提出, 因为卷积运算产生布局感受野, 所以同一标签的像素的相应特征可能不同, 这种差异将导致类别内的一致, 会对识别的准确率产生影响。为了使像素级别识别特征的能力得到提高, DANet 提出利用注意力机制建立特征之间的关联, 用来获取全局上下文信息。

然而, 最先进的实时语义分割仍然具有挑战性, 通常需要高端的 GPU。受到双分支方法的启发, Fast-SCNN 采用了共享的浅层网络路径来进行编码, 同时以低分辨率高效地学习上下文信息。

2 本文网络模型与结构

Fast-SCNN^[2] 是 2019 年提出的一种快速卷积

神经网络, 该模型在高分辨率图像上也能达到实时效果, 内存消耗低、很适合在嵌入式设备上运行, 模型比较小, 相对于其他语义分割网络具有参数少、运算速度快、效率高等优点, 并且也具有一定的准确性, 可以满足在道路场景上进行语义分割的实时性和准确率的要求。

但是 Fast-SCNN 网络也仍然具有升级的空间, 后期可以继续改进提高。本文对 Fast-SCNN 采用了模型剪枝和多尺度卷积优化的改进方法, 提出了一种多尺度特征道路场景语义分割的模型 CRFast-SCNN。

2.1 Fast-SCNN 网络介绍

Fast-SCNN 受双分支结构和经典的编码器解码器框架启发, 模型前面几个层是用来提取低层次的特征, 将第一部分改为一个浅层学习下采样的模块, 使上述 2 个框架的核心思路结合起来, 从而构建一个快速的语义分割模型。Fast-SCNN 的结构如图 1 所示。

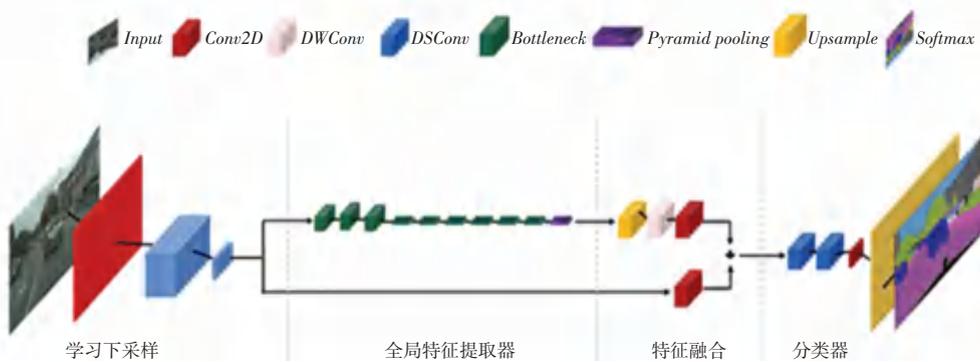


图 1 Fast-SCNN 网络结构

Fig. 1 Fast-SCNN network structure

由图 1 可知, 整个网络由学习下采样模块、全局特征提取器、特征融合模块和分类器四部分组成。这里给出研究阐释如下。

(1) 学习下采样模块^[9]: 该模块的卷积层总共有 3 个, 先是 1 个普通的卷积层, 后面 2 个则使用了深度可分离卷积层, 这 2 层主要是用来提高效率。一般情况下, 每个卷积层的步长设置为 2, 该模块输出特征的长度 (或宽度) 也大约为输入图像尺寸的 1/8。每一个卷积层卷积核大小均为 3×3, 每一个卷积层后都会紧接着一个 BN 层和一个 ReLU 激活函数。

(2) 全局特征提取器: Fast-SCNN 利用全局特征提取器来实时获取全局上下文信息, 全局特征提取器的结构有些类似于传统的两分支结构中的深度

分支。在传统的深度分支中, 其输入图像的分辨率一般很低, 而全局特征提取器的输入是在经过学习下采样模块处理后所得到的特征图。当输入图像和输出图像大小一致时, 使用残差连接以及深度可分离卷积, 可以减少参数数量以及浮点操作的数量。此外, 在模块最后还添加了一个金字塔池模块 (PPM), 高模块可以根据区域的不同来进行上下文信息的聚合。

(3) 特征融合模块: 特征融合模块的主要作用之一是融合 2 个重要分支的输出特征, Fast-SCNN 进行特征融合的结构一般比较清晰简单, 可以在一定时间内最大限度地提升计算效率。

继续融合时, 2 个深度分支上输出特征的大小必须一致, 因此, 就要同时对深度分支的输出特征进

行上采样操作。在2个分支的末端,都加上了一个尺寸为 1×1 的卷积操作,主要功能是进行最终通道数的调整,将2个分支最终的输出特征予以相加,再使用激活函数来进行非线性函数的变换操作。

(4)分类器:该模块主要由2个深度可分离卷积,加上一个尺寸为 1×1 的常规卷积组成,这样就使网络性能得到提高。模块中还有一个 $Softmax$ 操作^[10]。在进行网络结构训练时,使用 $Softmax$ 来计算损耗,在实际推理时,将 $Softmax$ 操作替换为 $Argmax$ 操作,就可以大幅提高推理的速度。

研究归纳后得到的前述4部分的网络结构参数见表1。表1中, t 表示瓶颈残差块的膨胀系数; c 表示该计算模块输出的特征图通道数; n 表示该模块重复的次数; s 表示卷积的步长。

表1 Fast-SCNN 网络结构参数

Tab. 1 Fast-SCNN network structure parameters

Input	Block	t	c	n	s
$1\ 024 \times 2\ 048 \times 3$	Conv2D	-	32	1	2
$512 \times 1\ 024 \times 32$	DSCConv	-	48	1	2
$256 \times 512 \times 48$	DSCConv	-	64	1	2
$128 \times 256 \times 64$	Bottleneck	6	64	3	2
$64 \times 128 \times 64$	Bottleneck	6	96	3	2
$32 \times 64 \times 96$	Bottleneck	6	128	3	1
$32 \times 64 \times 128$	PPM	-	128	-	-
$32 \times 64 \times 128$	FFM	-	128	-	-
$128 \times 256 \times 128$	DSCConv	-	128	2	1
$128 \times 256 \times 128$	Conv2D	-	19	1	1

2.2 Fast-SCNN 网络改进

道路场景语义分割对实时性和准确率都有一定的要求。因此,在选择深度学习的网络框架时,就要同时兼顾到实时性和准确率。相较于传统的语义分割网络模型,Fast-SCNN 在实时性和准确率上有着较好的表现,但仍具有一定可塑性,可进一步加以改进与提高。因此通过对原网络的详尽分析可知,为了更适用于道路场景语义分割,本文拟对 Fast-SCNN 网络从下面2个方面来进行改进:

(1)网络模型剪枝:通过多次实验,找到模型中存在的某个模块对整体效果影响不大,而且去掉之后还可以提高模型计算速度,对该部分进行微调,从而提高了推理速度。

(2)多尺度特征卷积:在残差网络中使用残差分层的方式进行连接,可以在更细粒度级别表达多尺度特征,可以增加每层网络的感受野大小,保证分割精度。

2.2.1 模型剪枝

由表1可知,在Fast-SCNN 模型中的全局特征提取器中相同膨胀系数的Bottleneck 层的重复次数都为3,经过多次对比实验,本文中的Bottleneck 改进对比实验结果见表2。由表2可知,减少Bottleneck 层的重复次数后,模型效果并未下降,计算速度还有了一定的提高,由此可以看出原模型中的全局特征提取部分存在一定的信息冗余,过多的Bottleneck 层会增加模型的参数量和计算量,虽然产生的影响较小,但是减少后却可以进一步降低参数量和计算量,以及提高网络的计算速度。这里即将模型剪枝后的网络模型命名为CFast-SCNN。

表2 Bottleneck 改进对比实验

Tab. 2 Bottleneck improved comparison experiments

Input	Block	t	c	n	s
$1\ 024 \times 2\ 048 \times 3$	Conv2D	-	32	1	2
$512 \times 1\ 024 \times 32$	DSCConv	-	48	1	2
$256 \times 512 \times 48$	DSCConv	-	64	1	2
$128 \times 256 \times 64$	Bottleneck	6	64	3	2
$64 \times 128 \times 64$	Bottleneck	6	96	3	2
$32 \times 64 \times 96$	Bottleneck	6	128	3	1
$32 \times 64 \times 128$	PPM	-	128	-	-
$32 \times 64 \times 128$	FFM	-	128	-	-
$128 \times 256 \times 128$	DSCConv	-	128	2	1
$128 \times 256 \times 128$	Conv2D	-	19	1	1

2.2.2 多尺度特征融合

Fast-SCNN 网络借用了PSPNet 网络中的金字塔池化部分,并使用多级池化技术实现了全局信息的获取,以及当前热门的多分支结构^[11-12],改进后的Fast-SCNN 网络结构如图2所示。由图2可知,具体来说,在对卷积进行下采样步骤后,使用多级池化方法从浅至深的方向提取出图像的多层次特征;同时,从第二分支中了解高图像分辨率级别的浅层信息特征,并有效填充因多级池化方法而引起的空间信息流失;网络中的卷积部分通常使用的是深度可分离卷积,这类卷积算法有助于缩小模型规模和提高推理速度。虽然Fast-SCNN 可以将通过多次卷积下采样操作所得的输出结果与金字塔池化进行融合,但最终得到的图像语义特征却容易丢掉了图像中某些小型物件的浅层特性。这将导致难以检测到过小尺寸的边缘特征信息。网络中经过池化层后再进行了4倍上采样,使得其 $1/8$ 图像的特征信息差值变大,使得网络对边缘像素进行标签预测时的效果也会变差。至此,将经过上采样后的特征信息

和从第二分支得到的通道信息进行了融合,而对于上采样后的特征信息,则不能进行通道上的特征识

别,这就使得整个网络的信息分割精度在一定程度上有所降低。

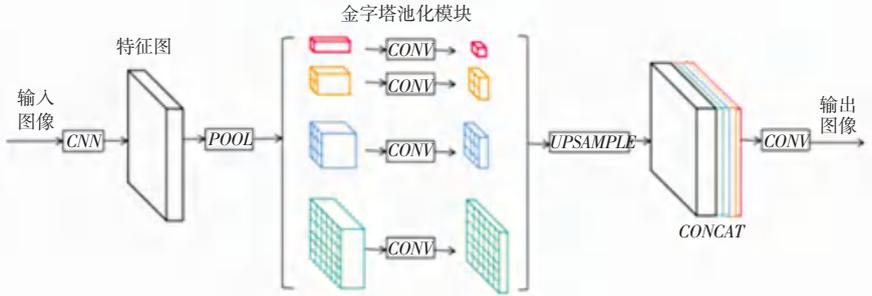


图 2 改进后的 Fast-SCNN 网络结构

Fig. 2 Improved Fast-SCNN network structure

改进后的结果可分述如下:

(1) 模型剪枝,减少全局特征提取器部分产生的冗余信息,降低参数量,减少计算量。

(2) 多次重复进行下采样,使网络更大概率地学习到图像中更深层的语义特征信息,针对细节信息,加强网络的识别精度。

(3) 为避免扩展计算产生的误差,在完成金字塔池化部分后,调整上采样的采样倍数。

(4) 基于 U-Net 网络中编码器和解码器对称的思路,在高层语义特征进行上采样后融合 3 组不同尺寸的特征信息,尺寸分别为 1/4、1/8、1/16,可以识别更多浅层细节信息。

(5) 获取高层语义特征后,将该特征信息进行上采样,再将其与浅层信息融合到一起,能更好地学习到通道间信息,使网络模型拥有区域不同通道信息的能力。

3 实验结果与分析

3.1 数据集

Cityscapes 数据集^[13]是一个大型公开的城市道路场景数据集,图像来源分别为在 50 个不同城市的街道上拍摄采集的城市街景图。Cityscapes 数据集图片示例见图 3。该数据集中包含了 5 000 张像素级标注的高质量图片,其中训练集有 2 795 张,测试集与验证集有 500 张,用于基准测试的有 1 525 张。此外,为了进一步地验证弱监督语义分割模型,还包含 20 000 张带有粗略标注的图像。数据集中包括 8 个类别。这 8 个类别涵盖了 19 个子类别。常用的 19 个语义类别是:道路 (Road)、人行道 (Sidewalk)、建筑物 (Building)、墙壁 (Wall)、栅栏 (Fence)、栏杆 (Pole)、交通灯 (Traffic light)、交通标志 (Traffic

sign)、草丛 (Vegetation)、地面 (Terrain)、天空 (Sky)、行人 (Person)、骑行者 (Rider)、汽车 (Car)、卡车 (Truck)、公交车 (Bus)、火车 (Train)、摩托车 (Motorcycle)、自行车 (Bicycle)。



图 3 Cityscapes 数据集图片示例

Fig. 3 An example of Cityscapes dataset pictures

3.2 评价标准

本文选用平均交并比 (Mean Intersection over Union, $MIoU$) 作为模型效果的评价指标。计算每一类的 IoU 然后求取平均值。在众多的语义分割评价指标中, $MIoU$ 因其具有简洁、代表性较强的特点,故而成为研究中最常用的度量标准。

假设图像中共有 $k+1$ 个类别, P_{ij} 表示本属 i 类却预测为 j 类的像素点总数, P_{ii} 表示第 i 类像素预测正确的个数, $MIoU$ 评价指标的计算可由如下公式进行计算:

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (1)$$

3.3 参数设置

文章仿真实验是基于 Ubuntu 18.014 操作系统,处理器为 Intel Xeon Silver 4210, RAM 为 64 G 内存。编译环境采用 Python3.6 编译环境及 Pytorch1.6 框架。本文使用 Fast-SCNN 和改进后的 CFast-SCNN 在 Cityscapes 数据集上进行训练和验证。使用标准

梯度下降法训练模型,参数设置:批量大小为 50,基础学习率为 0.12,学习动量为 0.9,权重衰减系数设置为 $4e-5$,这样一来每次训练都可以得到最佳性能。

3.4 实验结果

改进后的模型和原模型在 Cityscapes 数据集上的比较结果见表 3 所示。

表 3 各模型在 Cityscapes 数据集上训练的结果

Tab. 3 Results of each model's training on the Cityscapes dataset

Classes	Fast-SCNN	FCN	PSPNet	CFast-SCNN
Road	96.85	97.73	96.55	96.59
Sidewalk	77.48	68.87	79.24	81.14
Buiding	89.07	87.58	89.56	91.42
Wall	43.78	30.67	45.07	44.20
Fence	49.66	30.52	39.78	57.35
Pole	49.85	26.89	50.90	65.30
Traffic Light	48.20	31.47	62.09	68.81
Traffic Sign	63.36	47.01	70.73	65.70
Vegetation	89.84	90.48	90.30	89.59
Terrain	50.43	50.38	52.70	57.93
Sky	91.98	94.18	91.18	94.21
Person	71.36	69.88	75.58	76.27
Rider	45.62	38.52	51.48	40.63
Car	90.54	89.01	91.25	90.90
Truck	43.16	71.03	38.02	64.64
Bus	38.72	72.74	65.85	63.83
Train	24.95	68.10	33.18	19.75
Motocycle	30.24	39.77	48.06	51.67
Bicycle	63.51	58.57	72.55	74.34
<i>MIoU</i>	60.85	62.10	65.48	68.11

由表 3 中可以看出在 Cityscapes 数据集上,改进后的模型效果优于改进前的模型效果。改进后的 CFast-SCNN 相较于原模型在平均交并比 (*MIoU*) 上有 7.26% 的提升。改进后模型在准确率上也有所提高,达到了预期的实验效果。

在设置超参数和网络结构都为最优的情况下,随着迭代次数的不断增加,网络损失函数曲线图和 *PA* 变化曲线图分别如图 4、图 5 所示。

将图 4 与图 5 进行对比可以看出,改进后的模型 CFast-SCNN 经过 75 次迭代后,损失函数值下降很明显, *MIoU* 精度值在经过 50 次迭代后则有显著提升。损失值越低,在训练过程中产生的损失越少,网络准确率较高。 *MIoU* 值提升也证明了改进后 CFast-SCNN 网络的有效性。

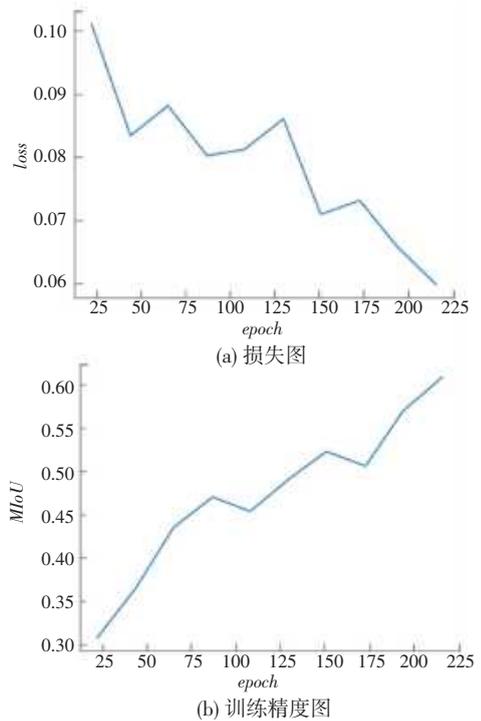


图 4 Fast-SCNN 的损失图和训练精度图

Fig. 4 Loss diagram and training accuracy diagram of Fast-SCNN

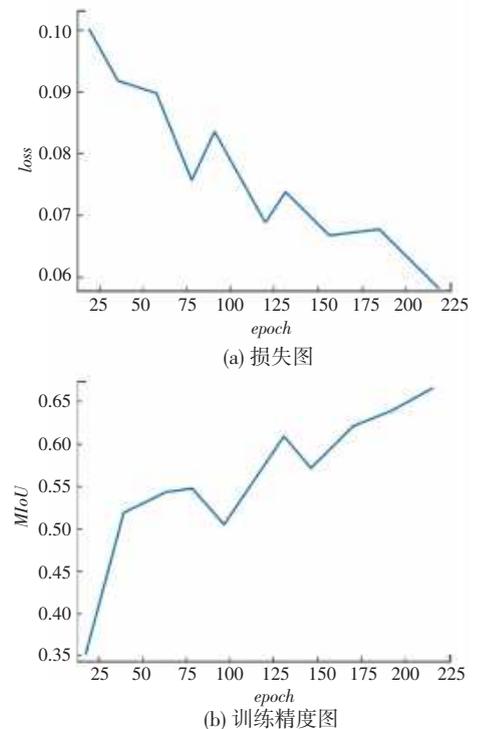


图 5 CFast-SCNN 的损失图和训练精度图

Fig. 5 Loss diagram and training accuracy diagram of CFast-SCNN

4 结束语

作为计算机视觉中图像理解的重要一环,图像语义分割技术不仅在工业界的需求正日益凸显,而且也是当下学术界的研究热点之一。本文提出的改

进后 CFast-SCNN 道路场景语义分割模型, 可以满足当下在实际环境中道路场景的分割准确率和实时性的需求。研究可知, 改进后的模型相比原模型在各方面性能上都有了一定的提高。未来也还会在网络的改进上开展更多研究。

参考文献

- [1] CHENG Guo, ZHENG Jiangyu, KILICARSLAN M. Semantic segmentation of road profiles for efficient sensing in autonomous driving[C]// 2019 IEEE Intelligent Vehicles Symposium (IV). Paris, France; IEEE, 2019: 564-569.
- [2] POUDAL R P K, LIWICKI S. Fast - SCNN: Fast Semantic Segmentation Network[C]. arXiv preprint arXiv:1902.04502v1, 2019.
- [3] LONG J, SHELHAMER E, DARRELL T. Fully Convolution Networks for Semantic Segmentation[C]// Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). Boston: IEEE, 2015: 3431-3440.
- [4] RONNEBERGER O, FISCHER P, BROX T. U - Net: Convolutional networks for biomedical image segmentation[M]// NAVAB N, HORNEGGER J, WELLS W, et al. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science. Cham: Springer, 2015, 9351: 234-241.
- [5] BADRINARAYANAN V, KENDALI A, CIPOLLA R. SegNet: A deep convolutional encoder - decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis &

- Machine Intelligence, 2017, 39(12): 2481-2495.
- [6] ZHAO Hengshuang, SHI Jianping, QI Xiaojun, et al. Pyramid scene parsing network[C]// I2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 6230-6239.
- [7] YU F, KOLTUN V, FUNKHOUSER T. Dilater residual network [C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 636-644.
- [8] FU Jun, LIU Jing, TIAN Haijie, et al. Dual attention network for scene segmentation [C]// IEEE Conference on Computer Vision and Pattern Recognition. Long Beach, CA, USA: IEEE, 2019: 472-480.
- [9] MAZZINI D. Guided upsampling network for real-time semantic segmentation[J]. arXiv preprint arXiv:1807.07466v1, 2018.
- [10] YU Changqian, WANG Jingbo, PENG Hao, et al. Bisenet: Bilateral segmentation network for real-time semantic segmentation [M]// FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision - ECCV 2018. ECCV 2018. Lecture Notes in Computer Science. Cham: Springer, 2018, 11217: 334-349.
- [11] 黄福蓉. 用于实时道路场景的语义分割算法 CBR-ENet[J]. 中国电子科学研究院学报, 2021, 16(03): 277-284.
- [12] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.
- [13] 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). Las Vegas: IEEE, 2016: 3213-3223.

(上接第 78 页)

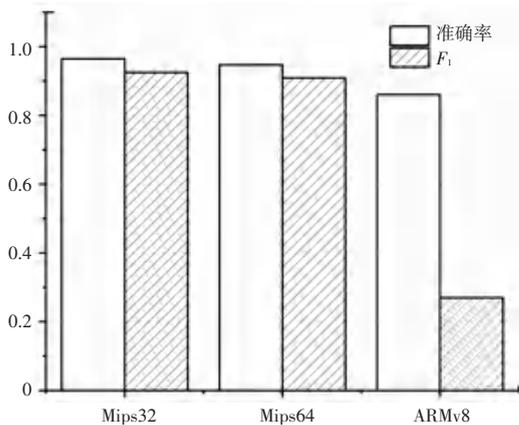


图 5 实验结果柱状图

Fig. 5 Histogram of experimental results

4 结束语

本文根据 LLVM 编译器架构中对指令描述的代码对指令进行分类, 从 Mips32 架构指令集手册中提取出指令的描述文本, 并根据上述分类原则对文本数据进行手工标注, 最后实现了对 BERT 模型的预训练和下游多标签分类任务的训练。经过对 Mips64 与 ARMv8 架构指令集的分类测试结果表明了本文提出基于 BERT 模型的指令集多标签分类研

究方法在指令多标签分类任务中的有效性。

参考文献

- [1] 董峰, 付宇卓. 基于 LLVM 架构的 ARM 后端移植[J]. 信息技术, 2007(07): 38-41.
- [2] 严佩敏, 唐婉琪. 基于改进 BERT 的中文文本分类[J]. 工业控制计算机, 2020, 33(07): 108-110, 112.
- [3] 杨飘, 董文永. 基于 BERT 嵌入的中文命名实体识别方法[J]. 计算机工程, 2020, 46(04): 40-45, 52.
- [4] DEVLIN J, CHANG M W, LEE K, et al. BERT: Pretraining of deep bidirectional transformers for language understanding [J]. Computation and Language, 2018, 23(2): 3-19.
- [5] 段瑞雪, 巢文字, 张仰森. 预训练语言模型 BERT 在下游任务中的应用[J]. 北京信息科技大学学报(自然科学版), 2020, 35(06): 77-83.
- [6] 王鹏, 陈影, 邢明杰. 基于 LLVM 的 RISC-V 自定义扩展指令支持方法[J]. 计算机系统应用, 2021, 30(11): 20-26.
- [7] 孔凡芑, 刘旭红, 刘秀磊, 等. 基于 BERT 模型的航天科技开源情报分类[J]. 北京信息科技大学学报(自然科学版), 2021, 36(01): 28-33.
- [8] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[EB/OL]. [2020-10-11]. <https://gluebenchmark.com/leaderboard>.
- [9] HOWARD J, RUDER S. Universal language model fine-tuning for text classification[J]. arXiv preprint arXiv:1801.06146, 2018.
- [10] 王国泰, 董晶晶, 高杨, 等. 基于 Bert 预训练模型的虚假新闻文本检测[J]. 信息技术, 2022(01): 137-142.