ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 39,No. 1,Jan. 2024,pp. 154－166
DOI:10. 16337/j. 1004-9037. 2024. 01. 014

http://sjcj. nuaa. edu. cn
E-mail:sjcj@nuaa. edu. cn
Tel/Fax:＋86-025-84892742

图引导的特征融合和分组对比学习的域自适应语义分割

赵伟枫[1]，谢明鸿[1]，张亚飞[1,2]，李华锋[1,2]

(1.昆明理工大学信息工程与自动化学院,昆明 650500;2.昆明理工大学云南省人工智能重点实验室,昆明 650500)

摘　要：在无监督域自适应语义分割任务中,有效地融合源域和目标域的特征以及解决不同类别像素数量分布不均衡的问题是提升跨域语义分割网络性能的关键。为了充分融合源域和目标域的特征,建立源域和目标域之间的长距离上下文关系,本文构建了双跨域图卷积网络,利用图卷积来引导源域和目标域的特征进行融合。本文分别构造了跨域位置相似矩阵和通道相似矩阵,提出了跨域位置图卷积和跨域通道图卷积。为了解决数据集中存在的类不平衡问题,同时提取到更多域不变特征,本文提出了分组对比学习策略,通过在组内构造正负样本,拉近2个域相同类之间的距离并拉远2个域不同类之间的距离。实验证明,本文提出的方法在数据集GTA5到Cityscapes和SYNTHIA到Cityscapes上的跨域语义分割均取得了良好的效果。

关键词：图卷积;对比学习;语义分割;域自适应

中图分类号：TP391.4　　**文献标志码**:A

Graph-Guided Feature Fusion and Group Contrastive Learning for Domain Adaptation Semantic Segmentation

ZHAO Weifeng[1], XIE Minghong[1], ZHANG Yafei[1,2], LI Huafeng[1,2]

(1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 2. Key Laboratory of Artificial Intelligence of Yunnan Province, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: Considering the problem of unsupervised domain adaption semantic segmentation, it is very important to establish a long-distance context relationship between the source domain and the target domain and how to solve the problem of unbalance distribution of different classes of pixels. we propose a dual cross-domain graph convolution network to exploit the long-distance context between source and target domain and fuse the feature of two domains. Specifically, we construct the position similarity matrix and channel similarity matrix of the cross domain and propose the cross-domain position graph convolution and cross-domain channel graph convolution. In order to solve the problem of unbalanced distribution of classes in the datasets and capture more domain invariant feature, we propose a group contrastive learning strategy to narrow the distance between the same class of two domains and widen the distance between the different classes of two domains by constructing positive and negative samples in the group. A large number of experiments show that our method achieves good performance on Urban Scene datasets GTA5 to Cityscapes and SYNTHA to Cityscapes.

基金项目:国家自然科学基金(62161015,61966021)。
收稿日期:2022-11-20;**修订日期**:2023-05-16

Key words: graph convolution; contrastive learning; semantic segmentation; domain adaptation

引　言

语义分割是对图像中的每一个像素进行分类,在自动驾驶、医疗图像分割以及场景识别等任务中有着广阔的应用前景。语义分割模型在训练过程中需要像素级标注的图像,而标注像素级的标签需要耗费大量的人力物力。目前,为解决语义分割数据的标注问题,主要采用计算机合成高质量的虚拟场景图像,同时自动生成像素级别的标签。在语义分割中常用的合成数据集有GTA5[1]和SYNTHIA[2]。但是,合成数据与现实场景数据之间存在着域的差异,利用合成数据训练的语义分割网络在现实场景数据上的表现不尽如人意。无监督域自适应方法可以将源域学到的知识迁移到目标域中[3-5],为了解决语义分割中的域偏移问题,可以将其应用到语义分割任务中。

无监督域自适应语义分割是利用带标签的源域样本和不带标签的目标域样本来训练分割网络,其本质是在不使用目标域标签的情况下,使网络学习到源域和目标域之间的域不变信息。文献[6-7]利用对抗思想进行域自适应语义分割,利用鉴别器和生成器之间的对抗学习来迫使生成器提取域不变信息。上述工作主要是从整体层面进行对抗,没有考虑图像中不同类别的像素数量分布不平衡的问题(类不平衡问题)。此外,在提取域不变信息的过程中,仅靠对抗损失来约束生成器提取两个域的公共特征来混淆鉴别器,导致网络不能提取到充足的域不变信息。文献[8]采用基于熵最小化的方法来进行无监督域自适应语义分割,利用目标域输出预测图的熵值大小来衡量预测是否准确。但这种方法会使预测概率值高的类别在熵损失函数中有较大的梯度,网络就会更倾向于迁移简单样本(像素数量占比较大的类别)而忽略难样本,导致难样本难以迁移,加剧了类不平衡问题,因此跨域效果较差。综上所述,以上方法都没有考虑两个域像素之间的关联性以及类不平衡问题,使语义分割网络的跨域性能较差。

在跨域语义分割任务中,同一类物体的风格(颜色、纹理等)不仅在不同域之间会有差异,在相同域中也会有一定的差异,这些差异会降低分类器分类像素的准确率。如果建立起域内和域间的长距离上下文关系,融合两个域的特征,就可以使两个域的像素之间相互关联,进而使编码器提取到更具有判别性的特征来提升分类器的性能。文献[9-11]利用图卷积建立了域内的长距离上下文依赖关系进行有监督的语义分割,并取得了良好的分割效果。利用图卷积可以建立长距离上下文依赖关系的特性,本文将其引入到无监督域自适应语义分割任务中来建立域间的长距离上下文依赖关系,使编码器能提取到更多的域不变信息。为了建立域内和域间的长距离上下文关系,融合源域和目标域的特征,本文提出了双跨域图卷积网络,如图1所示。图1中上半部分是现有方法的分割结果,下半部分是本文所提出的双域图卷积网络的分割结果,实验证明所提出的方法可以有效改善分割结果。

本文的主要贡献总结如下:

(1)构造了跨域的位置相似性矩阵和通道相似性矩阵,通过双跨域图卷积来更新图像特征图上的结点信息,建立域内和域间像素的长距离上下文依赖关系,使无监督域自适应分割网络能提取到更多的域不变信息。

(2)为了解决类不平衡问题,提出了分组对比学习方法,构造了分组对比损失函数,以进一步提取域不变特征。

(3)通过数据集GTA5到Cityscapes和SYNTHIA到Cityscapes的跨域语义分割实验,证明了所提出方法的有效性和优越性。

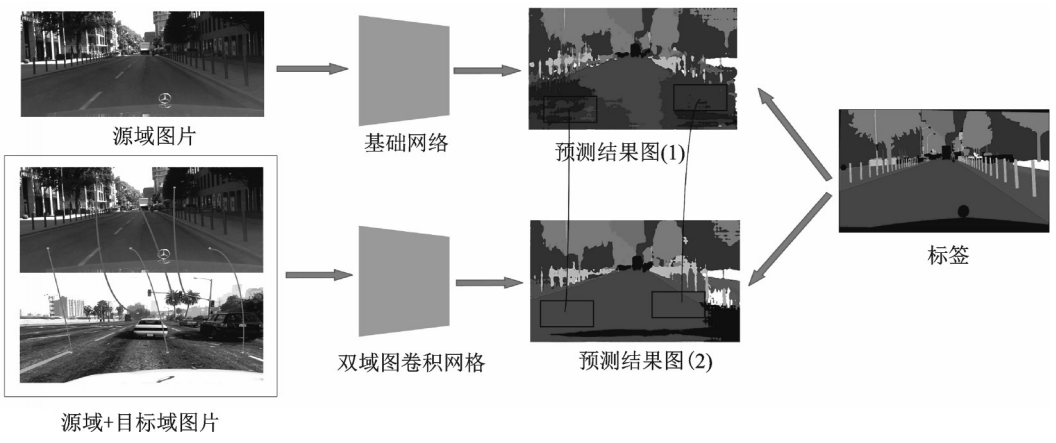

图1 双域图卷积网络效果图

Fig.1 Graph convolutional network rendering of two domains

1 相关工作

1.1 无监督域自适应语义分割

语义分割的基本任务是对图像中的每一个像素进行分类。卷积神经网络在有监督语义分割领域取得了极大的进展[12-16]。在无监督语义分割领域,文献[6,7,17,18]把对抗思想应用到跨域语义分割任务中并且取得了一定的效果。但是基于对抗的方法忽略了语义分割数据集的类不平衡问题,使分割模型对像素数量占比较大的类别过度拟合,造成像素数量占比较小的类别的跨域分割效果较差。自监督方法[19-20]也受到了许多研究者的关注,并应用到了跨域语义分割中。但自监督跨域语义分割的效果主要依赖目标域伪标签预测的准确性。如果伪标签预测错误,会对跨域分割网络的训练造成很大危害。特别是像素数量占较小的类别,伪标签预测的准确率较低,从而会影响跨域分割的效果。另一类是基于风格转换的方法[21-22]。该类方法先利用CycleGAN[23]把源域图像转换到目标域图像风格下,从风格上减少域差异,然后再利用对抗进行跨域图像分割。但这类方法的效果取决于风格转换的质量。如果风格转换的质量不好,那么跨域分割效果就会很差。Vu等[8]利用熵最小化的思想来做跨域语义分割,但是利用熵最小化做损失时,分割网络会对像素数量占比较大的类别过拟合。

1.2 图卷积

近年来,图卷积在计算机视觉任务中取得了很大的进展[24-27]。根据文献[24],图G由节点ν和边ε构成,图卷积被定义为

$$H^{(l+1)}=\sigma(AH^{(l)}W^{(l)}) \tag{1}$$

式中:σ为非线性激活函数;A为图的邻接矩阵;$W^{(l)}$为图卷积第l层的权重;$H^{(l)}$为第l层的图特征;$H^{(l+1)}$是经过图卷积节点更新之后得到的$l+1$层的图特征。图由于其特殊的结构,可以建立起像素之间长距离的上下文依赖关系,并且可以保留像素原有的位置信息。文献[9-11]将图卷积应用在了有监督的语义分割任务中并取得了良好的分割效果。上述工作将语义分割转化为图节点分类问题。基于节点的图卷积使用消息传播在邻居节点之间交换信息,因此可以建立起长距离的上下文关系[11]。同时,由于在节点的信息交换过程中没有节点消失,这样既扩大了感受野又避免了本地位置信息的丢失。在有监督的语义分割中,可以利用标签来指导网络的特征提取和图卷积的节点特征信息更新。但是,在无监督跨域语义分割中,图卷积如何在两个域之间进行信息传播和节点更新是需要解决的问题。

1.3 对比学习

近年来,对比学习在无监督领域取得了一些进展[28-30]。对比学习的核心思想是利用正样本和负样本在特征空间做对比,从而学习到样本的特征表示,其难点在于如何构造正负样本。对于语义分割这种逐像素的分类任务,正样本可以理解为同一类的像素点,负样本可以理解为不同类的像素点。文献[31]把对比学习的思想应用到了无监督跨域语义分割任务中,其利用源域的标签构造掩膜计算每个类的分布,通过对比损失函数[13]来迫使目标域向源域靠拢。但由于源域和目标域之间本身就存在域差异,根据源域的标签求出的类概率分布并不能反映目标域中相同类的概率分布,因此直接根据类分布让目标域向源域靠拢是不合理的,并且其没有对数据集中类的不平衡问题提出解决方案。本文从分类器的角度出发,利用双分类器分别预测源域类别的概率和目标域类别的概率,让目标域中类别的预测概率向源域类别的预测概率靠拢,从而实现跨域语义分割。

2 本文方法

本文有带标签的源域数据集 $\boldsymbol{X}_{\mathrm{S}}=\{(x_i^{(\mathrm{s})},Y_i^{(\mathrm{s})})\}_{i=1}^{n_{\mathrm{s}}}$ 和无标签的目标域数据集 $\boldsymbol{X}_{\mathrm{T}}=\{x_j^{(\mathrm{t})}\}_{j=1}^{n_{\mathrm{t}}}$,其中 n_{s} 和 n_{t} 分别为源域样本的数量和目标域样本的数量。无监督域自适应语义分割就是在不利用目标域标签的情况下训练分割网络,使网络在目标域数据集上测试时能具有较好的性能。

源域数据集和目标域数据集之间具有相同的语义类别以及相似的空间位置关系。利用图卷积可以建立长距离的上下文关系和保留像素之间的位置关系这两个特性[9-11],本文提出了跨域位置相似矩阵和通道相似矩阵的构造方法。利用构造的跨域位置相似矩阵和通道相似矩阵分别进行跨域位置图卷积和跨域通道图卷积,以更新空间上像素点的信息并建立通道之间的相互关系。此外,为了缓解数据的类不平衡的问题,本文提出了分组对比学习方法,根据文献[32]中对城市道路场景数据集中类别的数目统计结果和类别之间的空间位置关系进行分组,组一中的建筑物、道路、植物、天空和人行道这些类别的占比较高,同时这些类别所占的面积较大,类别与类别之间可以通过图卷积操作建立起长距离的上下文关系。同样,组二中的行人、汽车、自行车和卡车等类别都存在于道路上,而且一般存在图像的中间部位,因此根据数量占比和空间位置关系把这些类别分为一组。组三中的交通标志、交通信号灯、栏杆和地台等类别普遍存在于图像的两侧,把这些类别分为一组可以较好地利用类别之间的位置关系。具体地,在每个分组内分别做对比损失,拉近两个域中同类之间的距离,拉远两个域中不同类之间的距离。此外,对不同组的对比损失函数,本文赋予不同的权重来进一步缓解类不平衡问题。本文方法的总体框架如图2所示,主要包括基础网络、图卷积模块和组对比学习模块3部分,浅色箭头表示源域图像在基础网络中的流向,深色箭头表示目标域图像在基础网络中的流向,L_{esg} 和 $L_{\mathrm{contrastive\_gk}}$ 分别为交叉熵损失函数和组对比损失函数。本文利用源域特征图 Z_{S} 和目标域特征图 Z_{T} 构造双域的位置和通道相似性矩阵。图2中左下部分是图卷积模块示意图,利用构造的双域相似性矩阵分别在源域特征图和目标域特征图上做图卷积,输出 $\tilde{Z}_{\mathrm{S}}$ 和 $\tilde{Z}_{\mathrm{T}}$。图2右下部分是所提出的分组对比损失的原理示意图,利用源域分类器 C_{S} 和目标域分类器 C_{T} 分别输出源域类别的概率预测和目标域类别的概率预测,再利用分组对比损失来提取到不同域之间的域不变信息。

2.1 位置图卷积

随着卷积神经网络层数的加深,网络提取到的特征更倾向于高级语义特征[9]。因此,在特征图层面上构造图并做图卷积操作,可以使不同类的特征更加突出。假设源域和目标域的特征图分别表示为 $\mathbf{Z}_{\mathrm{S}}\in\mathbf{R}^{N\times D}$ 和 $\mathbf{Z}_{\mathrm{T}}\in\mathbf{R}^{N\times D}$。其中,$D$ 为特征图的通道数,$N=H\times W$ 为特征图每个通道上的像素点数目。为了捕捉长距离的上下文关系,本文利用 Z_{S} 和 Z_{T} 构造源域和目标域的位置相似矩阵 $\tilde{\boldsymbol{A}}_{\mathrm{S}}^{\mathrm{p}}$ 和 $\tilde{\boldsymbol{A}}_{\mathrm{T}}^{\mathrm{p}}$ 为

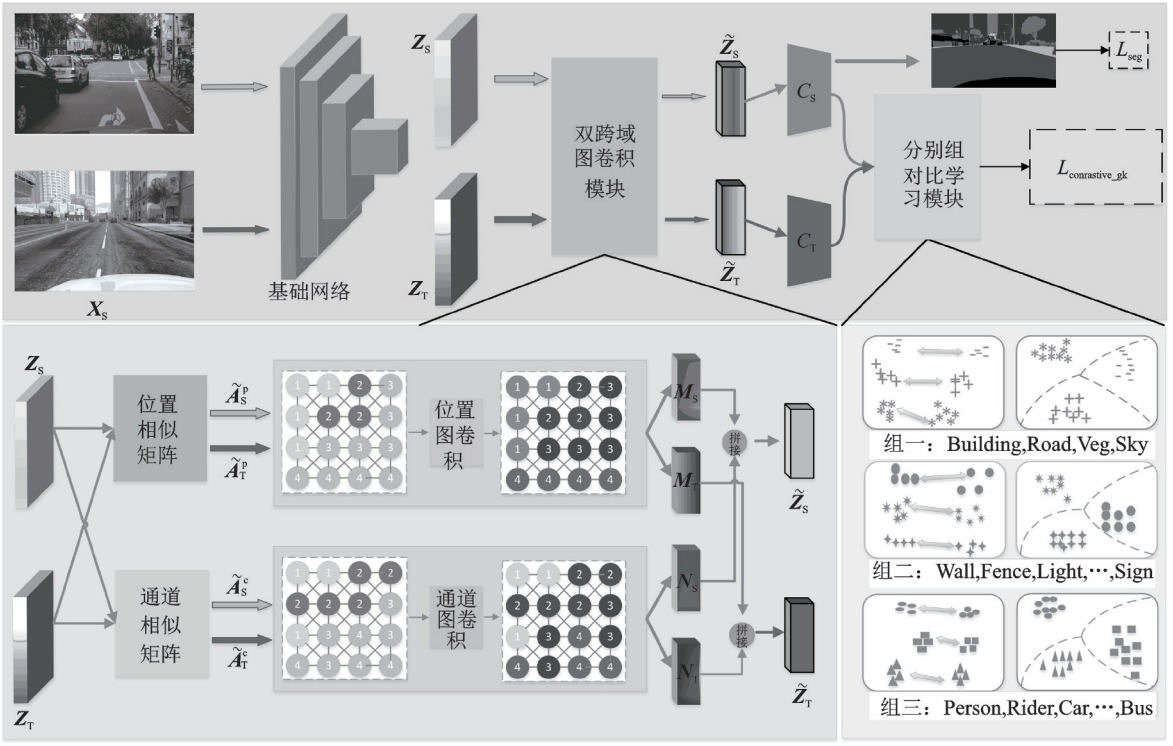

图2 整体网络结构图

Fig.2 Overall network structure diagram

$$\tilde{A}_S^p = \phi(Z_S)\tilde{\Lambda}(Z_S)\phi(Z_T)^T \tag{2}$$

$$\tilde{A}_T^p = \partial(Z_T)\tilde{\Lambda}(Z_T)\partial(Z_S)^T \tag{3}$$

式中：$\phi()$和$\partial()$表示1×1卷积；函数$\tilde{\Lambda}()$用来求输入特征图的对角矩阵，使位置相似矩阵与输入具有关联性，以保留图像的语义信息。计算出跨域位置相似矩阵后，根据式(1)做图卷积操作，即跨域位置图卷积，以更新节点信息建立长距离上下文依赖关系，融合两个域的特征为

$$M_S = \sigma(\tilde{A}_S^p \upsilon(Z_S)W_S) \tag{4}$$

$$M_T = \sigma(\tilde{A}_T^p \upsilon(Z_T)W_T) \tag{5}$$

式中：M_S和M_T分别为源域特征图和目标域特征图经过图卷积之后的输出；W_S和W_T为位置图卷积的权重；$\upsilon()$为1×1卷积。

2.2 通道图卷积

由文献[33-34]可知，通道之间的特征聚合同样对提升语义分割效果有很大的帮助。为了利用特征图中通道之间的关系，本文在通道层面上构造跨域通道相似矩阵。为了减少计算量，将输入的源域特征图$Z_S\in\mathbf{R}^{N\times D}$经过$1\times1$卷积转换成$Z_S'\in\mathbf{R}^{N\times D_1}$，新的特征图有$N$个节点，每个节点的特征向量维度是$D_1$。对于目标域特征图$Z_T\in\mathbf{R}^{N\times D}$经过$1\times1$卷积转换成$Z_T'\in\mathbf{R}^{N\times D_2}$，新的特征图有$N$个节点，每个节点的特征向量维度是$D_2$。与跨域位置相似矩阵的构造原理相似，本文利用$Z_S'$和$Z_T'$来构造跨域的通道相似矩阵为

$$\tilde{A}_S^c = \varphi(Z_S')\tilde{\Lambda}(Z_S')\varphi(Z_T')^T \tag{6}$$

$$\tilde{A}_T^c = \vartheta(Z_T')\tilde{\Lambda}(Z_T')\vartheta(Z_S')^T \tag{7}$$

式中:$\tilde{\boldsymbol{A}}_{\mathrm{S}}^{\mathrm{c}}$和$\tilde{\boldsymbol{A}}_{\mathrm{T}}^{\mathrm{c}}$分别为源域和目标域的通道相似矩阵;$\varphi()$和$\vartheta()$为$1\times1$卷积。计算出跨域通道相似矩阵之后,根据式(1)做图卷积,即跨域通道图卷积,更新节点的特征图,即

$$\boldsymbol{N}_{\mathrm{S}}=\sigma(\tilde{\boldsymbol{A}}_{\mathrm{S}}^{\mathrm{c}}\upsilon(\boldsymbol{Z}_{\mathrm{S}}^{l})\boldsymbol{W}_{\mathrm{S}}^{\mathrm{c}}) \tag{8}$$

$$\boldsymbol{N}_{\mathrm{T}}=\sigma(\tilde{\boldsymbol{A}}_{\mathrm{T}}^{\mathrm{c}}\upsilon(\boldsymbol{Z}_{\mathrm{T}}^{l})\boldsymbol{W}_{\mathrm{T}}^{\mathrm{c}}) \tag{9}$$

式中:$\boldsymbol{N}_{\mathrm{S}}$和$\boldsymbol{N}_{\mathrm{T}}$分别为源域特征图和目标域特征图经过图卷积之后的输出;$\boldsymbol{W}_{\mathrm{S}}^{\mathrm{c}}$和$\boldsymbol{W}_{\mathrm{T}}^{\mathrm{c}}$为通道图卷积的权重。在进行图卷积操作更新节点信息后,每个像素点之间的位置关系都可以得到保留,并且在更新节点的过程中,相同类节点的相似度会越来越大,不相同类节点的相似度会越来越小。$\boldsymbol{M}_{\mathrm{S}}$和$\boldsymbol{N}_{\mathrm{S}}$分别为目标域特征图利用位置图卷积和通道图卷积融合特征之后得到的特征图,$\boldsymbol{M}_{\mathrm{T}}$和$\boldsymbol{N}_{\mathrm{T}}$分别为目标域特征图利用位置图卷积和通道图卷积融合特征之后得到的特征图,在进行分组对比学习之前,需要将特征图经过图卷积更新后的位置特征图和通道特征图融合在一起才能综合利用位置信息和通道信息,因此本文把$\boldsymbol{M}_{\mathrm{S}}$和$\boldsymbol{N}_{\mathrm{S}}$拼接为$\tilde{\boldsymbol{Z}}_{\mathrm{S}}$,把$\boldsymbol{M}_{\mathrm{T}}$和$\boldsymbol{N}_{\mathrm{T}}$拼接为$\tilde{\boldsymbol{Z}}_{\mathrm{T}}$。

2.3 分组对比学习

在经过双跨域图卷积更新节点信息之后,本文采用分组对比学习来使编码器进一步提取域不变特征,同时解决数据集中存在的类不平衡问题。根据类的空间分布位置以及类的占比,把不同的类别分为3组,每组分别有C_1、C_2、C_3个类别,分别在组内用对比损失来约束网络。同时,由于C_2和C_3中的大部分类别数量占比较小,本文赋予组2和组3损失函数更大的权重,以进一步缓解类不平衡的问题。

源域分类器C_{S}和目标域分类器C_{T}的输出分别为源域的概率预测值$P_{\mathrm{S}}^{(i,j)}$和目标域的概率预测值$P_{\mathrm{T}}^{(i,j)}$,本文引入分组对比损失为

$$L_{\mathrm{GCL}}^{k}(C_{\mathrm{S}}(\tilde{\boldsymbol{Z}}_{\mathrm{S}}),C_{\mathrm{T}}(\tilde{\boldsymbol{Z}}_{\mathrm{T}}))=-\sum_{i=1}^{H\times W}\sum_{j=1}^{C_k}P_{\mathrm{T}}^{(i,j)}\ln P_{\mathrm{S}}^{(i,j)}+$$

$$\alpha\sum_{i=1}^{H\times W}\sum_{j=1}^{C_k}\sum_{l=1}^{C_k}P_{\mathrm{T}}^{(i,j)}\ln P_{\mathrm{S}}^{(i,l)}\quad l=j,\alpha=0,k=1,2,3 \tag{10}$$

式(10)第1项的作用是使两个域中相同类的预测输出值更加接近,从而拉近两个域相同类的距离;第2项作用是使两个域中不同类的预测输出值差异更大,从而拉远两个域不同类的距离。α为条件值,当第2项中源域分类器的输出和目标域分类器的输出为同一类时$\alpha=0$,不同类时$\alpha=1$。$C_k(k=1,2,3)$表示第k组中的类别数,i为概率图中的第i个像素点,j为源域C_k中的第j个类,l为目标域C_k中的第l个类。

需要指出的是,经过跨域位置图卷积和通道图卷积得到的特征图$\tilde{\boldsymbol{Z}}_{\mathrm{S}}$和$\tilde{\boldsymbol{Z}}_{\mathrm{T}}$建立了域之间的长距离上下文关系,融合了两个域的特征。通过对源域的预测标签图进行有监督训练,可以训练网络提取到一定程度的域不变特征信息。本文利用源域的标签$\boldsymbol{Y}_{\mathrm{S}}$与网络输出的特征图$G(\boldsymbol{Z}_{\mathrm{S}})$做交叉熵损失,即

$$L_{\mathrm{seg}}(G(\boldsymbol{Z}_{\mathrm{S}}),\boldsymbol{Y}_{\mathrm{S}})=-\sum_{i=1}^{H\times W}\sum_{j=1}^{C}\boldsymbol{Y}_{\mathrm{S}}^{(i,j)}G(\boldsymbol{Z}_{\mathrm{S}})^{(i,j)} \tag{11}$$

式中:G为分割网络;C为类别数目;i为概率图中的第i个像素点;j为C中的第j个类。由于目标域没有标签,本文构造了伪标签$\boldsymbol{Y}_{\mathrm{T}}^{*}$,对目标域的输出预测图$G(\boldsymbol{Z}_{\mathrm{T}})$进行自监督训练,即

$$L_{\mathrm{seg}}(G(\boldsymbol{Z}_{\mathrm{T}}),\boldsymbol{Y}_{\mathrm{T}}^{*})=-\sum_{i=1}^{H\times W}\sum_{j=1}^{c}\boldsymbol{Y}_{\mathrm{T}}^{(i,j)*}G(\boldsymbol{Z}_{\mathrm{T}})^{(i,j)} \tag{12}$$

式中$L_{\mathrm{seg}}(C(\boldsymbol{Z}_{\mathrm{T}}),\boldsymbol{Y}_{\mathrm{T}}^{*})$为目标域利用伪标签的交叉熵损失函数。

本文所提出网络的总损失函数为

$$L_{\mathrm{total}}=L_{\mathrm{seg}}(G(\boldsymbol{Z}_{\mathrm{S}}),\boldsymbol{Y}_{\mathrm{S}})+L_{\mathrm{seg}}(G(\boldsymbol{Z}_{\mathrm{T}}),\boldsymbol{Y}_{\mathrm{T}}^{*})+\sum_{k=1}^{3}\lambda_k L_{\mathrm{GCL}}^{k}(C_{\mathrm{S}}(\tilde{\boldsymbol{Z}}_{\mathrm{S}}),C_{\mathrm{T}}(\tilde{\boldsymbol{Z}}_{\mathrm{T}})) \tag{13}$$

3 实验验证

3.1 实验设置

本文在深度学习框架PyTorch上进行实验，使用NVIDIA 3090 GPU进行训练和测试工作。受限于GPU内存，在训练过程中把源域数据集GTA5中的图像剪裁为720像素×1 280像素，把目标域数据集Cityscapes中的图片剪裁为512像素×1 024像素。为了进行实验对比，本文的Baseline是利用AdaptSegNet方法[6]的超参数，同时采用了IBN(Instance and batch normalization)[35]和自监督方法训练的网络。Baseline网络包含VGG16和ResNet101两种网络架构。具体来说，本文采用SGD (Stochastic gradient descent)优化器[36]来优化分割网络，利用Poly策略[37]来更新学习率。对于式(13)中的超参数λ_1、λ_2和λ_3，由于λ_1项的类别占比较多，在初始值为1的基础上，对λ_1值的调整方向是逐步减小；λ_2和λ_3项的类别占比较小，在初始值为1的基础上，对λ_2和λ_3值的调整方向是逐步增加。经过大量实验发现，当$\lambda_1=0.85$，$\lambda_2=1.35$，$\lambda_3=1.35$时，数据集GTA5到Cityscapes的跨域语义分割结果取得比较好的效果，因此本文中λ_1、λ_2和λ_3的值分别为0.85、1.35和1.35。为了验证所提出方法的有效性，本文做了数据集GTA5到Cityscapes跨域实验，采用指标mIoU评测实验结果。mIoU指标是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。

3.2 定量实验结果

表1展示了对比方法和本文方法在GTA5数据集到Cityscapes数据集19个公共类的跨域实验的定量评价结果。本文分别对比了AdaptSegNet[6]、CLAN[38]、BDL[7]、Advent[8]、MaxSquare[39]、CCM[40]和FADA[41]这7种目前主流的跨域语义分割方法。其中AdaptSegNet和CLAN主要采取对抗损失的方法来对齐两个域的特征分布。BDL利用CycleGAN对两个域的图像进行风格转换，以迫使网络可以提取

表1 数据集GTA5到Cityscapes的跨域语义分割对比实验结果

Table 1 GTA5 to Cityscapes: Comparative experimental results of cross domain semantic segmentation

方法	主干网络	道路	人行道	建筑	墙	围栏	杆子	交通灯	交通标志	绿化带	地	天空	行人	骑手	汽车	货车	公共汽车	火车	摩托车	自行车	mIoU
AdaptSegNet[6]	ResNet101	86.5	36.0	79.9	23.4	23.3	23.9	35.2	14.8	83.4	33.3	75.6	58.5	27.6	73.7	32.5	35.4	3.9	30.1	28.1	42.4
CLAN[38]		87.0	27.1	79.6	27.3	23.3	28.3	35.5	24.2	83.6	27.4	74.2	58.6	28.0	76.2	33.1	36.7	6.7	31.9	31.4	43.2
BDL[7]		91.0	44.7	84.2	34.6	27.6	30.2	36.0	36.0	85.0	**43.6**	83.0	58.6	31.6	83.3	35.3	49.7	3.3	28.8	35.6	48.5
Advent[8]		87.6	21.4	82.0	34.8	26.2	28.5	35.6	23.0	84.5	35.1	76.2	58.6	30.7	84.8	34.2	43.4	0.4	28.4	35.2	44.8
MaxSquare[39]		89.4	43.0	82.1	30.5	21.3	30.3	34.7	24.0	85.3	39.4	78.2	63.0	22.9	84.6	36.4	43.0	5.5	34.7	33.5	46.4
CCM[40]		93.5	**57.6**	84.6	39.3	24.1	25.2	35.0	17.3	85.0	40.6	86.5	58.7	28.7	**85.8**	**49.0**	**56.4**	5.4	31.9	43.2	49.9
FADA[41]		91.0	50.6	86.0	**43.4**	29.8	**36.8**	**43.4**	25.0	**86.8**	38.3	**87.4**	**64.0**	38.0	85.2	31.6	46.1	6.5	25.4	37.1	50.1
本文方法		**94.2**	55.2	**87.3**	39.5	**36.7**	35.1	42.2	**38.9**	85.7	42.3	86.5	63.5	**38.5**	83.3	38.1	44.4	**8.5**	**36.9**	**47.5**	**52.9**
AdaptSegNet[6]	VGG16	87.3	29.8	78.6	21.1	18.2	22.5	21.5	11.0	79.7	29.6	71.3	46.8	6.50	80.1	23.0	26.9	0.00	10.6	0.30	37.6
CLAN[38]		88.0	30.6	79.2	23.4	20.5	26.1	23.0	14.8	81.6	34.5	72.0	45.8	7.90	80.5	**26.6**	29.9	0.00	10.7	0.00	39.3
Advent[8]		86.8	28.5	78.1	27.6	24.2	20.7	19.3	8.90	78.8	29.3	69.0	47.9	5.90	79.8	25.9	34.1	0.00	11.3	0.30	36.6
CBST[42]		66.7	26.8	73.7	14.8	9.50	28.3	25.9	10.1	75.5	15.7	51.6	47.2	6.20	71.9	3.70	2.20	5.40	18.9	**32.4**	39.0
BDL[7]		89.2	40.9	81.2	29.1	19.2	14.2	29.0	19.6	83.7	35.9	80.7	54.7	23.3	82.7	25.8	28.0	2.30	**25.7**	19.9	35.4
FADA[41]		92.3	51.1	**83.7**	33.1	**29.1**	28.5	28.0	**21.0**	82.6	32.6	**85.3**	**55.2**	**28.8**	**83.5**	24.4	**37.4**	0.00	21.1	15.2	43.8
本文方法		**92.4**	**54.3**	83.2	**35.6**	**29.1**	**31.2**	**33.9**	19.1	**84.5**	**40.2**	82.2	**55.2**	24.7	**83.5**	24.1	20.3	**10.2**	21.9	15.2	**47.5**

到与风格无关的域不变的特征。ADVENT利用熵值最小化的思想,通过约束目标域预测图的熵值损失函数来对齐两个域。MaxSquare改进了KL散度损失以及熵损失函数,通过改进的域对齐函数来对齐域的分布。CCM利用两个域的语义一致性来进行跨域。FADA针对类不平衡问题对分类器进行了改进。

从表1可以看出,本文方法对道路、建筑、交通标志、栅栏、火车和自行车等分割结果的IoU达到最高值,对植被、天空等类别的识别率也领先于其他大多数方法,对行人、交通信号和交通灯等小目标的识别率也有所提升,这说明所提出的双域图卷积网络和分组对比学习损失有效。此外,从所有类分割结果的mIoU均值可以看出,本文方法优于对比方法。这说明所提出的方法确实建立起了长距离的上下文依赖关系,不仅对主要类别的像素级分类起到了正向的作用,同时对次要类别以及占比极小类别的分类正确率也有一定的提升。

本文同时也做了SYNTHIA数据集到Cityscapes数据集的跨域实验。值得一提的是,SYNTHIA数据集到Cityscapes数据集之间的域差异比GTA5到Cityscapes数据集之间的差异要大。因此,在SYNTHIA数据集到Cityscapes数据集的跨域语义分割非常有挑战性。表2给出了13个公共类的mIoU值。通过观察可以发现,本文方法在SYNTHIA到Cityscapes的跨域语义分割中同样能得到比较好的结果。在ResNet101作为基础网络时,对比了AdaptSegNet、CLAN、BDL、ADVENT、MaxSquare和FADA这6种方法,其中对于道路、人行道、行人这3个类别本文方法均达到了最高的的准确率。在VGG16作为基础网络时,对比了AdaptSegNet、CLAN、ADVENT、BDL、CBST[42]和FADA这6种方法,其中对于道路、人行道、交通信号灯和植被这4种类别本文方法达到了最高的准确率。同时,ResNet101和VGG16作为基础框架时,本文方法的mIoU值均优于对比方法。

表2 数据集SYNTHIA到Cityscapes的跨域语义分割对比实验结果

Table 2 SYNTHIA to Cityscapes: Comparative experimental results of cross domain semantic segmentation

方法	主干网络	道路	人行道	建筑	交通灯	交通标志	绿化带	天空	行人	骑手	汽车	公共汽车	摩托车	自行车	mIoU
AdaptSegNet[6]	ResNet101	84.3	42.7	77.5	4.70	7.00	77.9	82.5	54.3	21.0	72.3	32.2	18.9	32.3	46.7
CLAN[38]		81.3	37.0	80.1	16.1	13.7	78.2	81.5	53.4	21.2	73.0	32.9	22.6	30.7	47.8
BDL[7]		86.0	46.7	80.3	14.1	11.6	79.2	81.3	54.1	**27.9**	73.7	42.2	25.7	**45.3**	51.4
Advent[8]		85.6	42.2	79.7	5.40	8.10	80.4	**84.1**	57.9	23.8	73.3	36.4	14.2	33.0	48.0
MaxSquare[39]		82.9	40.7	80.0	12.8	18.2	82.5	82.2	53.1	18.0	79.0	31.4	10.4	35.6	48.2
CCM[40]		79.6	36.4	80.6	22.4	14.9	81.8	77.4	56.8	25.9	80.7	45.3	29.9	52.0	52.9
FADA[41]		84.5	40.1	**83.1**	**20.1**	**27.2**	**84.8**	84.0	53.5	22.6	**85.4**	**43.7**	**26.8**	27.8	52.5
本文方法		**92.8**	**53.3**	82.9	19.2	21.4	82.8	81.9	**59.1**	26.6	84.5	38.7	21.1	44.2	**54.5**
AdaptSegNet[6]	VGG16	78.9	29.2	75.5	0.10	4.80	72.6	76.7	43.4	8.80	71.1	16.0	3.60	8.40	37.6
CLAN[38]		80.4	30.7	74.7	1.40	8.00	77.1	79.0	46.5	8.90	**73.8**	18.2	2.20	9.90	39.3
Advent[8]		67.9	29.4	71.9	0.60	2.60	74.9	74.9	35.4	9.60	67.8	21.4	4.10	15.5	36.6
BDL[7]		72.0	30.3	74.5	10.2	**25.2**	80.5	80.0	**54.7**	**23.2**	72.7	24.0	7.50	**44.9**	39.0
CBST[42]		69.6	28.7	69.5	11.9	13.6	82.0	81.9	49.1	14.5	66.0	6.60	3.70	32.4	35.4
FADA[41]		80.4	35.9	**80.9**	7.90	22.3	81.8	**83.6**	48.9	16.8	77.7	**31.1**	**13.5**	17.9	46.0
本文方法		**90.9**	**48.3**	77.2	**15.1**	11.5	**82.1**	81.5	39.1	21.1	72.8	27.0	9.30	41.5	**47.5**

3.3 视觉实验结果

为了从视觉上展示本文方法的有效性，图3展示了GTA5到Cityscapes的跨域语义分割的视觉效果，其中图3(a)为测试图像，图3(b)为只有源域参与训练的网络的分割结果，图3(c～e)为Advent[8]、BDL[7]和CLAN[38]的分割结果，图3(f)为本文方法的分割结果，图3(g)为标签图像。在第1行图像中，由于栅栏像素数量占比较小，很难有效地分割道路两旁出现的栅栏，并且对比方法在道路的边界部分都出现了错误分类的情况，以上情况在本文方法中都得到了有效的改善。在第2行图像中，由于天空和建筑物的界限不明确以及天空占比较小，对比方法的分割结果出现了很严重的错误分类现象，并且在对比方法中人行道也没有很好地识别出来，但是在本文方法的分类结果中，天空、建筑物和人行道都能很好地被分割开，这说明本文方法能很好地解决类不平衡问题，并且人行道分类准确性有所提高，正是由于考虑到不同类别之间的空间位置关系，根据类别数量占比和空间位置关系把类别进行分组，才会在视觉效果中看到了属于组一的天空和建筑相较于之前的方法有了明显的改善。第3行图像中，对比方法对建筑物、天空出现了明显的错误分类情况，并且对人行道的分割效果也比较差，这在本文方法中都得到了明显的改善。从第4行图像中的卡车分类效果可以看出，本文方法对卡车整体分割的比较完善。第5行图像中本文方法对交通信号标志牌也不存在错分类的情况，交通信号标志属于组三，在以往的方法中，组三中的类别数量占比较小，同时由于形状复杂，容易出现错分类的情况。但是在本文方法的结果中，交通标志牌被完整地分割了出来，没有出现错误分类的情况。

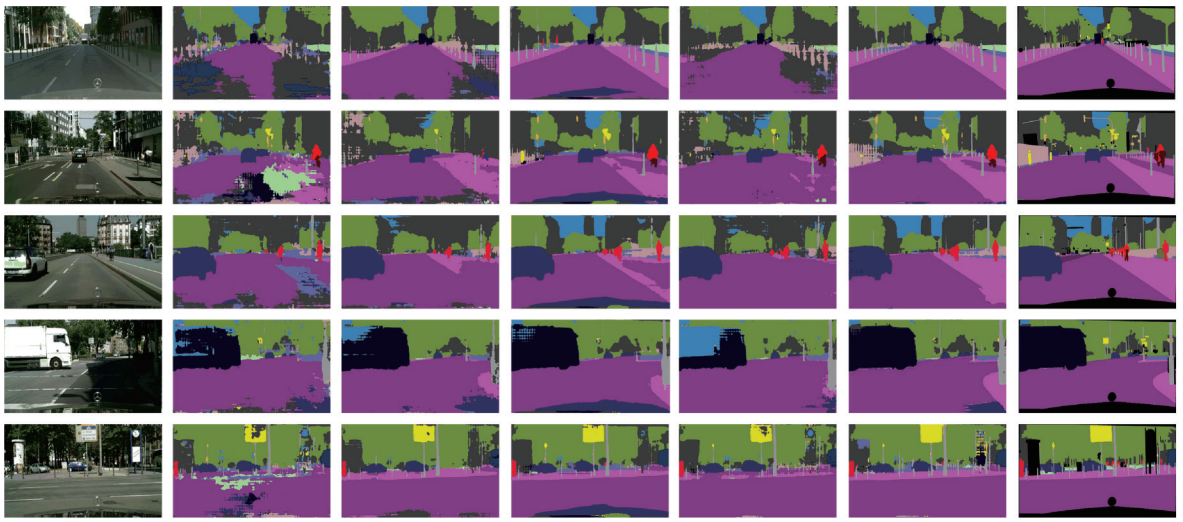

图3 GTA5到Cityscapes实验效果图

Fig.3 Experimental rendering of GTA5 to Cityscapes

3.4 消融实验

本节分别采用了5个网络进行消融实验，以验证本文所提出的跨域位置图卷积、跨域通道图卷积和分组对比学习方法的有效性。图4为数据集GTA5到Cityscapes上消融实验的视觉评价结果，图4(a)为Cityscapes中的测试图片；图4(b)为Baseline的分割结果；图4(c)为Baseline加上分组对比学习(Group contrastive learning, GCL)的分割结果；图4(d)为Baseline加上分组对比学习和跨域位置图卷积(Cross-domain position graph convolution, CPGC)的分割结果；图4(e)为Baseline加上分组对比学习和跨域通道图卷积的分割结果(Cross-domain channel graph convolution, CCGC)的分割结果；图4(f)展

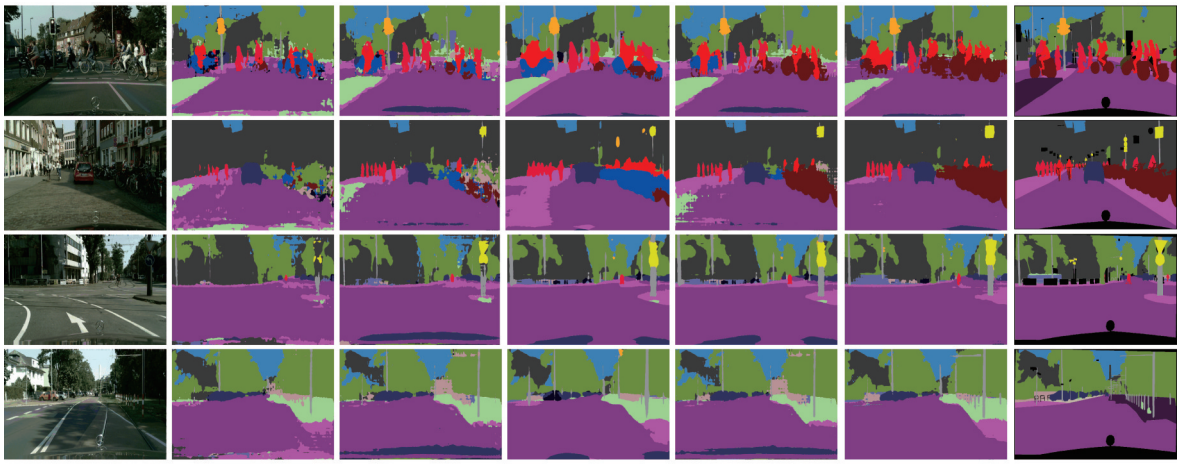

图4 消融实验:GTA5到Cityscapes

Fig.4 Ablation experiment:GTA5 to Cityscapes

示Baseline加上分组对比学习、跨域位置图卷积和跨域通道图卷积的分割结果;图4(g)为标签。

图4中,第1行图像选取了行人过马路的复杂街道场景,在Baseline中像行人、自行车以及骑车的人这些次要类别并没有很好地被分类出来,但是在引入本文方法后,人行道上的次要类别分类准确性有了显著的提升。第2行图像是比较复杂的交通场景图像,可以看出道路两侧有复杂的小物体以及行人等,本文方法可以有效地将小物体进行正确分类,并且相较于Baseline,在引入组对比损失后,分割的效果明显提升,第2列中道路上的错误分类现象有了明显的改善。在引入双域图卷积之后,本文方法不仅识别出了自行车这类小物体,并且识别出了难以识别的行人以及骑自行车的人,通过与最后1列的标签相比,可以发现本文方法有效地提高了网络对于复杂场景道路物体的分割效果。第3行和第4行选取了较为开阔的城市街道场景图像,通过观察第3行和第4行Baseline的分割结果,可以看出图像上部对于天空、建筑以及树木的边缘分割效果并不好,但是随着本文方法的引入,这些类别的分类准确性有所提高,同时由于建立了域间和域内的长距离上下文关系,本文方法对类与类之间的边缘分割更加准确。从第2行和第3行的交通标志分割结果来看,随着本文方法的引入,交通标志的分割准确性也有所提升,这也证明了本文方法确实可以提升次要类别的识别率。从图4可以看出,本文所提出的分组对比学习、跨域位置图卷积和跨域通道图卷积均能对跨域语义分割起到积极的作用。

表3为数据集GTA5到Cityscapes在VGG16和Resnet101作为基础框架时的消融实验的定量评价结果。从表3可以看出当Baseline上分别添加所提出的分组对比学习、跨域位置图卷积和跨域通道图卷积时,分割结果的mIoU值较Baseline均有所提升。VGG16作为基础框架时,本文方法最终的mIoU值相较于Baseline提高了3.9%,Resnet101作为基础框架时,本文方法最终的

表3 数据集GTA5到Cityscapes的消融实验

Table 3 Ablation experiment of GTA5 to Cityscapes

网络	GCL	CPGC	CCGC	mIoU
VGG16 (Baseline)				40.4
	√			41.5
	√	√		42.7
	√		√	42.3
	√	√	√	44.3
Resnet 101 (Baseline)				48.1
	√			49.3
	√	√		51.1
	√		√	50.7
	√	√	√	52.9

mIoU较Baseline提升了4.8%。

4 结束语

为了提升图像跨域语义分割网络的性能,本文构建了双跨域相似矩阵并且利用图卷积来挖掘域内和域间的长距离上下文关系,融合源域和目标域的特征,并通过分组对比学习来解决类不平衡问题。最后,通过大量实验证明了本文方法的有效性,并且在现有的跨域语义分割方法中取得了领先的性能。但是,本文方法也存在一定的局限性,如域之间的位置信息挖掘的不够充分,导致模型在一些类别较复杂的场景仍会出现错误分类的情况。接下来的工作会寻找更加有效的方法以建立类与类之间的位置关系,进一步改善跨域语义分割网络的性能。

参考文献:

[1] RICHTER S R, VINEET V, ROTH S, et al. Playing for data: Ground truth from computer games[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2016: 102-118.

[2] ROS G, SELLART L, MATERZYNSKA J, et al. The SYNTHIA dataset: A large collection of synthetic images for semantic segmentation of urban scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: [s.n.], 2016: 3234-3243.

[3] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[C]//Proceedings of International Conference on Machine Learning. Lille, France: PMLR, 2015: 1180-1189.

[4] HOFFMAN J, TZENG E, PARK T, et al. Cycada: Cycle-consistent adversarial domain adaptation [C]//Proceedings of International Conference on Machine Learning. Stockholm, Sweden: PMLR, 2018: 1989-1998.

[5] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial discriminative domain adaptation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 7167-7176.

[6] TSAI Y H, HUNG W C, SCHULTER S, et al. Learning to adapt structured output space for semantic segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 7472-7481.

[7] LI Y, YUAN L, VASCONCELOS N. Bidirectional learning for domain adaptation of semantic segmentation[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 6936-6945.

[8] VU T H, JAIN H, BUCHER M, et al. Advent: Adversarial entropy minimization for domain adaptation in semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 2517-2526.

[9] ZHANG L, LI X, ARNAB A, et al. Dual graph convolutional network for semantic segmentation[EB/OL]. (2020-08-26) [2022-01-30]. https://arxiv.org/abs/1909.06121v2.

[10] LI X, YANG Y, ZHAO Q, et al. Spatial Pyramid based graph reasoning for semantic segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 8950-8959.

[11] LU Y, CHEN Y, ZHAO D, et al. Graph-FCN for image semantic segmentation[C]//Proceedings of International Symposium on Neural Networks. Cham: Springer, 2019: 97-105.

[12] XU P B, QU A G, WANG K F, et al. A survey of panoptic segmentation methods[J]. Acta Automatica Sinica, 2021, 47(3): 549-568.

[13] HUANG T H, NIE ZHUO Y, WANG Q G, et al. Real-time image semantic segmentationbased on block adaptive feature fusion[J]. Acta Automatica Sinica, 2021, 47(5): 1137-1148.

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

[15] 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, 2017, 40(4): 834-848.

[16] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017: 2881-2890.

[17] HOFFMAN J, WANG D, YU F, et al. FCNs in the wild: Pixel-level adversarial and constraint-based adaptation[EB/OL]. (2016-12-08)[2022-01-30]. https://arxiv.org/abs/1612.02649.

[18] CHEN Y C, LIN Y Y, YANG M H, et al. Crdoco: Pixel-level domain transfer with cross-domain consistency[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 1791-1800.

[19] SAPORTA A, VU T H, CORD M, et al. ESL: Entropy-guided self-supervised learning for domain adaptation in semantic segmentation[EB/OL]. [2022-01-30]. https://arxiv.org/abs/2006.08658.

[20] LEE D H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks[C]//Proceedings of Workshop on Challenges in Representation Learning. [S.l.]: ICML, 2013: 896.

[21] HOFFMAN J, TZENG E, PARK T, et al. Cycada: Cycle-consistent adversarial domain adaptation[C]//Proceedings of International Conference on Machine Learning. [S.l.]: PMLR, 2018: 1989-1998.

[22] WU Z, HAN X, LIN Y L, et al. DCAN: Dual channel-wise alignment networks for unsupervised scene adaptation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: [s.n.], 2018: 518-534.

[23] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Honolulu,USA: IEEE, 2017: 2223-2232.

[24] KIPF T N, WELLING M. Semi-supervised classification with graph convolutional networks[EB/OL]. (2017-02-22)[2022-01-30]. https://arxiv.org/abs/1609.02907.

[25] CHEN Y, ROHRBACH M, YAN Z, et al. Graph-based global reasoning networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 433-442.

[26] LI Y, GUPTA A. Beyond grids: Learning graph representations for visual recognition[J]. Advances in Neural Information Processing Systems, 2018, 31: 9225-9235.

[27] LIANG X, HU Z, ZHANG H, et al. Symbolic graph reasoning meets convolutions[J]. Advances in Neural Information Processing Systems, 2018, 31: 1853-1863.

[28] HADSELL R, CHOPRA S, LECUN Y. Dimensionality reduction by learning an invariant mapping[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006: 1735-1742.

[29] OORD A, LI Y, VINYALS O. Representation learning with contrastive predictive coding[EB/OL]. (2019-01-22)[2022-01-30]. https://arxiv.org/abs/1807.03748v1.

[30] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 9729-9738.

[31] LI S, XIE B, ZANG B, et al. Semantic distribution-aware contrastive adaptation for semantic segmentation[EB/OL]. (2021-05-11)[2022-01-30]. https://arxiv.org/abs/2105.05013.

[32] CHOI S, KIM J T, CHOO J. Cars can't fly up in the sky: Improving urban-scene segmentation via height-driven attention networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020: 9373-9383.

[33] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 3146-3154.

[34] WANG X, GIRSHICK R, GUPTA A, et al. Non-local neural networks[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018: 7794-7803.

[35] PAN X, LUO P, SHI J, et al. Two at once: Enhancing learning and generalization capacities via ibn-net[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: [s.n.], 2018: 464-479.

[36] BOTTOU L. Large-scale machine learning with stochastic gradient descent[C]//Proceedings of COMPSTAT'2010. [S.l.]: Physica-Verlag, 2010: 177-186.

[37] LIU W, RABINOVICH A, BERG A C. Parsenet: Looking wider to see better[EB/OL]. (2015-11-19)[2022-01-30]. https://arxiv.org/abs/1506.04579.

[38] LUO Y, ZHENG L, GUAN T, et al. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019: 2507-2516.

[39] CHEN M, XUE H, CAI D. Domain adaptation for semantic segmentation with maximum squares loss[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Long Beach, USA: IEEE, 2019: 2090-2099.

[40] LI G, KANG G, LIU W, et al. Content-consistent matching for domain adaptive semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Glasgow, UK: [s.n.], 2020: 440-456.

[41] WANG H, SHEN T, ZHANG W, et al. Classes matter: A fine-grained adversarial approach to cross-domain semantic segmentation[C]//Proceedings of the European Conference on Computer Vision (ECCV). Glasgow, UK: [s.n.], 2020: 642-659.

[42] ZOU Y, YU Z, KUMAR B V K, et al. Unsupervised domain adaptation for semantic segmentation via class-balanced self-training[C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: [s.n.], 2018: 289-305.

作者简介:

赵伟枫(1995-),男,硕士研究生,研究方向:图像去雾、语义分割,E-mail:745655411l@qq.com。

谢明鸿(1983-),**通信作者**,男,博士,副教授,研究方向:行人重识别、超分辨率重建、图像融合等,E-mail:minghongxie@163.com。

张亚飞(1986-),女,博士,副教授,研究生导师,研究方向:计算机视觉、机器学习等,E-mail:zyfeimail@163.com。

李华锋(1986-),男,博士,教授,研究方向:行人重识别、计算机视觉、机器学习等,E-mail:hfchina99@163.com。

(编辑:张黄群)