计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2023,32(10):235−241 [doi: 10.15888/j.cnki.csa.009276]

E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于类间距离蒸馏的语义分割①

邓文革[1], 王亚军[1], 隋立林[1], 孙国栋[1], 张正博[2]

[1](国能数智科技开发 (北京) 有限公司, 北京 100011)
[2](武汉大学 测绘遥感信息工程国家重点实验室, 武汉 430079)
通信作者: 隋立林, E-mail: 20070885@ceic.com

摘　要: 知识蒸馏被广泛应用于语义分割以减少计算量. 以往的语义分割知识提取方法侧重于像素级的特征对齐和类内特征变化提取, 忽略了对语义分割非常重要的类间距离知识的传递. 为了解决这个问题, 本文提出了一种类间距离提取方法, 将特征空间中的类间距离从教师网络转移到学生网络. 此外, 语义分割是一个位置相关的任务, 因此本文开发了一个位置信息提取模块来帮助学生网络编码更多的位置信息. 在 Cityscapes、Pascal VOC 和 ADE20K 这 3 个流行的语义分割数据集上的大量实验表明, 该方法有助于提高语义分割模型的精度, 取得了较好的性能.

关键词: 知识蒸馏; 语义分割; 模型压缩

引用格式: 邓文革,王亚军,隋立林,孙国栋,张正博.基于类间距离蒸馏的语义分割.计算机系统应用,2023,32(10):235–241. http://www.c-s-a.org.cn/1003-3254/9276.html

Distilling Inter-class Distance for Semantic Segmentation

DENG Wen-Ge[1], WANG Ya-Jun[1], SUI Li-Lin[1], SUN Guo-Dong[1], ZHANG Zheng-Bo[2]

[1](CHN Energy Digital Intelligence Technology Development (Beijing) Co. Ltd., Beijing 100011, China)
[2](State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China)

Abstract: Knowledge distillation is widely adopted in semantic segmentation to reduce the computation cost. The previous knowledge distillation methods for semantic segmentation focus on pixel-wise feature alignment and intra-class feature variation distillation, neglecting to transfer the knowledge of the inter-class distance, which is important for semantic segmentation. To address this issue, this study proposes an inter-class distance distillation (IDD) method to transfer the inter-class distance in the feature space from the teacher network to the student network. Furthermore, since semantic segmentation is a position-dependent task, thus this study exploits a position information distillation module to help the student network encode more position information. Extensive experiments on three popular semantic segmentation datasets: Cityscapes, Pascal VOC, and ADE20K show that the proposed method is helpful to improve the accuracy of semantic segmentation models and achieves great performance.

Key words: knowledge distillation; semantic segmentation; model compression

语义分割旨在为输入图像的每个像素分配一个标签. 它是计算机视觉中一项基本而富有挑战性的任务, 在许多领域都有广泛的应用, 如自动驾驶[1]、地物变化检测[2] 等. 近年来, 由于深度学习在计算机视觉中的成功, 基于卷积神经网络 (CNN)[3] 的方法大大提高了语义分割的准确性. 然而, 基于神经网络的语义分割算法通常计算成本较高, 这限制了其在实际中的应用, 尤其是对于要求高效率的实际任务.

为了解决这个问题, 国内外研究人员开发了许多轻量级模型, 例如 ENet[4]、ESPNet[5]、ICNet[6] 和 STDC[7].

① 收稿时间: 2023-03-27; 修改时间: 2023-04-27; 采用时间: 2023-05-23; csa 在线出版时间: 2023-08-21
CNKI 网络首发时间: 2023-08-22

尽管研究人员设计了优秀的网络来降低计算成本, 但在准确性和模型大小之间的平衡性很难达成令人满意的结果. 为解决这个问题, 本文没有重新设计主干网络, 而是采用知识蒸馏策略, 在教师网络的指导下训练学生网络, 并获得了与上文提到的网络差不多的结果.

知识蒸馏[8] 作为一种模型压缩方法, 最初用于图像分类任务, 能够显著简化繁琐的模型. 由于知识蒸馏方法的优势, 一些语义分割方法使用知识蒸馏来减少模型大小, 它们使学生模型向教师网络逐像素学习特征和类内特征变化. 其中的代表性算法, 类内特征方差提取 (IFVD)[9] 侧重于将类内特征的变化从教师网络转移到学生网络. 通道式知识蒸馏 (CD)[10] 强调提取每个通道中最重要的区域. 值得注意的是, 语义分割是具有各种类别的像素级类别预测任务, 因此特征空间中的类间距离在语义分割中是普遍存在的. 由于有众多的参数和复杂的网络结构, 教师网络在特征空间中具有更强的分类能力和更大的类间距离. 然而, 过去用于语义分割的知识蒸馏方法忽略了将教师网络的特征空间中的类间距离转移到学生网络中.

另一方面, CNN 能够隐式地对位置信息进行编码[11]. 由于语义分割是一个位置相关的任务, 一般来说, 由于网络结构简单, 参数较少, 学生网络无法像教师网络那样编码丰富的位置信息.

为了解决上述问题, 本文采取提取特征空间中的类间距离和从教师网络到学生网络的位置信息. 本文提出了一种新方法, 称为类间距离蒸馏 (简称为 IDD 方法), 它由两个主要部分组成.

(1) 类间距离蒸馏模块 (IDDM): 本文首先设计了一个图对类间距离进行编码, 让学生网络模仿老师网络的大的类间距离.

(2) 位置信息提取模块 (PIDM): 本文采用位置信息网络来提取隐含在特征图中的位置信息. 教师网络和学生网络都将通过该网络预测绝对坐标掩码. 通过最小化其输出结果, 学生网络可以编码更多的位置信息. 通过 IDD 方法, 使得学生网络学习更多关于类间距离和位置信息的知识, 以提升学生网络的分割精度.

1 相关工作

1.1 语义分割

卷积神经网络的模型的出现极大地促进了语义分割的发展. 许多研究者尝试不同的方法来使深度学习模型习得丰富的上下文信息[12–14]. Zhao 等人[15] 提出了一种从多个尺度收集上下文信息的金字塔池策略; DeepLabV2[16] 采用 atrous 空间金字塔池方法获取丰富的上下文信息; 同时, 学者们采用编码器-解码器模块捕获多级特征和上下文信息; OCNet[17] 利用了一种自我注意机制来捕捉所有像素之间的关系.

为了满足移动平台的实时语义分割需求, 研究人员们也提出了一些轻量级网络. ENet[4] 采用了非对称的编解码结构和卷积核分解运算, 大大减少了参数和浮点运算的数量; ESPNet[5] 中采用了点态卷积和空间金字塔的膨胀卷积以降低计算量; ICNet[6] 通过设计一种高效的网络结构来处理不同分辨率的图像, 实现了快速语义分割; STDC[7] 通过减少网络冗余设计了一种新的实时分段架构. 与此不同, 本文使用知识蒸馏得到了轻量级的语义分割网络, 避免了重新设计网络结构, 提高了效率.

1.2 针对语义分割的知识蒸馏

Hinton 等人[8] 提出了知识蒸馏的概念, 将软标签从教师网络转移到学生网络以提高学生网络性能. 由于知识蒸馏的卓越性能, 一些研究者将知识蒸馏应用于语义分割. Liu 等人[18] 使用结构化知识发现方法从教师网络中传输像素式、成对式和整体式知识; He 等人[19] 设计了一个自动编码器, 将知识转换为学生网络更容易学习的紧凑形式; Wang 等人[9] 提出了一种类内特征变异提取方案, 使学生网络模拟教师网络的类内特征分布; Shu 等人[10] 开发了一种简单而有效的方法来最小化教师网络和学生网络之间的渠道差异. 与上述方法不同, 本文的方法注重提取特征空间中的类间距离, 这是对先前提取的逐像素特征对齐和类内特征变化的补充.

2 方法

本文首先给出模型框架概述以区分现有语义分割的知识蒸馏方法的工作以及 IDD 模型, 然后详细描述 IDDM 和 PIDM 方法.

2.1 模型概述

知识蒸馏是通过让学生网络通过蒸馏的过程学习教师网络暗知识, 进而提高学生网络精度的方法. 通常情况下, 知识蒸馏可以按照知识类型, 蒸馏方法和网络架构进行分类. 蒸馏过程中的知识可以是教师网络的 logit、中间层的输出、不同激活层或样本对里包含的

特征.

语义分割是一项密集的预测任务, 旨在为每个像素分配一个标签. 虽然现有基于知识蒸馏的语义分割方法已经取得了很好的进展, 但是它们主要集中在对齐像素特征和类内特征方差上. 它们的损失函数通常可以表示为:

$$L = L_{tar}(D(GT), D(F^s)) + \lambda L_{dis}\left(\varphi\left(F^T\right), \varphi\left(F^S\right)\right) \quad (1)$$

$$L = L_{tar}\left(D(GT), D\left(F^S\right)\right) = -\sum_{k=1}^{N} D(GT_k)\log\left(D\left(F_k^S\right)\right) \quad (2)$$

其中, L_{tar} 是交叉熵损失, GT 是真值, F^S 和 F^T 分别表示学生网络和教师网络的特征图. $\varphi(\cdot)$ 表示映射函数. $D(GT)$ 和 $D(F^S)$ 分别表示所有像素的真值和学生网络的类别概率分布. N 是像素数, $D(GT_k)$ 表示第 k 个像素的类别概率分布的真值, $D\left(F_k^S\right)$是学生网络生成的第 k 个像素的类别概率分布. λ 是一个超参数, 用于控制损失的权重. L_{dis} 是损失函数, 例如均方误差损失.

显然, 现有方法忽略了将教师网络中的类间间距离转移到学生网络中. 因此, 如图 1 所示, 本文使用 IDD 方法将类间距离和位置信息从教师传递给学生.

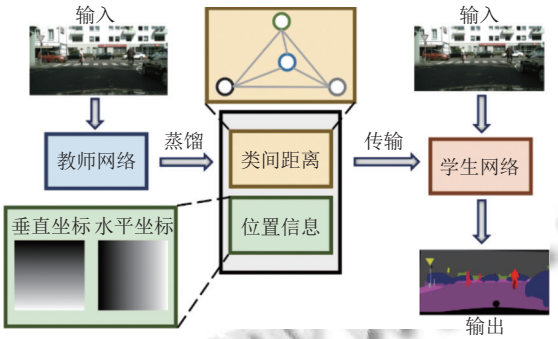

图 1 IDD 整体框架

2.2 类间距离蒸馏模块

语义分割是一项基于像素的分类任务. 受网络结构简单、参数少的限制, 学生网络的性能相对较差, 类间距离较小. 为了让学生网络学习到教师网络在特征空间中较大的类间距离, 本文提出了一种类间距离蒸馏模块, 通过让学生网络模拟教师网络输出的方法来解决这一问题.

如图 1 所示, 首先构造一个图 $g=\{v, e\}$来对类间类别距离进行编码, 其中 $v=\{v_i \mid i=1,\cdots, N\}$是一组节点, N 表示被处理图像的类别总数, $e=\{e_{i,j} \mid i=1,\cdots, N; j=1,\cdots, N; i\neq j\}$表示一组边. v_i 表示第 i 类的标记, v_i 是通过对具有相同类别标签 i 的所有像素的特征求平均而获得的. $e_{i,j}$ 是 i 类和 j 类的类间代表之间的欧几里德距离, 其被定义为:

$$e_{i,j} = Dis\left(v_i, v_j\right) \quad (3)$$

式 (3) 表示 i 类和 j 类之间的特征距离, 而 Dis 是欧氏距离. 由于网络较深, 参数众多, 教师网络的课际距离较大. 受这一特性的启发, 为了使学生网络能够在类间距离方面更好地模拟教师网络, 采用类间距离损失函数 L_{id}, 其定义为:

$$L_{id} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\left(e_{i,j}^T - e_{i,j}^S\right)^2,\ i \neq j \quad (4)$$

其中, $e_{i,j}^T$和$e_{i,j}^S$分别代表教师网络和学生网络中的 $e_{i,j}$. 在网络的训练过程中, 我们用一个二维数组来存储图 $g=\{v, e\}$中的节点和边长, 通过上述的损失函数 L_{id} 让学生网络存储图的二维数组来模仿教师网络中的改图. 在学生网络的持续性训练过程中, 通过最小化损失函数 L_{id}, 能够有效学习到教师网络在特征空间中较大的类间距离, 从而有效提高学生网络的语义分割精度.

2.3 位置信息蒸馏模块

语义分割是一个依赖于位置信息的任务. 例如, 同样的一个人在自行车上是的类别是骑行者, 在马路上的类别是行人. Islam 等人[11] 证明了 CNN 具有对位置信息进行编码的能力. 受其启发, 本文进一步引入了位置信息提取模块, 以增强学生网络预测位置信息的能力. 因此, 学生网络可以在其输出特征中编码更多的位置信息, 这可以用来提高语义分割网络的精度.

具体来说, 本文使用 $A\in \mathrm{R}^{C\times H\times W}$ 来表示输入特征图. 首先, 将 A 输入到预先训练的位置信息网络中, 我们的位置信息网络由一个 3 层的 MLP 组成, MLP 网络未经过预训练. 如图 2 所示, 以获得分别代表横坐标和纵坐标的位置信息掩膜 $P^{HOR}\in \mathrm{R}^{H\times W}$ 和 $P^{VER}\in \mathrm{R}^{H\times W}$. 在 $P^{HOR}\in \mathrm{R}^{H\times W}$ 中, 每一列都有相同的值, 我们使用 V_j^{HOR} ($j\in[1, H]$) 来表示列 j 的值, 其中 $V_j^{HOR}=j$. 在 P^{VER}, 每一行都有相同的值, 我们使用 V_i^{VER} ($i\in[1, W]$) 来表示行 i 的值, 其中 $V_i^{VER}=i$.

采用损失函数 L_{pi} 将教师网络的位置信息传递给学生网络:

$$L_{pi}=\frac{1}{2}L_{pi}^{HOR}+L_{pi}^{VER} \tag{5}$$

其中,

$$L_{pi}^{HOR}=\sum_{j=1}^{H}\left\|\frac{Q_j^{HOR_T}}{\left\|Q_j^{HOR_T}\right\|_2}-\frac{Q_j^{HOR_S}}{\left\|Q_j^{HOR_S}\right\|_2}\right\|_2 \tag{6}$$

$$L_{pi}^{VER}=\sum_{i=1}^{W}\left\|\frac{Q_i^{VER\_T}}{\left\|Q_i^{VER\_T}\right\|_2}-\frac{Q_i^{VER\_S}}{\left\|Q_i^{VER\_S}\right\|_2}\right\|_2 \tag{7}$$

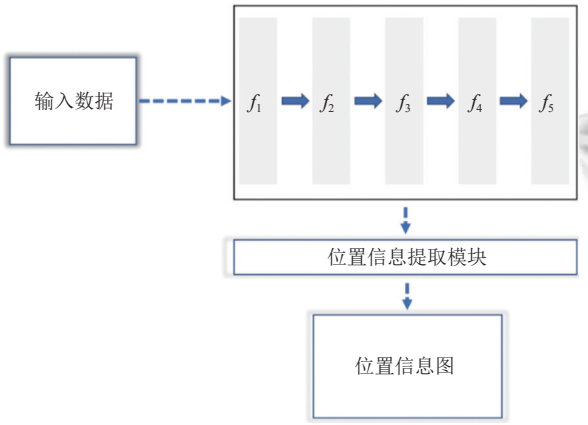

图 2 位置信息提取网络, 用于提取网络编码的位置信息

分别在水平和垂直方向上表示 L_{pi}. $Q_j^{HOR_T}$和$Q_j^{HOR_S}$以向量化的形式表示由教师网络和学生网络产生的 P^{VER} 的第 j 列. 与上述的表达方式类似, $Q_i^{VER_T}$和$Q_i^{VER_S}$以矢量化的形式表示由教师网络和学生网络产生的 P^{HOR} 的第 i 行. 在学生网络的训练过程中, 通过最小化该损失函数, 能够让学生网络在水平和竖直方向上学习教师网络编码的位置信息, 从而提高学生网络编码位置信息的能力, 进而增强学生网络语义分割的精度.

2.4 损失函数

本文同时还采用通道损失函数 L_{cw} 来最小化教师网络和学生网络之间的通道概率图的 Kullback-Leibler (KL) 散度. IDD 方法的最终损失函数公式如下:

$$L=L_{skd}+\lambda_1\cdot L_{cw}+\lambda_2\cdot L_{id}+\lambda_3\cdot L_{pi} \tag{8}$$

其中, L_{skd} 是用于语义分割的结构化损失函数, λ_1, λ_2 和 λ_3 是用于平衡不同损失的超参数.

3 实验分析

为验证方法的有效性, 本文在 3 个主流的公开数据集: Cityscapes[20]、Pascal VOC[21] 和 ADE20K[22] 上进行了全面的实验.

3.1 数据集介绍

(1) Cityscapes

Cityscapes 包含 5 000 张经过精细注释的城市驾驶场景图像. 它由 2 975、500 和 1 525 幅图像组成, 分别用于训练、验证和测试. 它被标注了 19 个语义类别. 每张图像的分辨率为 2048×1024. 在本文的实验中, 本文不使用粗糙标记的图像.

(2) Pascal VOC

Pascal VOC 由 1 464 幅用于训练的图像、1 449 幅用于验证的图像和 1 456 幅用于测试的图像组成. 它涵盖了 20 个前景对象类和 1 个背景类.

(3) ADE20K

ADE20K 是麻省理工学院发布的一个具有挑战性的场景解析数据集, 它包含 20k、2k、3k 图像和 150 个用于训练、验证和测试的类.

3.2 评价指标

使用每个类别的交集/并集 (IoU) 和所有类别的平均 IoU (mIoU) 作为指标衡量分割精度. 模型参数 (Params) 的总数用于测量模型大小. 本文采用分辨率为 512×1024 的输入图像来计算每秒浮点运算次数 (FLOPs), 这是衡量模型复杂性的一个通用指标. 来评价检测器的性能.

3.3 实验细节

本文在与文献 [18] 相同的教师和学生网络上进行实验. 具体来说, 在本文所有实验中, 教师网络使用的是在 ImageNet 上预先训练好的 PSPNet. 对于学生网络, 本文在不同的分段体系结构上进行实验, 如具有代表性的 PSPNet 和 DeepLab 模型以及 ResNet18 和 ESPNet 的主干, 以验证本文的 IDD 方法的有效性.

实验使用 PyTorch 平台来实现本文算法. 与文献 [18] 方法类似, 本文通过小批量随机梯度下降 (SGD) 训练本文的学生网络, 迭代次数为 40 000 次. 动量和重量衰减分别设置为 0.9 和 0.000 5, 同时应用多项式学习率策略, 基本学习率和功率分别设置为 0.01 和 0.9. 输入图像裁剪为 512×512, 同时应用随机缩放和随机翻转来扩充数据.

本文还在 Cityscapes 测试集上评估了不同的知识蒸馏的方法的性能, 例如 SKD、IFVD 和 CD. 采用的学生网络为 ESPNet、PSPNet-R18(0.5) 和 PSPNet-R18. 实验结果列于表 1. 当采用 ESPNet 作为学生网络时,

本文的方法在验证集和测试集上分别获得了 7.47% 和 7.05% 的显著提高. 与传递类内特征方差的 SKD 和传递通道特征的 CD 相比, 该方法的性能分别提高了 3.74% 和 1.60%. 使用 IDD 后, PSPNet-R18(0.5) 的性能从 61.17% 提高到 69.76%, 在验证集上分别比 IFVD 和 CD 提高了 6.41% 和 1.19%. 当采用 PSPNet-R18 作为学生网络时, 加入 IDD 方法, 增益达到 7.50% (70.09% vs. 77.59%), 分别比 IFVD 和 CD 高出 3.05% 和 1.69%. 实验结果表明, IDD 在语义分割方面优于以往的知识蒸馏方法.

表 1 基于不同知识蒸馏方法的语义分割模型在 Cityscapes 数据集

Method	mIoU (%)		Params	FLOPs
	Val	Test	(M)	(G)
T:PSPNet-R101	78.50	78.40	70.43	574.9
S:ESPNet	61.40	60.30	0.363 5	4.422
ESPNet+SKD	63.80	62.00	0.363 5	4.422
ESPNet+IFVD	65.13	63.07	0.363 5	4.422
ESPNet+CD	67.27	65.32	0.363 5	4.422
ESPNet+IDD	68.37	67.01	0.363 5	4.422
S:PSPNet-R18(0.5)	61.17	—	3.271	31.53
PSPNet-R18(0.5)+SKD	61.60	60.05	3.271	31.53
PSPNet-R18(0.5)+IFVD	63.35	63.68	3.271	31.53
PSPNet-R18(0.5)+CD	68.57	66.75	3.271	31.53
PSPNet-R18(0.5)+IDD	69.33	68.45	3.271	31.53
S:PSPNet-R18	70.09	67.60	13.07	125.8
PSPNet-R18+SKD	72.70	71.40	13.07	125.8
PSPNet-R18+IFVD	74.54	72.74	13.07	125.8
PSPNet-R18+CD	75.90	74.58	13.07	125.8
PSPNet-R18+IDD	77.43	76.23	13.07	125.8

如图 3 所示, 采用点状图来描述不同网络 (OCRNet[23]、DeepLabV3、FCN、ANN 和 PSPNet) 的参数和精度. 通过使用 IDD 方法, 以 PSPNet-R18(1.0) 作为学生网络主干的模型比 FCN 和 ANN分别高出 6.79% 和 0.08%.

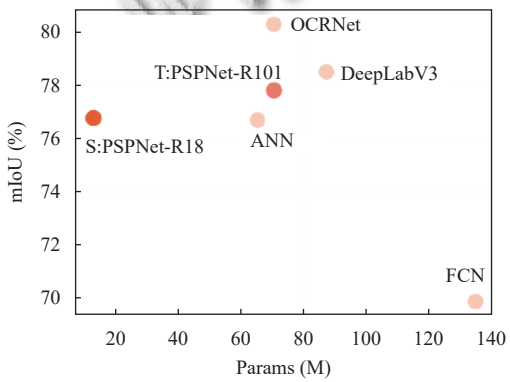

图 3 Pascal VOC 验证集上不同模型的参数和 mIoU 的比较

采用 ResNet18 和 MobileNetV2 作为学生网络, 在验证集上评估本文的方法. 结果如图 4. 以 ResNet18 作为学生网络的主干, 本文的方法将未经蒸馏的模型精度提高了 6.01%, 比 SKD、IFVD 和 CD 分别提高了 3.74%、2.74% 和 1.47%. 对于 MobileNetV2, IDD 比基准模型提高了 4.66%, SKD、IFVD 和 CD 分别提高了 2.88%、2.21% 和 0.89%.

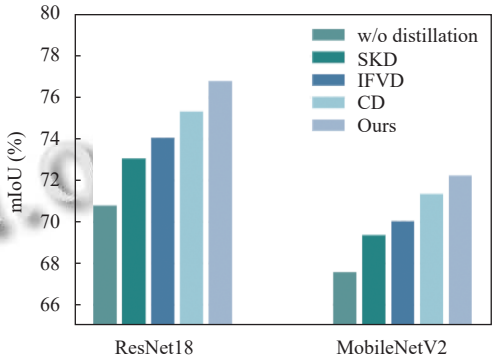

图 4 Pascal VOC 验证数据集上不同语义分割的知识蒸馏策略的比较

为了进一步验证方法的有效性, 本文在具有挑战性的数据集 ADE20K 上进行实验. 定量结果记录在表 2 中. 当学生模型建立在 ResNet18 上时, IDD 方法将学生模型的精度从 24.65% 提高到 27.65%, 比 SKD、IFVD 和 CD 分别提高了 2.67%、1.87% 和 0.89%. 以 MobileNetV2 为学生网络的主干, 与基准模型相比, IDD 获得了 5.72% 的性能提升, SKD、IFVD 和 CD 分别提高了 4.04%、3.50% 和 1.19%.

表 2 针对语义分割的不同知识蒸馏方法在 ADE20K 验证数据集上的精度比较

Method	mIoU (%)	Params (M)
T:PSPNet-R101	44.94	70.43
S:PSPNet-R18	24.65	13.07
PSPNet-R18+SKD	25.02	13.07
PSPNet-R18+IFVD	25.82	13.07
PSPNet-R18+CD	26.80	13.07
PSPNet-R18+IDD	27.69	13.07
S:PSPNet-MNV2	23.21	2.15
PSPNet-MNV2+SKD	24.89	2.15
PSPNet-MNV2+IFVD	25.43	2.15
PSPNet-MNV2+CD	27.74	2.15
PSPNet-MNV2+IDD	28.93	2.15

此外, 如图 5 所示, 本文使用 PSPNet-R18(1.0) 作为学生网络来计算每个类别的 mIoU, 并与两种最先进的方法进行了比较. 由于本文算法使得学生网络具有大的类间距离和丰富的位置信息, 它在一些特定类别

的识别上表现良好, 如骑手, 轿车, 公交车等.

同时, 本文还在 Cityscapes 上进行了消融实验, 以评估本文提出的不同模块的效果. 本文的损失函数由 4 部分组成 L_{skd}, L_{cw}, L_{id}, 和 L_{pi}. $L = L_{skd} + \lambda_1 \cdot L_{cw} + \lambda_2 \cdot L_{id} + \lambda_3 \cdot L_{pi}$.

为了探索每个损失项目的有效性, 使用评估指标 mIoU 在 Cityscapes 验证数据集上进行消融实验. 教师网络采用 PSPNet, 主干网络采用 ResNet101 (T:PSPNet-R101); 学生模型采用 PSPNet, 主干网络为 ResNet18 (S:PSPNet-R18), 也在 ImageNet 中进行预训练. 结构化 KD 损失 L_{skd} 将学生网络 S:PSPNet-R18 的性能从 70.09% 提高到 73.03%. 通道式 KD 损失 L_{cw} 进一步将学生模型提高到 75.78%. 通过采用本文的类间距离蒸馏方法, 增益增加了 5.34% (76.43% vs. 70.09%).

此外, 在应用本文的位置信息损失 L_{pi}, 轻量级学生网络 S:PSPNet-R18 的准确度达到 77.59%, 接近教师网络 T:PSPNet-R101 的准确度, 其 mIoU 值为 78.56%. 实验结果证明了本文方法的有效性.

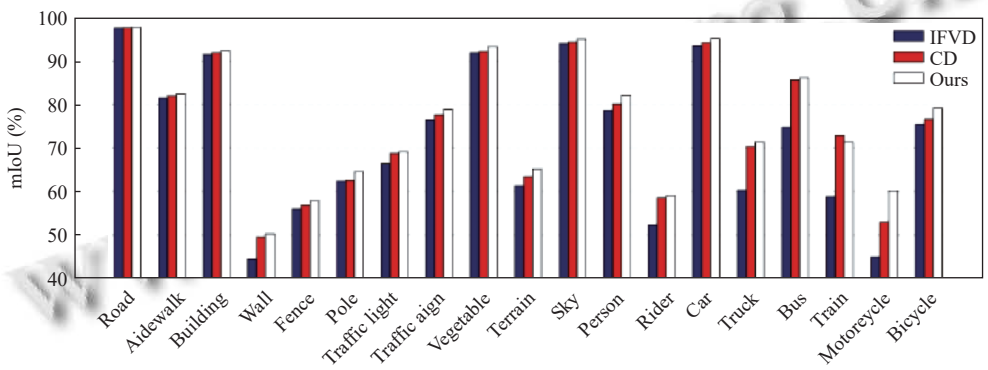

图 5 Cityscapes 数据集上基于知识蒸馏的语义分割方法的不同类别的 mIoU

4 结论

本文提出了一种新的语义分割知识提取方法 (IDD), 帮助学生模型在特征空间中具有较大的类间距离和丰富的位置信息. 具体来说, 本文提出了类间距离提取模块 (IDDM) 和位置信息提取模块 (PIDM), 以将类间距离和位置信息从教师网络转移到学生网络. 在 3 个著名的语义分割数据集上, 分别验证了 IDD 方法的有效性, 该方法不仅获得了针对知识蒸馏的语义分割方案中最高的精度, 而且对其他语义分割模型也是有用的.

消融实验表明, 本文探索的两个模块能使学生网络更好地模拟教师网络. 通过在 Cityscapes、Pascal VOC 和 ADE20K 这 3 个公开数据集上进行大量实验, 验证了该方法的有效性.

参考文献

1 Dong GS, Yan Y, Shen CH, *et al*. Real-time high-performance semantic image segmentation of urban street scenes. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3258–3274. [doi: 10.1109/TITS.2020.2980426]

2 Kemker R, Salvaggio C, Kanan C. Algorithms for semantic segmentation of multispectral remote sensing imagery using deep learning. ISPRS Journal of Photogrammetry and Remote Sensing, 2018, 145: 60–77. [doi: 10.1016/j.isprsjprs.2018.04.014]

3 Tu ZG, Li HY, Zhang DJ, *et al*. Action-stage emphasized spatiotemporal VLAD for video action recognition. IEEE Transactions on Image Processing, 2019, 28(6): 2799–2812. [doi: 10.1109/TIP.2018.2890749]

4 Paszke A, Chaurasia A, Kim S, *et al*. ENet: A deep neural network architecture for real-time semantic segmentation. arXiv:1606.02147, 2016.

5 Mehta S, Rastegari M, Caspi A, *et al*. ESPNet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 561–580.

6 Zhao HS, Qi XJ, Shen XY, *et al*. ICNet for real-time semantic segmentation on high-resolution images. Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich: Springer, 2018. 418–434.

7 Fan MY, Lai SQ, Huang JS, *et al*. Rethinking BiSeNet for real-time semantic segmentation. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern

Recognition. Nashville: IEEE, 2021. 9711–9720.
8 Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv:1503.02531, 2015.
9 Wang YK, Zhou W, Jiang T, *et al*. Intra-class feature variation distillation for semantic segmentation. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 346–362.
10 Shu CY, Liu YF, Gao JF, *et al*. Channel-wise knowledge distillation for dense prediction. Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal: IEEE, 2021. 5291–5300.
11 Islam MA, Jia S, Bruce NDB. How much position information do convolutional neural networks encode? Proceedings of the 8th International Conference on Learning Representations. Addis Ababa: OpenReview. net, 2020.
12 水文泽, 孙盛, 余旭, 等. 轻量化卷积神经网络在 SAR 图像语义分割中的应用. 计算机应用研究, 2021, 38(5): 1572–1575, 1580. [doi: 10.19734/j.issn.1001-3695.2020.05.0150]
13 熊炜, 童磊, 金靖熠, 等. 基于卷积神经网络的语义分割算法研究. 计算机应用研究, 2021, 38(4): 1261–1264. [doi: 10.19734/j.issn.1001-3695.2019.12.0705]
14 冯兴杰, 孙少杰. 一种融合多级特征信息的图像语义分割方法. 计算机应用研究, 2020, 37(11): 3512–3515. [doi: 10.19734/j.issn.1001-3695.2019.07.0249]
15 Zhao HS, Shi JP, Qi XJ, *et al*. Pyramid scene parsing network. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 6230–6239.
16 Chen LC, Papandreou G, Kokkinos I, *et al*. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFS. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834–848. [doi: 10.1109/TPAMI.2017.2699184]
17 Yuan YH, Huang L, Guo JY, *et al*. OCNet: Object context network for scene parsing. arXiv:1809.00916, 2018.
18 Liu YF, Chen K, Liu C, *et al*. Structured knowledge distillation for semantic segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 2599–2608.
19 He T, Shen CH, Tian Z, *et al*. Knowledge adaptation for efficient semantic segmentation. Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach: IEEE, 2019. 578–587.
20 Cordts M, Omran M, Ramos S, *et al*. The Cityscapes dataset for semantic urban scene understanding. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 3213–3223.
21 Everingham M, Eslami SMA, van Gool L, *et al*. The Pascal visual object classes challenge: A retrospective. International Journal of Computer Vision, 2015, 111(1): 98–136. [doi: 10.1007/s11263-014-0733-5]
22 Zhou BB, Zhao H, Puig X, *et al*. Scene parsing through ADE20K dataset. Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 5122–5130.
23 Yuan YH, Chen XL, Wang JD. Object-contextual representations for semantic segmentation. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 173–190.

(校对责编: 孙君艳)