

基于空洞卷积鉴别器的语义分割迁移算法

杨培健¹ 吴晓富¹ 张索非² 周全¹

1 南京邮电大学通信与信息工程学院 南京 210003

2 南京邮电大学物联网学院 南京 210003

(peijiany@163.com)

摘要 近年来,基于卷积神经网络的有监督图像语义分割方法的研究取得了巨大进展。针对该方法所依赖的手动标签繁琐、费时的问题,一种流行的解决方法是通过游戏视频来收集类似于真实场景的图像并自动生成标签,随后利用迁移学习将合成场景训练的模型迁移到真实场景。由于域偏移,简单地将合成场景(源域)上学习的模型应用到真实场景(目标域)一般会出现较高的泛化误差。针对该问题,提出一种新的图像语义分割的无监督迁移算法。该算法首先基于传统的图像风格转换网络对源域图像集进行风格转换预处理,使得图像风格能对齐于目标域,有效降低域间差异;然后,采用生成对抗训练实现源域与目标域特征的对齐。针对现有生成对抗训练中鉴别网络视野受限的问题,提出通过空洞卷积来设计鉴别网络,从而有效提升鉴别网络的分辨能力。在两个典型城市道路数据集 GTA5 以及 SYNTHIA 上的实验表明:相比于经典的 AdaptSegNet 算法,所提算法在 GTA5 数据集上的平均交并比(mIoU)提高了 4.5%,在 SYNTHIA 数据集上的平均交并比提高了 2.6%。

关键词:深度学习;语义分割;迁移学习;域适应;生成对抗网络;空洞卷积

中图分类号 TP391

Semantic Segmentation Transfer Algorithm Based on Atrous Convolution Discriminator

YANG Pei-jian¹, WU Xiao-fu¹, ZHANG Suo-fei² and ZHOU Quan¹

1 School of Telecommunication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

2 School of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract Supervised semantic segmentation with convolutional neural networks has made great progress in recent years. Since the pix-level labeling required by supervised semantic segmentation is tedious and labor intensive, one way that becomes recently prevalent is to collect photo-realistic synthetic data from video games, where pixel-level annotation can be automatically generated. Despite this, the intrinsic domain difference between synthetic and real images usually causes a significant performance drop when applying the learned model to real world scenarios. To solve this problem, we propose a novel domain adaptive semantic segmentation method. It firstly performs image style conversion over the source domain for reducing the domain difference. Then, the generative adversarial network is employed for feature alignment between source and target domains. In particular, we propose to use the atrous convolution for constructing the powerful discriminator network with the enlarged field of view. Extensive experiments show that the proposed algorithm can achieve 4.5% mIoU improvement on the GTA5 dataset and 2.6% on the SYNTHIA dataset, compared with the classic AdaptSegNet algorithm.

Keywords Deep learning, Semantic segmentation, Transfer learning, Domain adaptation, Generative adversarial network, Atrous convolution

1 引言

在大型数据集上训练的神经网络极大地推动了图像语义分割的发展^[1-5],但是,为大型数据集手动注释像素级标签十分耗费人力。经过学者们的不懈努力^[6-7],近来实现了在计算机自动生成标签的合成图像(源域)上训练神经网络。但是,当使用真实图片(目标域)进行测试时,源域与目标域之间的差

异会破坏网络的性能。域适应技术旨在纠正这种不匹配问题,使得在测试时网络性能表现更佳,因此获得了极大的关注。

许多工作希望通过最小化域分布差异来学习域不变表示。Carlucci 等^[8]和 Mancini 等^[9]使用平均值和协方差作为特征分布的统计数据来匹配两个不同的域,然而,当源域和目标域数据的特征分布不是高斯分布时,仅匹配均值和协方差不足以使两个不同的域很好地对齐。对抗学习是另一种常用

收稿日期:2019-11-03 返修日期:2020-06-06 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61372123,61701252)

This work was supported by the National Natural Science Foundation of China (61372123,61701252).

通信作者:吴晓富(xfuwu@njupt.edu.cn)

的域适应方法,通过迫使鉴别器无法鉴别特征是来自源域还是目标域,达到减小域间差异的目的。Hoffman 等^[10]最先将对抗学习的思想引入语义分割,使用全卷积网络(Fully Convolutional Networks, FCN)^[2]作为分割模型,在特征输出空间使用鉴别网络,在两个域之间进行全局和局部对齐。但是,将生成对抗网络用于语义分割这种密集型分类任务时,训练过程将变得非常不稳定。Sankaranarayanan 等^[11]提出将辅助分类损失函数用于生成对抗网络的训练,使得分割网络收敛更加平稳。Tsai 等^[12]在文献^[11]的基础上提出了多级对抗学习方法。该方法将多个鉴别网络分别接在分割网络的不同输出层,对多个级别的特征进行域适应对齐,增强了网络的泛化性能。以上方法均为直接在源域和目标域间对齐特征,部分解决了域偏移问题。但是,合成图片与真实图片在视觉(如外形、光线等)上的差异,导致很难仅从特征层的域适应角度有效减小域间差异。Zhu 等^[13]提出的图像风格转换网络可以将合成图片映射为与真实图片风格类似的形式,有效减小了两个域的图片在视觉上的差异。Hoffman 等^[14]在文献^[13]的基础上引入了语义一致性损失函数,在图像转换网络训练的过程中加入更多的语义信息,使得转换过程更加稳定,减小了转换后图片的内容损失。Wu 等^[15]提出了一种端到端的图

片风格转换网络,并在每个输出通道上对齐特征,保留了更多的空间结构及语义信息。以上方法从像素级域适应角度证明了在训练分割网络之前转换图片风格的有效性。

本文将考虑从特征域适应以及像素域适应两个角度来解决域偏移问题,整体网络结构图如图 1 所示。首先,基于文献^[13]构造像素域适应模块,该模块接受源域图片作为输入,通过迭代训练生成与目标域风格类似的图片,此时源域与目标域图片在光线、亮度等视觉上的域间差异已经减小。进一步地,将转换所得的目标域风格图片作为新的源域图片输入特征域适应模块,特征域适应模块由分割网络与鉴别网络两部分组成。文献^[11-12]均使用 5 个步长等于 2 的卷积层构造鉴别网络,这种连续的下采样层在学习抽象特征表示的同时会丢失空间信息,从而影响分割性能。为解决这种问题,本文设计了一种基于空洞卷积^[1]的鉴别网络,使用多个并行相加的空洞卷积层,在不损失分辨率的同时获取更多的上下文信息。与文献^[12]一致,本文采用 Deeplab-V2^[1]作为分割网络框架。分割网络训练旨在使鉴别网络无法正确鉴别输入图片是来自源域还是目标域,鉴别网络则通过训练来正确鉴别图片来源,通过分割网络与鉴别网络的对抗学习,进一步降低源域与目标域间的特征差异。

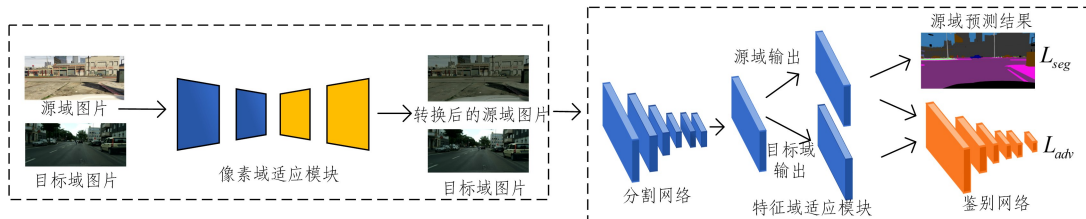


图 1 整体网络结构图

Fig. 1 Overview of our network architecture

2 语义分割迁移算法

本文所提语义分割迁移网络框架包含像素域适应模块和特征域适应模块两部分。像素域适应模块整体基于文献^[13]的结构,特征域适应模块选用文献^[12]的结构,本文重点对核心的鉴别网络重新进行设计。

2.1 像素域适应模块

像素域适应模块采用生成对抗网络结构,具体由转换网络 G 、转换网络 F 和鉴别网络 D 构成。整体结构如图 2 所示,各网络的功能如下。

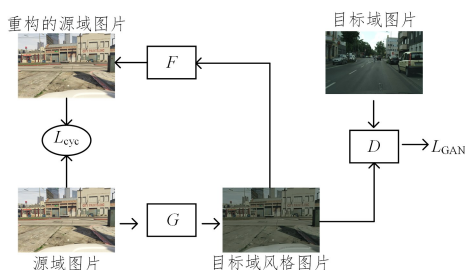


图 2 像素域适应模块的结构

Fig. 2 Structure of pixel domain adaptation module

- 1) 转换网络 G : 将源域图片转换为目标域风格图片。
- 2) 转换网络 F : 将转换网络 G 生成的目标域风格图片

还原为源域风格图片。

3) 鉴别网络 D : 判别是真实的目标域图片还是转换所得的目标域风格图片。

给定源域数据 \mathbf{X}_S 和目标域数据 \mathbf{X}_T , 转换网络 G 的映射表示为 $\mathbf{X}_S \rightarrow \mathbf{X}_T$, 转换网络 F 的映射表示为 $\mathbf{X}_T \rightarrow \mathbf{X}_S$ 。首先,使用转换网络 G 将源域图片转换为目标域风格图片,转换后的图片表示为 $\{G(x_s)\}$ 。鉴别网络 D 的目标在于鉴别输入图片是来自原始的目标域图片 $\{x_t\}$, 还是经过转换网络转换所得到的目标域风格图片 $\{G(x_s)\}$, 由此可以构造出生成对抗网络。对抗损失函数的定义如下:

$$L_{GAN}(\mathbf{G}, \mathbf{D}, \mathbf{X}_S, \mathbf{X}_T) = E_{x_t \sim \mathbf{X}_T} [\log \mathbf{D}(x_t)] + E_{x_s \sim \mathbf{X}_S} [\log(1 - \mathbf{D}(G(x_s)))] \quad (1)$$

转换网络 G 的目标是最小化上述损失函数以获得好的转换图片。判别网络 F 的目标是最大化上述损失函数,以正确判别结果。由此,优化方程可表示为:

$$\min_G \max_D L_{GAN}(\mathbf{G}, \mathbf{D}, \mathbf{X}_T, \mathbf{X}_S) \quad (2)$$

优化方程(2),将源域图片转为目标域风格图片。但在实际训练中,很容易将源域图片映射为目标域风格中的某一张图片。因此,需要将转换后的目标域风格图片通过转换网络 F 尽可能还原出源域图片,即 $x_s \rightarrow G(x_s) \rightarrow F(G(x_s)) \approx x_s$ 。

循环一致性损失函数的定义如下:

$$L_{cyc}(\mathbf{G}, \mathbf{F}, \mathbf{X}_S) = E_{x_s \sim X_S} [\| \mathbf{F}(\mathbf{G}(x_s)) - x_s \|_1] \quad (3)$$

其中, $\mathbf{F}(\mathbf{G}(x_s))$ 表示转换网络输出的源域风格图片。优化式(3), 可使源域风格图片逼近源域图片。

综上, 可以得到像素域适应模块的目标函数为:

$$L(\mathbf{X}_S, \mathbf{X}_T, \mathbf{G}, \mathbf{F}, \mathbf{D}) = L_{GAN}(\mathbf{G}, \mathbf{D}, \mathbf{X}_S, \mathbf{X}_T) + \lambda L_{cyc}(\mathbf{G}, \mathbf{F}, \mathbf{X}_S, \mathbf{X}_T) \quad (4)$$

其中, 参数 λ 用于调整对抗损失函数与循环一致性损失函数的比例。

像素域适应模块的优化方程表示为:

$$\mathbf{G}^*, \mathbf{F}^*, \mathbf{D}^* = \min_{\mathbf{G}, \mathbf{F}} \max_{\mathbf{D}} L(\mathbf{X}_S, \mathbf{X}_T, \mathbf{G}, \mathbf{F}, \mathbf{D}) \quad (5)$$

分别使用 GTA5^[6] 和 SYNTHIA^[5] 作为源域数据集, 使用 Cityscapes^[16] 作为目标域数据集, 进行转换操作, 转换结果如图 3 所示。



图 3 像素域适应模块的转换结果

Fig. 3 Conversion result of pixel domain adaptation modul

2.2 特征域适应模块

特征域适应模块由分割网络 G 和鉴别网络 D 组成。本文采用在 ImageNet^[17] 数据集预训练的 Deeplab-V2 作为分割

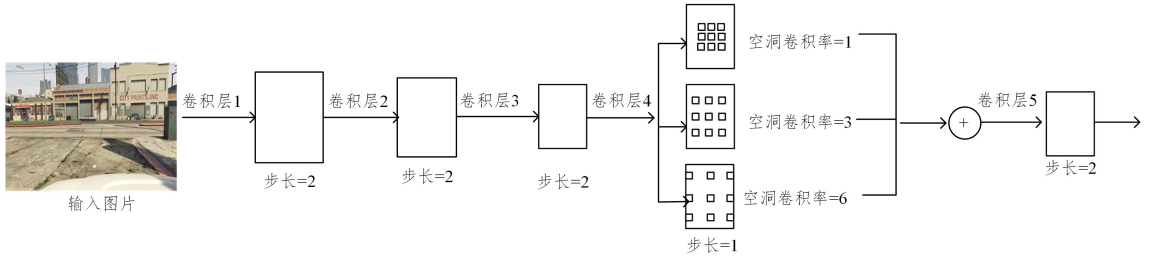


图 4 基于空洞卷积的鉴别网络结构

Fig. 4 Discriminator network structure based on atrous convolution

表 1 堆叠与相加方法参数及性能的对比

Table 1 Comparison of parameters and performance between plus and concat

方法	参数/M	mIoU
堆叠	20.7	42.6
相加	16.12	43.8

鉴别网络的损失函数定义如下:

$$L_d(P) = -\frac{1}{h\omega} \sum (1-z) \log(\mathbf{D}(P)^{hw}) + z \log(\mathbf{D}(P)^{hw}) \quad (10)$$

其中, P 表示分割网络的输出, 当鉴别网络的输入来自源域时 $z=0$, 当鉴别网络的输入来自目标域时 $z=1$ 。鉴别网络的训练目标表示如下:

网络框架。对于转换网络 G 生成的目标域风格图片, 重新将其表示为 $\{I_s\}$, 并将 $\{I_s\}$ 作为新的源域数据输入分割网络。分割网络的损失函数定义如下:

$$L_{seg}(I_s) = -\frac{1}{h\omega} \sum_{h,w} \sum_{c \in C} Y_s^{hw} \log(P_s^{hw}) \quad (6)$$

其中, Y_s 为源域数据标签; $P_s = \mathbf{G}(I_s)$ 表示分割网络的输出值; h, ω 和 c 分别表示输入图片的高、宽和通道数。

将目标域图片 I_t 输入分割网络, 输出结果表示为 $P_t = \mathbf{G}(I_t)$ 。为了使 P_t 分布接近于 P_s , 定义对抗损失函数为:

$$L_{adv}(I_t) = -\frac{1}{h\omega} \sum \log(\mathbf{D}(P_t)^{hw}) \quad (7)$$

用对抗损失函数来训练分割网络, 以最大化鉴别网络鉴别出错的概率。构造分割的联合损失函数, 定义如下:

$$L(I_s, I_t) = L_{seg}(I_s) + \lambda_{adv} L_{adv}(I_t) \quad (8)$$

参数 λ_{adv} 用于调整两个损失函数之间的比例。分割网络的训练目标表示如下:

$$\min_G L(I_s, I_t) \quad (9)$$

本文设计了一种基于空洞卷积的鉴别网络, 其结构如图 4 所示。前 3 层为步长为 2 的普通卷积层。第 4 层由 3 个空洞卷积率分别为 1, 3, 6、步长为 1 的空洞卷积并行组成, 该层的 3 个空洞卷积同时接受卷积层 3 的输出特征作为输入, 并将各自的输出特征相加, 在不降低输入分辨率的同时获取了更多的空间信息。如表 1 所列, 与传统 ASPP 结构^[1] 堆叠的方式相比, 相加操作减少了网络参数量, 更利于生成对抗网络的收敛。输出卷积层 5 为步长为 2 的普通卷积, 接受卷积层 4 的相加特征作为输入, 输出特征用于分类输入图片是来自源域还是目标域。

$$\min_D L_d(P) \quad (11)$$

即通过更新鉴别网络参数来最小化鉴别网络损失函数, 达到正确判别结果的目的。

3 实验

将合成数据集 GTA5 和 SYNTHIA 作为源域数据训练模型, 并在真实数据集 Cityscapes 上进行测试。与文献^[12, 14-15] 相同, 使用 $mIoU$ 来评估本文方法的性能。

3.1 数据集介绍

GTA5 数据集包含 24966 张像素为 1914×1052 的图片。GTA5 数据集与 Cityscapes 数据集有 19 个共同类别, 与文献^[12, 14-15] 一致, 使用这 19 个共同类别进行训练, 其他类别

归为背景类。对于 SYNTHIA 数据集,本文将使用其子集 SYNTHIA-RAND-CITYSCAPES^[6] 作为训练集,其中包含 9400 张像素为 1280×760 的图片,与文献[11]一致,将使用 16 个共同类别进行训练。真实数据集 Cityscapes 被划分为训练集、验证集和测试集,实验中将使用训练集作为目标域数据进行训练,其中包含 2975 张像素为 2048×1024 的图片,并使用 500 张验证集图片进行测试。

3.2 实现方法

在训练像素域适应模块时,输入图片将被随机裁剪至 440×440 的大小,并训练 15 个轮,式(4)中的固定为 20。在训练特征域适应模块时,对于分割网络,使用 SGD 作为优化器,初始学习率设置为 2.5×10^{-4} ;对于判别网络,使用 Adam 作为优化器,初始学习率设置为 1×10^{-4} ,动量为 0.9 和

0.99。式(8)中的 λ_{adv} 设置为 0.001。

3.3 性能对比

本节将本文方法与其他文献方法进行对比,主要比较两项任务:1) GTA5 迁移到 Cityscapes;2) SYNTHIA 迁移到 Cityscapes。表 2 展示了 GTA5 到 Cityscapes 的域适应结果。其中,文献[12]主要运用不同的对抗损失函数来从特征级降低域间差异;文献[13]从像素域适应的角度转换图片风格,减小域偏移带来的影响。通过对比可以看出本文方法的优越性。图 5 展示了在 GTA5 数据集上训练的模型的分割结果,并将本文方法与文献[12]的结果进行对比。表 3 展示了 SYNTHIA 到 Cityscapes 的域适应结果,公平起见,本文方法结果展示与文献[11-12,15]一致,表中仅列出了 13 个类别的测试结果。

表 2 从 GTA5 迁移到 Cityscapes 时不同算法的 IoU 性能比较

Table 2 IoU performance comparison of different algorithms when migrating from GTA5 to Cityscapes

method	road	sidewalk	building	wall	fence	pole	light	sign	veg	terrain	sky	person	rider	car	truck	bus	train	mbike	bike	<i>mIoU</i>
AdaptSegNet ^[12]	86.3	22.1	80.4	21.6	24.6	25.2	33.6	15.3	82.6	30.5	76.2	57.1	27.6	77.5	24.7	36.2	1.5	29.6	21.2	40.7
DCAN ^[15]	85.0	30.8	81.3	25.8	21.2	22.2	25.4	26.6	83.4	36.7	76.2	58.9	24.9	80.7	29.5	42.9	2.5	26.9	11.6	41.7
Cycada ^[14]	86.7	35.6	80.1	19.8	17.5	38.0	39.9	41.5	82.7	27.9	73.6	64.9	19	65.0	12.0	28.6	4.5	31.1	42.0	42.7
Ours	89.2	42.0	84.1	27.9	23.6	30.2	35.0	28.4	83.3	36.6	84.9	59.0	29.7	83.0	31.1	43.3	0.0	25.0	21.6	45.2

注:加粗字体为每列最优值,AdaptSegNet^[12]结果为根据作者提供源码实验所得

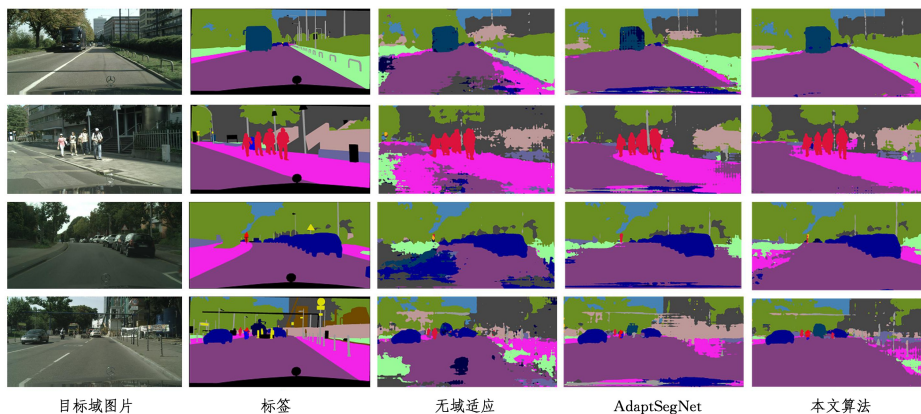


图 5 Cityscapes 测试集的可视化效果图

Fig. 5 Visualization of Cityscapes test set

表 3 从 SYNTHIA 迁移到 Cityscapes 时不同算法的 IoU 性能比较

Table 3 IoU performance comparison of different algorithms when migrating from SYNTHIA to Cityscapes

method	road	sidewalk	building	light	sign	Veg	sky	person	rider	car	bus	mbike	bike	<i>mIoU</i>
LSD ^[11]	80.1	29.1	77.5	11.1	18.0	78.1	76.7	48.2	15.2	70.5	17.4	8.7	16.7	40.9
DCAN ^[15]	81.5	33.4	72.4	8.6	10.5	71.0	68.7	51.5	18.7	75.3	22.7	12.8	28.1	42.7
AdaptSegNet ^[12]	75.8	32.6	76.0	12.1	10.6	76.7	79.3	52.1	18.7	65.5	21.6	20.8	28.3	43.9
Ours	82.4	34.5	77.3	11.7	15.4	78.1	80.6	49.8	19.9	70.2	30.8	16.3	37.1	46.5

注:加粗字体为每列最优值,AdaptSegNet^[12]结果为根据作者提供源码实验所得

3.4 实验结果分析

本文模型中各方法对总体性能的影响如表 4 所列。在将 resnet^[18] 作为基本框架的基础上,使用对抗域适应方法进行域对齐,并将所得实验结果作为基准。在从 GTA5 迁移到 Cityscapes 的任务中,所提空洞卷积鉴别网络(Atrous Convolutional Discriminator, ACD)、像素域适应模块(Pixel Domain Adaptation, PDA)分别带来了 3.1% 和 1.4% 的性能提升,证明了本文方法的有效性。

表 4 模型中各方法对总体性能的影响

Table 4 Performance contribution of each design in our framework

Method	ACD	PDA	<i>mIoU</i>
Baseline(resnet)	—	—	40.7
+ ACD	Y	—	43.8
+ PDA	Y	Y	45.2

注:“—”表示未使用该项方法,Y 表示使用了该项方法

结束语 本文介绍了一个基于空洞卷积鉴别器的语义分割迁移算法模型,提出从特征域适应以及像素域适应两个角

度考虑解决域偏移问题。此外,在特征域适应模块中,提出使用空洞卷积设计鉴别网络,在不损失像素分辨率的同时获取多级别上下文信息,促进了分割网络与鉴别网络的收敛。在两个合成数据集 GTA5 以及 SYNTHIA 上进行的验证证实了本文方法的有效性。所提方法在 GTA5 数据集上的 *mIoU* 值达到了 45.2%,在 SYNTHIA 数据集上的 *mIoU* 达到了 46.5%。然而,本文算法仍然存在需要进一步改进的地方,如测试时模型对远处物体和小物体的分割结果不太理想,这可能是由于下采样过程造成了位置信息的丢失。接下来可以进一步优化网络结构,提升模型对小物体及远处物体的分割能力。

参 考 文 献

- [1] CHEN L C, PAPANDEOU 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.
- [2] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015:3431-3440.
- [3] LIU Z, LI X, LUO P, et al. Semantic image segmentation via deep parsing network[C] // *Proceedings of the IEEE International Conference on Computer Vision*. 2015:1377-1385.
- [4] LIN G, SHEN C, VAN DEN HENGEL A, et al. Efficient piecewise training of deep structured models for semantic segmentation[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:3194-3203.
- [5] ZHENG S, JAYASUMANA S, ROMERA-PAREDES B, et al. Conditional random fields as recurrent neural networks[C] // *Proceedings of the IEEE International Conference on Computer Vision*. 2015:1529-1537.
- [6] 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 Conference on Computer Vision and Pattern Recognition*. 2016:3234-3243.
- [7] RICHTER S R, VINEET V, ROTH S, et al. Playing for data: Ground truth from computer games[C] // *European Conference on Computer Vision*. Cham: Springer, 2016:102-118.
- [8] CARIUCCI F M, PORZI L, CAPUTO B, et al. Autodial: Automatic domain alignment layers[C] // *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, 2017: 5077-5085.
- [9] MANCINI M, PORZI L, ROTA BULÒ S, et al. Boosting domain adaptation by discovering latent domains [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018:3771-3780.
- [10] HOFFMAN J, WANG D, YU F, et al. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation [J]. *arXiv: 1612.02649*, 2016.
- [11] SANKARANARAYANAN S, BALAJI Y, JAIN A, et al. Learning from synthetic data: Addressing domain shift for semantic segmentation[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018:3752-3761.
- [12] TSAI Y H, HUNG W C, SCHULTER S, et al. Learning to adapt structured output space for semantic segmentation[C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018:7472-7481.
- [13] 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*. 2017:2223-2232.
- [14] HOFFMAN J, TZENGE E, PARK T, et al. Cycada: Cycle-consistent adversarial domain adaptation [J]. *arXiv: 1711.03213*, 2017.
- [15] 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)*. 2018:518-534.
- [16] 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*. 2016:3213-3223.
- [17] DENG J, DONG W, SOCHER R, et al. Imagenet: A large-scale hierarchical image database [C] // *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009: 248-255.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:770-778.



YANG Pei-jian, born in 1995, postgraduate. His main research interests include semantic segmentation and transfer learning.



WU Xiao-fu, born in 1975, Ph.D, professor. His main research interests include computer vision, face recognition and transfer learning.