计算机科学
COMPUTER SCIENCE

基于多教师网络模型的半监督语义分割方法

许华杰，肖毅烽

引用本文

许华杰，肖毅烽. 基于多教师网络模型的半监督语义分割方法[J]. 计算机科学，2023，50(12)：279-284.

XU Huajie, XIAO Yifeng. Semi-supervised Semantic Segmentation Method Based on Multiple Teacher Network Model [J]. Computer Science, 2023, 50(12): 279-284.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221000245

基于多教师网络模型的半监督语义分割方法

许华杰[1,2,3,4]　肖毅烽[1]
1 广西大学计算机与电子信息学院　南宁 530004
2 广西多媒体通信与网络技术重点实验室　南宁 530004
3 广西高校并行分布与智能计算重点实验室　南宁 530004
4 广西智能数字服务工程技术研究中心　南宁 530004

摘　要　基于一致性正则化的方法在半监督语义分割任务中展现出了较好的性能,这类方法通常涉及两个角色:一个显式或隐式的教师网络和一个学生网络。其中学生网络通过最小化两个网络对不同扰动样本预测结果之间的一致性损失实现训练。但是来自单个教师网络的不可靠预测可能会导致学生网络学习到错误的信息。通过将平均教师模型 MT 的单教师网络扩展为多教师网络,提出了多平均教师网络(Multiple Mean Teacher Network,MMTNet)模型,使学生网络从多个教师网络的平均预测结果进行学习,有效降低单个教师网络预测错误的影响。此外,MMTNet 通过对无标签数据进行强、弱数据增强的方式对无标签数据进行数据扰动,增加了无标签数据的多样性,在一定程度上缓解了学生网络和教师网络之间存在的耦合问题,避免了学生网络对教师网络的过度拟合,从而进一步降低了教师网络进行伪标签预测错误时所产生的影响。在 PASCAL VOC 2012 扩充数据集上的实验结果表明,所提出的多平均教师网络 MMTNet 模型可获得比其他目前主流的半监督语义分割方法更高的平均交并比,且实际分割效果更优。

关键词:半监督学习;语义分割;平均教师模型;多教师网络;一致性正则化
中图法分类号　TP391

Semi-supervised Semantic Segmentation Method Based on Multiple Teacher Network Model

XU Huajie[1,2,3,4] and XIAO Yifeng[1]
1 College of Computer and Electronic Information,Guangxi University,Nanning 530004,China
2 Guangxi Key Laboratory of Multimedia Communications and Network Technology,Nanning 530004,China
3 Key Laboratory of Parallel,Distributed and Intelligent Computing,Nanning 530004,China
4 Guangxi Intelligent Digital Services Research Center of Engineering Technology,Nanning 530004,China

Abstract　The methods based on consistency regularization show better performance in semi-supervised semantic segmentation task. Such methods usually involve two roles,an explicit or implicit teacher network,and a student network which is trained by minimizing the consistency loss between the prediction results of two networks for different perturbation samples. But unreliable predictions from a single-teacher network may cause the student network to learn wrong information. By extending the mean teacher(MT) model to the multiple teacher network,multiplemeanteacher network(MMTNet) is proposed to make the student network learn from the average prediction results of multiple teacher networks,which can effectively reduce the impact of single-teacher network prediction errors. In addition,MMTNet implements data perturbation of unlabeled data by applying strong data augmentation and weak data augmentation to the unlabeled data,which increases the diversity of unlabeled data,alleviates the coupling problem between student network and teacher network to a certain extent and avoids the overfitting of student network to teacher network,so as to further reduce the impact of pseudo-label prediction errors in the teacher network. Experimental results on VOC 2012 augmented dataset show that the proposed multiple mean teacher network model MMTNet can achieve higher mean intersection over union than other mainstream semi-supervised semantic segmentation methods,and the actual segmentation performance is better.

Keywords　Semi-supervised learning,Semantic segmentation,Mean teacher model,Multi-teacher network,Consistency regularization

到稿日期:2022-10-31　返修日期:2023-03-29
基金项目:广西科技计划项目(2017AB15008);崇左市科技计划项目(FB2018001)
This work was supported by the Science and Technology Plan Project of Guangxi(2017AB15008) and Science and Technology Plan Project of Chongzuo(FB2018001).
通信作者:许华杰(hjxu2009@163.com)

1 引言

全监督语义分割通过从大量的语义分割标签数据中学习，取得了巨大的成功[1-2]。然而，全监督语义分割一方面需要大量的数据样本进行学习，另一方面需要人工参与，为数据样本进行像素级标注。像素级标注任务需要耗费大量的人力和时间成本，据研究表明，在一个对象上绘制语义分割注释比绘制边界框注释要多耗费 8 倍的时间，比只标注该对象的类别要多耗费 78 倍的时间[3]。为了突破上述限制，半监督语义分割成为一个重要的研究方向，即使用少量有标签数据和大量无标签数据共同完成分割模型的训练，从而降低语义分割任务对大量像素级标注数据的需求。

在半监督语义分割领域，一致性正则化方法得到了广泛的研究[4-6]，它基于平滑假设和聚类假设，强制网络对包括网络扰动、特征扰动以及数据扰动下的样本进行一致性的预测。一致性正则化方法的基础是假设网络对未标注的图像数据进行准确预测，使得对图像施加的微弱扰动不会影响网络的预测结果，即网络对不同扰动下的样本应该进行一致性的预测，尽管这一假设在半监督语义分割任务中并不一定都能得到满足。在一致性正则化的半监督方法中，平均教师模型(Mean Teacher network，MT)[7]最初用于半监督分类任务，CutMix-Seg[4]对 MT 模型进行扩展后将其用于半监督语义分割任务中。MT 模型包括一个学生网络和一个教师网络，学生网络通过学习教师网络预测输出的结果进行训练。然而，来自教师网络的预测结果可能是不可靠的[8]，从而导致学生网络从单个教师网络的预测输出结果中所学习到的信息可能是错误的。另一方面，教师网络的参数是通过学生网络参数的指数移动平均(Exponential Moving Average，EMA)进行更新的，因此学生网络和教师网络对同一个输入可能会得到相似的预测，从而导致两者之间出现严重的耦合问题。

针对单个教师网络预测结果作为伪标签可能不可靠的问题，本文通过将 MT 模型中的单教师网络扩展为多教师网络，取多个教师网络预测结果的平均值得到的伪标签作为无标签数据的训练目标，提出了多平均教师网络 MMTNet 模型。同时，受到文献[9]在自训练方法中通过数据扰动的方式缓解分割模型在多阶段重训练过程中存在的耦合问题的启发，本文提出的多教师网络模型 MMTNet，在训练过程中通过数据扰动的方式增加无标签数据的多样性，从而在一定程度上缓解了学生网络和教师网络存在的耦合问题，避免了学生网络对教师网络的过度拟合，从而进一步降低了教师网络进行伪标签预测错误时所产生的影响。

实验结果表明，本文提出的多平均教师网络模型 MMTNet 在 PASAL VOC 2012 扩充数据集上取得了优于其他目前主流方法的半监督语义分割效果。

2 相关工作

半监督学习目前的方法主要有两个分支，分别是基于自训练的半监督方法和基于一致性正则化的半监督方法。其中，基于自训练的半监督方法使用少量有标签数据训练得到的模型为无标签数据进行伪标签的预测，进而利用无标签数据进行半监督学习，代表性方法有 FixMatch[10]和 Dash[11]。而与本文密切相关的是基于一致性正则化的半监督方法，其基于平滑假设和聚类假设，强制网络对各种扰动下的样本进行一致性的预测，其中的扰动包括数据扰动、特征扰动、网络扰动 3 种形式。数据扰动是通过对输入数据进行随机的数据增强来实现扰动，然后让模型对不同数据增强的同源图像进行一致性预测。例如，FixMatch[10]使用弱数据增强样本的网络预测来监督强数据增强样本的学习，使网络对弱数据增强和强数据增强的样本预测尽可能相似。特征扰动指使用多个解码器对图像的中间特征进行一致性预测，从而进行特征层面的扰动。例如，CCT[6]使用不同的解码器对编码器输出的特征进行一致性约束。网络扰动指使用两个结构相同但参数不同的网络对输入数据进行预测，并在两个网络预测之间进行一致性约束。例如，MT[7]通过最小化学生网络预测输出和教师网络预测输出的损失来进行一致性学习。

半监督语义分割的早期方法使用对抗生成网络 GAN 作为无标签数据的监督信号[12-13]。然而 GAN 存在训练不稳定、梯度消失等问题。除了 GAN 外，部分研究通过其他的辅助网络来解决半监督语义分割存在的不同问题，例如 US-RN[14]提出无偏子类正则化网络，通过从平衡子类中探索类无偏分割来解决类不平衡问题；ELN[15]提出误差定位网络，将图像及其分割预测图作为输入来识别伪标签中可能预测错误的像素点。受到半监督学习的成功启发，半监督语义分割采用自训练方法和一致性正则化方法来利用无标签数据。在自训练方法中，Ke 等[16]提出了一个三阶段的自训练框架，在不同阶段生成伪标签并提高伪标签的质量；U2PL[17]在自训练过程中，基于熵将预测图中的可靠像素点作为伪标签，将不可靠像素点视为相应类别的负样本。而在与本文密切相关的、基于一致性正则化的半监督语义分割方法中，CPS[5]基于网络扰动，让两个初始化不同但结构相同的网络互相使用对方生成的预测结果进行交叉伪监督；CutMix-Seg[4]对 MT[7]模型进行了扩展，将经过不同数据扰动的图像分别输入学生网络和教师网络中，然后使用教师网络预测结果对学生网络预测结果进行监督，并将其用于语义分割任务中；Pseudo-Seg[18]沿用 FixMatch[10]的框架，将弱数据增强到强数据增强的一致性扩展到分割领域中。

3 基于多教师网络模型的半监督语义分割方法

3.1 问题定义

给定一个带有 n 张有标签图像的数据集 D_l 和一个带有 m 张无标签图像的数据集 D_u，其中 $n \ll m$，半监督语义分割旨在使用数据集 D_l 和 D_u 共同进行分割模型的训练。半监督语义分割方法的优化目标通常包括最小化有标签数据的损失和无标签数据的损失，如式(1)所示：

$$L = L_s + \lambda L_u \tag{1}$$

其中，L_s 是有标签数据 D_l 的监督损失，L_u 是无标签数据 D_u 相关的损失，λ 用于权衡有标签数据 D_l 的监督损失和无标签数据 D_u 相关的损失在网络训练过程中的重要性。

3.2 多教师网络模型

针对单个教师网络的预测结果可能不可靠的问题，将

MT 模型中的单教师网络扩展为多教师网络，提出了多教师网络模型 MMTNet，其包括一个学生网络 $S(\theta)$ 和 3 个教师网络 $T(\theta_1)$，$T(\theta_2)$ 和 $T(\theta_3)$，其中学生网络和 3 个教师网络拥有相同的网络结构，设 θ，θ_1，θ_2，θ_3 分别是 $S(\theta)$，$T(\theta_1)$，$T(\theta_2)$ 和 $T(\theta_3)$ 的网络参数，如图 1 所示。在多教师网络模型 MMTNet 的训练过程中，有标签数据 X_l 和无标签数据 X_u 只用于 MMTNet 学生网络的参数训练，3 个教师网络的参数通过学生网络参数的指数移动平均 EMA 进行更新。

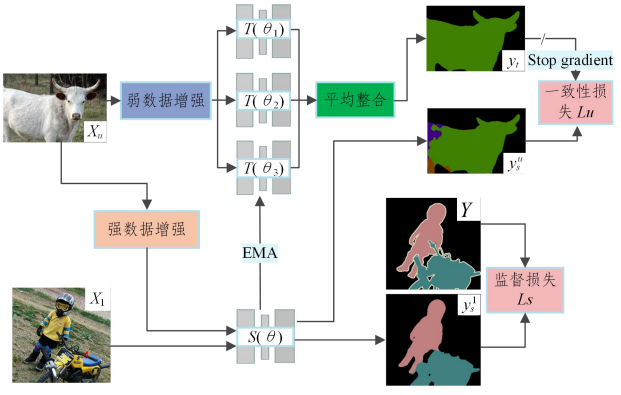


图 1 多平均教师网络模型(MMTNet)

Fig. 1 Multiple mean teacher network model(MMTNet)

对于有标签数据 X_l，学生网络使用标准的像素级交叉熵损失函数在具有真实标签的数据上进行监督学习；对于无标签数据 X_u，MMTNet 首先通过对无标签数据 X_u 进行强、弱数据增强的方式实现对无标签数据 X_u 的数据扰动，然后使用强、弱数据增强后的无标签样本进行学生网络的训练，相应核心部分的伪代码如算法 1 所示。

算法 1 MMTNet training on unlabeled data

Input: Unlabeled training set $D^u=\{X_u\}_{u=1}^m$, Student network $S(\theta)$ and three teacher network $\{T(\theta_i)\}_{i=1}^3$, Weak/Strong augmentations A^w/A^s

Output: Full train segmentation models $S(\theta)$ on D^u

1. for minibatch $\{x_k\}_{k=1}^t$ D^u do
2. for k ←1 to t do
3. $x_k \leftarrow A^w(x_k)$
4. for j←1 to 3 do
5. $Y_{tj}= T(x_k;\theta_j)$
6. Pseudo-labels $Y_t=\frac{1}{3}\sum_{j=1}^{3} Y_{tj}$
7. $x_k \leftarrow A^s(x_k)$
8. $S(\theta)$ prediction $Y_s^u= S(x_k;\theta)$
9. Update $S(\theta)$ to minimize $Lce(Y_s^u, Y_t)$
10. return $S(\theta)$

具体来说，将无标签数据 X_u 经过弱数据增强后输入教师网络中，得到 3 个教师网络的预测结果，然后 MMTNet 通过取 3 个教师网络预测结果的平均值获取伪标签 Y_t；同时，无标签数据 X_u 经过强数据增强后输入学生网络中，并得到预测结果 Y_s^u；获取上述的伪标签和预测结果后，按照式(2)，使用交叉熵函数最小化 Y_s^u 和 Y_t 之间的一致性损失 L_u，从而使用无标签数据 X_u 对学生网络的参数进行训练。

$$L_u=\frac{1}{|D_u|}\sum_{x\in D_u}\frac{1}{W\times H}\sum_{i=0}^{W\times H}(Lce(Y_s^u i, Y_t i)) \tag{2}$$

其中，$Lce(\ )$ 为交叉熵损失函数，W 和 H 是图像的宽和高，i 表示图像中像素点的序号。上述弱数据增强只改变图像的大小和位置关系，包括随机翻转、随机裁剪等策略；而强数据增强会改变图像的色彩性质，包括随机灰度、随机改变图像的亮度、对比度、饱和度等策略。

3.3 方法性能分析

作为方法的核心，算法 1 的主体包括一个学生网络和 3 个教师网络，它们具有相同的网络架构并且可以使用现成的网络模型，例如 DeeplabV3＋网络[1]。对于单个卷积神经网络而言，其整体的时间复杂度为 $O(\sum_{l=1}^{D} M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l)$，空间复杂度为 $O(\sum_{l=1}^{D} K_l^2 \cdot C_{l-1} \cdot C_l + \sum_{l=1}^{D} M_l^2 \cdot C_l)$，其中 D 表示网络层数，l 表示网络的第 l 个卷积层，M 表示卷积核输出特征图的边长，K 和 C 分别表示卷积核的大小以及输出的通道数。根据算法 1 的流程可知，算法 1 在使用带有 m 张图像的无标签数据集 D^u 进行学生网络一轮迭代训练时，其时间复杂度主要由以下 3 部分组成：1)无标签数据 D^u 进行强、弱数据增强的耗时 $A=O(m \cdot W \cdot H)$，其中 W 和 H 分别表示输入图像的宽和高；2)3 个教师网络为无标签数据 D^u 计算预测图，进而得到伪标签的耗时 $B=O(3 \cdot m \cdot \sum_{l=1}^{D} M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l)$；3)单个学生网络的训练耗时 $C=O(m \cdot \sum_{l=1}^{D} M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l)$。因此算法 1 总的时间复杂度为 $A+B+C=(4 \cdot m \cdot \sum_{l=1}^{D} M_l^2 \cdot K_l^2 \cdot C_{l-1} \cdot C_l + m \cdot W \cdot H)$；而空间复杂度是单个学生网络和 3 个教师网络的空间复杂度之和，即 $O(4 \cdot \sum_{l=1}^{D} K_l^2 \cdot C_{l-1} \cdot C_l + 4 \cdot \sum_{l=1}^{D} M_l^2 C_l)$。

与取单个教师网络预测结果获取的伪标签相比，MMTNet 取多个教师网络预测结果的平均值得到的伪标签综合了多个教师网络预测结果的信息，其质量相对更稳定。当某个教师网络预测错误时，来自其他教师网络的正确预测结果的分歧信息能够在一定程度上纠正单个教师网络的预测错误，从而解决单个模型难以识别自身所产生的伪监督错误的问题。此外，MMTNet 通过在训练过程中对无标签数据进行强、弱数据增的方式实现数据扰动，能够增加无标签数据的多样性，在一定程度上缓解学生网络和教师网络存在的耦合问题，从而进一步降低教师网络进行伪标签预测错误时所产生的影响。

4 实验及结果分析

4.1 实验数据及评价指标

原始的 Pascal VOC 2012 数据集由 1 464 张训练集图像和 1 449 张验证集图像组成，包括 20 个前景类别和 1 个背景类别，其中前景类别包括人、动物、交通工具和家庭用品[19]。参照相关文献的做法，首先通过引入语义边界数据集(Semantic Boundaries Dataset，SBD)进行数据扩充，将训练集的图像扩充至 10 582 张，作为本文实验采用的 PASAL VOC 2012 扩充数据集的训练集；然后对训练集分别按照 1/16，1/8，1/4 的比例设置有标签数据集，对于其余数据，则忽略其标签以充当无标签数据使用[1,5]。实验过程中，在上述不同的

数据集划分标准下，将划分后的有标签数据集和无标签数据集作为训练集进行半监督学习，用于训练语义分割模型；而全监督基线模型仅使用按照比例划分后训练集中的有标签数据进行训练，即在 1/16，1/8，1/4 比例的数据划分标准下，分别使用 662，1 323，2 646 张图像进行模型的训练。所有模型的训练结果均在来源于原始 Pascal VOC 2012 数据集的验证集上进行分割效果的验证和评估。

性能评价指标方面，本文采用图像语义分割任务中常用的平均交并比 mIoU(mean Intersection over Union)对模型的分割效果进行评估。mIoU 是真实值和预测值集合的交集与并集之比，其计算式如式(3)所示：

$$mIoU=\frac{1}{K+1}\sum_{i=0}^{K}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{3}$$

其中，K 为数据集样本前景类别的总数，p_{ij} 为将第 i 类别的像素点预测为第 j 类别的数量。

4.2 实验环境及设置

本文基于 PyTorch 深度学习框架实现所提 MMTNet 模型，实验运行的硬件环境为 12th Gen Intel(R) Core(TM) i5-12400F 处理器(16 GB 运行内存)，NVIDIA GeForece RTX 3060 GPU(12 GB 显存)；软件环境为 PyTorch1.10.2，Python3.9。

为了将所提方法与其他目前主流的半监督语义分割方法进行公平比较，本文参照相关文献中的做法，对所提模型中的学生网络和 3 个教师网络均使用以 ResNet-50 作为骨干网络的 Deeplabv3＋网络[5]。对于 Deeplabv3＋编码器部分的 ResNet-50，其参数通过在 ImageNet 上预训练的 ResNet-50 模型进行初始化，初始学习率设置为 0.001；Deeplabv3＋的分割网络部分使用 Kaiming 初始化方法进行参数的随机初始化，初始学习率设置为骨干网络部分初始化学习率的 10 倍[5]。训练过程中，采用带动量的小批量随机梯度下降(Stochastic Gradient Descent，SGD)方法进行学生网络参数的迭代优化，其中 SGD 的权重衰减率设置为 0.001，批处理大小设置为 8，并采用 Poly 策略进行学习率的动态调整[1]。模型训练过程中的所有训练样本来自进行了弱数据增强的数据，包括对图像进行随机翻转，在 0.5～2.0 倍的范围内对图像的长和宽进行调整，将图像随机裁切成固定分辨率 321×321 的图像块。对于无标签数据，除了进行上述的弱数据增强外，还进行了强数据增强(Strong Data Augmentations，SDA)，包括对图像进行随机灰度化、模糊化以及随机改变图像的亮度、对比度和饱和度。按照上述实验设置，所提出的多教师网络模型 MMTNet 在 1/16，1/8，1/4 这 3 种比例的数据集划分标准下，分别进行 3 组实验。每组实验中，MMTNet 使用划分后的有标签数据集以及无标签数据集共同作为训练集，进行 120 轮次的迭代训练，然后将训练好的学生网络在 PASAL VOC 2012 数据集的验证集上使用 mIoU 评估指标进行模型训练效果的评估。

4.3 实验结果比较及分析

4.3.1 与主流方法的对比实验

为了验证所提方法的有效性和先进性，将提出的多教师网络模型 MMTNet 与其他当前主流的半监督语义分割方法进行比较，结果如表 1 所列。表 1 中，对比方法的实验结果均来源于文献[5]，其中，分数(如“1/16”)和数字(如“662”)分别表示有标签数据的比例和数量，Baseline 表示全监督基线模型，CPS† 指在 CPS 中使用 CutMix 方法[20]以进一步提升分割性能。

表 1 与其他方法在 mIoU 指标上的对比

Table 1 mIoU index comparison with other methods

(单位：%)

Method\label number	1/16(662)	1/8(1323)	1/4(2646)
Baseline	65.12	68.99	70.67
MT[7]	66.77	70.78	73.22
CCT[6]	65.22	70.87	73.43
CutMix-Seg[4]	68.90	70.70	72.46
CPS[5]	68.2	73.2	74.2
CPS†[5]	71.98	**73.67**	74.90
MMTNet(ours)	**72.07**	73.54	**75.18**

通过对比表 1 中的全监督基线模型 Baseline 与其他半监督语义分割方法的结果可知，由于后者是在前者有标签训练样本的基础上，利用加入的无标签样本提供的信息提高模型训练的质量，因此分割性能相对于前者有所提高，其中尤以所提出的 MMTNet 模型的分割性能提升幅度最大。具体来说，在 1/16，1/8，1/4 比例的数据划分标准下，相比全监督基线模型，MMTNet 的分割效果在 mIoU 上分别提高了 6.95%，4.55%，4.51%，表明所提出的 MMTNet 能够利用额外的无标签数据进行半监督学习，从而增加训练样本的多样性，进而有效提高全监督模型的语义分割效果，而且有标签数据占比越小，效果的提升就越明显。

从表 1 的实验结果可知，在 1/16 比例的数据划分标准下，MMTNet 的分割 mIoU 比 CPS†，CPS，CutMix-Seg，CCT，MT 分别高出 0.09%，3.87%，3.17%，6.85%，5.30%；在1/8 比例的数据划分标准下，MMTNet 的分割 mIoU 比 CPS，CutMix-Seg，CCT，MT 分别高出 0.34%，2.84%，2.67%，2.76%，比 CPS† 低 0.13%；在 1/4 比例的数据划分标准下，MMTNet 的分割 mIoU 比 CPS†，CPS，CutMix-Seg，CCT，MT 分别高出 0.28%，0.98%，2.72%，1.75%，1.96%。上述实验结果表明，本文基于 MT 模型改进得到的、使用多个教师网络预测输出平均值作为无标签数据伪标签的多教师网络模型 MMTNet 相比改进前的原始 MT 模型的分割效果有很大的提升，分割性能优于大部分对比方法。虽然 MMTNet 在 1/8 比例的数据划分标准下的分割效果比 CPS† 低 0.13%，但是 MMTNet 在模型训练时，只需要计算学生网络进行反向传播时所需要的梯度，其计算量仅相当于需要进行双网络交叉伪监督的 CPS 所需计算量的一半，而 CPS† 通过 CutMix 方法对 CPS 的改进带来了额外的计算量，也就是说，在这种情况下 MMTNet 利用不到 CPS† 一半的计算量就达到了与其相当的分割性能。

4.3.2 有效性分析实验

相对于原始的 MT 模型，所提出的多教师网络模型 MMTNet 有以下两个改进点：1)将 MT 模型的单教师网络扩展为多教师网络，取多个教师网络预测结果平均值得到的伪标签作为无标签数据的训练目标；2)通过对无标签数据进行

数据扰动，在一定程度上缓解了学生网络和教师网络的耦合问题，从而进一步降低了教师网络进行伪标签预测错误时所产生的影响。为了验证 MMTNet 两个改进点的有效性，进行了两组相关的验证性实验。

1）多教师网络的有效性

多平均教师网络 MMTNet 和单教师网络 MT 之间的差别在于如何获取无标签数据的伪标签。图 2 给出了在 1/16，1/8，1/4 比例的数据划分标准下训练得到 3 个教师网络 MT1，MT2 和 MT3，分别通过以下两种方式获取伪标签的质量（用伪标签与真实标签之间的 mIoU 表示）：MT1，MT2，MT3 表示分别通过取对应的单个教师网络的预测输出获取伪标签；AveI（Average Integration，MMTNet 获取无标签的方法）表示通过取 3 个教师网络预测输出的平均值获取伪标签。从图 2 可以看到，通过 AveI 方法获取的伪标签，其质量比通过取单个教师网络的预测输出获取的伪标签的质量高，而取单个教师网络的预测输出获取的伪标签的质量往往较差且不稳定。

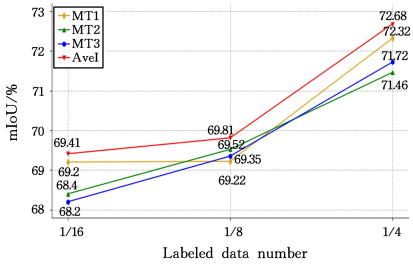
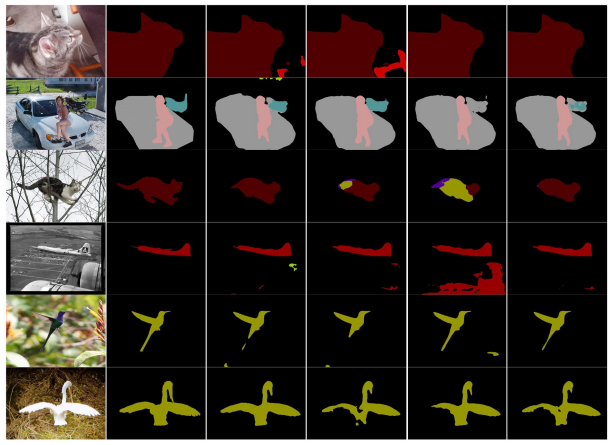


图 2 两种方式获得伪标签的质量（mIoU）对比

Fig. 2 Comparison of pseudo-label quality（mIoU）obtained by two ways

为了更直观地观察多教师网络 MMTNet 和单教师网络 MT 获取伪标签方式之间效果的差别，图 3 给出了在 1/8 比例的数据划分标准下，通过上述两种方法分别从 PASAL VOC 2012 扩充数据集的部分图像中所获取的伪标签示例。图 3 中“（a）”列对应的是输入图像，“（b）”列对应的是 Ground true，“（c）”—“（e）”列分别是通过“MT1”—“MT3”这 3 个教师网络获取的伪标签，“（f）”列为 MMTNet 通过取 3 个教师网络预测输出的平均值所获取的伪标签。

如图 3 所示，从单个教师网络获取的伪标签存在如下问题：（1）类别预测错误。例如，第一行的图像中，教师网络 MT1 和教师网络 MT2 将部分属于背景类别（“木根”）的像素点错误地预测为“椅子”类别；第二行的图像中，教师网络 MT3 将部分属于“摩托车”的像素点预测为“汽车”类别；第三行的图像中，教师网络 MT2 和教师网络 MT3 将属于“猫”的部分像素点预测为其他类别。而 MMTNet 所采用的 AveI 方法生成的伪标签由于综合了 3 个教师网络预测输出的结果，并没有出现上述类别预测错误的情况。（2）物体轮廓预测不完整。从第五行和第六行的图像可以看到，从单个教师网络预测结果获取的部分伪标签，存在物体轮廓预测不完整的情况。而 AveI 方法通过结合 3 个教师网络的预测输出生成伪标签，充分利用了 3 个教师预测输出之间的分歧信息，使伪标签中预测的物体具有较完整的轮廓信息。



（a）Input images （b）Ground truth （c）MT1 （d）MT2 （e）MT3 （f）AveI

图 3 两种方式获得的伪标签示例

Fig. 3 Examples of pseudo-labels obtained by two ways

2）数据扰动的有效性

MMTNet 在训练的过程中，通过对无标签数据进行强、弱数据增强的方式实现对无标签数据的数据扰动。为了验证数据扰动的有效性，本文对 MMTNet 进行是否使用数据扰动的对比实验，结果如表 2 所列。

表 2 数据扰动的对比实验结果

Table 2 Comparative experimental results of data perturbation

（单位：%）

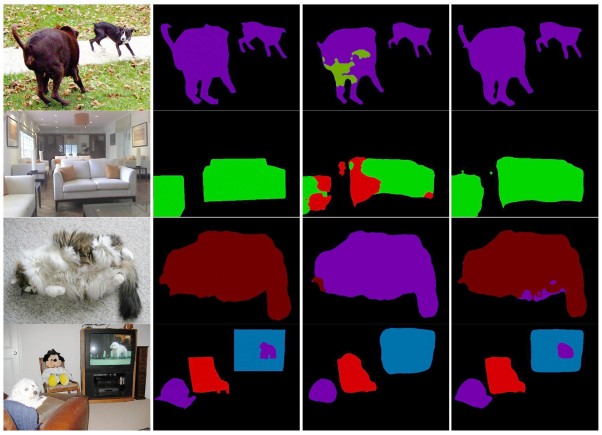
数据扰动	1/16（662）	1/8（1 323）	1/4（2 646）
×	69.12	71.43	73.76
√	72.07	73.54	75.18

从表 2 可知，在 1/16，1/8，1/4 比例的数据划分标准下，通过对无标签数据进行数据扰动，能够将多教师网络 MMTNet 的分割 mIoU 分别提高 2.95%，2.11% 和 1.42%，这表明对无标签数据进行数据扰动对模型性能的提高有明显作用。参考数据扰动在自训练方法领域中的应用以及文献[9]对数据扰动效果的分析可知，MMTNet 使用数据扰动后分割性能提升的主要原因是数据扰动能够增加无标签样本的多样性，在一定程度上缓解了学生网络和教师网络之间存在的耦合问题，从而降低教师网络进行伪标签预测错误时所产生的影响。

4.4 分割效果对比

图 4 给出了全监督基线模型和 MMTNet 模型在 PASAL VOC 2012 扩充数据集上部分图像的分割效果对比。从图 4 可见，全监督基线模型（对应于“SupOnly”列）在有标签数据不足的情况下，训练得到的模型对图像的预测存在以下问题：1）相似物体的类别预测错误。例如，在第一行的预测结果中，部分属于“狗”的像素点被预测为“牛”；在第二行的预测结果中，部分属于“沙发”的像素点被预测为“椅子”；在第三行的预测结果中，完全错误地将“猫”的像素点预测为“狗”。而所提出的 MMTNet 模型（对应于“MMTNet”列）并没有出现上述错误。全监督模型出现上述错误的可能原因是，少量的有标签数据不足以训练得出能够对相似物体进行区分的模型；而 MMTNet 模型通过数据抖动可有效地从无标签数据中学习到更多相似物体之间的差异信息，从而提高对相似物体的

辨别能力，使其能够预测出相似物体的正确类别。2）物体识别能力不足。在第四行的预测结果中，全监督模型没有识别出电视里面的“狗”，而 MMTNet 模型则可有效识别，主要得益于其采用的半监督学习方式可有效地从无标签数据中学习到更多对识别有帮助的信息。



(a) Input images (b) Ground truth (c) SupOnly (d) MMTNet

图 4 分割效果对比

Fig. 4 Comparison of segmentation effects

结束语 本文受 MT 模型的启发，提出了用于半监督语义分割的多平均教师网络模型 MMTNet。MMTNet 通过取多个教师网络预测结果的平均值的方式为无标签数据获取质量较高的伪标签，进而有效地利用无标签数据进行半监督学习；此外，MMTNet 通过对无标签数据进行数据扰动，能够增加无标签数据的多样性，缓解学生网络和教师网络之间存在的耦合问题，从而进一步降低教师网络进行伪标签预测错误时所产生的影响。在 PASCAL VOC 2012 扩充数据集上进行的相关实验验证了所提出的 MMTNet 模型的有效性，说明其可在训练的过程中充分利用无标签数据提高半监督语义分割模型的质量，改善语义分割的效果。

参 考 文 献

[1] CHEN L C, ZHU Y, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]// European Conference on Computer Vision (ECCV). 2018: 801-818.

[2] TIAN X, WANG L, DING Q. Review of Image Semantic Segmentation Based on Deep Learning[J]. Journal of Software, 2019, 30(2): 440-468.

[3] BEARMAN A, RUSSAKOVSKY O, FERRARI V, et al. What's the point: Semantic segmentation with point supervision[C]// European Conference on Computer Vision. Cham: Springer, 2016: 549-565.

[4] FRENCH G, LAINE S, AILA T, et al. Semi-supervised semantic segmentation needs strong, varied perturbations[C]// British Machine Vision Conference. 2020: 1-7.

[5] CHEN X, YUAN Y, ZENG G, et al. Semi-supervised semantic segmentation with cross pseudo supervision[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2613-2622.

[6] OUALI Y, HUDELOT C, TAMI M. Semi-supervised semantic segmentation with cross-consistency training[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 12674-12684.

[7] TARVAINEN A, VALPOLA H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[J]. Advances in Neural Information Processing Systems, 2017, 30: 2-7.

[8] REN Z, YEH R, SCHWING A. Not all unlabeled data are equal: Learning to weight data in semi-supervised learning[J]. Advances in Neural Information Processing Systems, 2020, 33: 21786-21797.

[9] YANG L, ZHUO W, QI L, et al. St++: Make self-training work better for semi-supervised semantic segmentation[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 4268-4277.

[10] SOHN K, BERTHELOT D, CARLINI N, et al. Fixmatch: Simplifying semi-supervised learning with consistency and confidence[J]. Advances in Neural Information Processing Systems, 2020, 33: 596-608.

[11] XU Y, SHANG L, YE J, et al. Dash: Semi-supervised learning with dynamic thresholding[C]// International Conference on Machine Learning. PMLR, 2021: 11525-11536.

[12] LIU S P, HONG J M, LIANG J P, et al. Medical Image Segmentation Using Semi-supervised Conditional Generative Adversarial Nets[J]. Journal of Software, 2020, 31(8): 2588-2602.

[13] LI Z X, ZHANG J, WU J L, et al. Semi-supervised adversarial learning based semantic image segmentation[J]. Journal of Image and Graphics, 2022, 27(7): 2157-2170.

[14] GUAN D, HUANG J, XIAO A, et al. Unbiased subclass regularization for semi-supervised semantic segmentation[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 9968-9978.

[15] KWON D, KWAK S. Semi-supervised semantic segmentation with error localization network[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 9957-9967.

[16] KE R, AVILES-RIVERO A I, PANDEY S, et al. A three-stage self-training framework for semi-supervised semantic segmentation[J]. IEEE Transactions on Image Processing, 2022, 31: 1805-1815.

[17] WANG Y, WANG H, SHEN Y, et al. Semi-supervised semantic segmentation using unreliable pseudo-labels[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 4248-4257.

[18] ZOU Y, ZHANG Z, ZHANG H, et al. PseudoSeg: Designing Pseudo Labels for Semantic Segmentation[C]// International Conference on Learning Representations. 2021: 1-8.

[19] EVERINGHAM M, ESLAMI S A, VAN GOOL L, et al. The pascal visual object classes challenge: A retrospective[J]. International Journal of Computer Vision, 2015, 111(1): 98-136.

[20] YUN S, HAN D, OH S J, et al. Cutmix: Regularization strategy to train strong classifiers with localizable features[C]// IEEE/CVF International Conference on Computer Vision. 2019: 6023-6032.

XU Huajie, born in 1974, Ph.D, associate professor, is a senior member of China Computer Federation. His main research interests include artificial intelligence and machine vision.

（责任编辑：何杨）