

# 基于编码器-解码器的半监督图像语义分割<sup>①</sup>



刘贝贝, 华 蓓

(中国科学技术大学 计算机科学与技术学院, 合肥 230027)  
通讯作者: 华 蓓, E-mail: bhua@ustc.edu.cn

**摘 要:** 基于深度卷积神经网络的图像语义分割方法需要大量像素级标注的训练数据, 但标注的过程费时又费力. 本文基于生成对抗网络提出一种编码-解码结构的半监督图像语义分割方法, 其中编码器-解码器模块作为生成器, 整个网络通过耦合标准多分类交叉熵损失和对抗损失进行训练. 为充分利用浅层网络包含的丰富的语义信息, 本文将编码器中不同尺度的特征输入到分类器, 并将得到的不同粒度的分类结果融合, 进而优化目标边界. 此外, 鉴别器通过发现无标签数据分割结果中的可信区域, 以此提供额外的监督信号, 来实现半监督学习. 在 PASCAL VOC 2012 和 Cityscapes 上的实验表明, 本文提出的方法优于现有的半监督图像语义分割方法.

**关键词:** 图像语义分割; 编码器-解码器; 深度学习; 生成对抗网络; 半监督学习

引用格式: 刘贝贝, 华蓓. 基于编码器-解码器的半监督图像语义分割. 计算机系统应用, 2019, 28(11): 182-187. <http://www.c-s-a.org.cn/1003-3254/7159.html>

## Encoder-Decoder for Semi-Supervised Image Semantic Segmentation

LIU Bei-Bei, HUA Bei

(School of Computer Science and Technology, University of Science and Technology of China, Hefei 230027, China)

**Abstract:** Image semantic segmentation methods based on deep convolutional neural network requires a large number of pixel-level annotation training data, but the labeling process is time-consuming and laborious. In this study, a semi-supervised image semantic segmentation method with encoder-decoder based on generative adversarial networks is proposed, in which the encoder-decoder as the generator. The entire network is trained by coupling the standard multi-class cross entropy loss with the adversarial loss. In order to make full use of the rich semantic information contained in the shallow layers, this study puts the features of multi-scales in the encoder into the classifier, and fuses the obtained classification results with different granularities to optimize the object boundaries. In addition, the discriminator enables semi-supervised learning by discovering the trusted regions in the unlabeled data segmentation results to provide additional supervisory signals. Experiments on PASCAL VOC 2012 and Cityscapes show that the proposed method is superior to the existing semi-supervised image semantic segmentation methods.

**Key words:** image semantic segmentation; encoder-decoder; deep learning; generative adversarial networks; semi-supervised learning

### 1 概述

图像语义分割是为图像中的每一个像素点指定一个语义类别, 是一种像素级的空间密集型预测任务, 被广泛应用到无人驾驶技术、无人机领域和医疗图像检

测与分析等<sup>[1]</sup>任务中. 深度卷积神经网络 (Convolutional Neural Network, CNN) 的出现, 尤其是全卷积神经网络的出现 (Fully Convolutional Network, FCN)<sup>[2]</sup>, 极大地推动了图像语义分割的发展, 目前大多数先进

<sup>①</sup> 收稿时间: 2019-04-17; 修改时间: 2019-05-21; 采用时间: 2019-05-23; csa 在线出版时间: 2019-11-06

的图像语义分割方法都是在 FCN 的基础上进行调整, 通过使用更深的网络、空洞卷积<sup>[3]</sup>、空洞空间金字塔池化模块 (Atrous Spatial Pyramid Pooling, ASPP)<sup>[4]</sup>等技术不断刷新分割精度。

虽然基于 CNN 的图像语义分割方法已经取得了惊人的性能, 但训练这些方法需要大量的有标签数据。不同于图像分类、目标检测任务, 图像语义分割需要的是像素级的标注, 训练数据的获取费时又费力。为缓和这一问题, 半监督/弱监督方法已应用到图像语义分割任务中。在弱监督图像语义分割方法中, 分割网络并不使用具有像素级标注的数据进行训练, 而是用的更容易获取的各种弱监督信号, 例如图像类别<sup>[5]</sup>、物体边界框等<sup>[6]</sup>。文献<sup>[5]</sup>提出, 基于卷积神经网络将多实例学习框架从检测任务扩展到图像语义分割任务, 此时图像语义分割可以被定义为给定一张图像和其对象类别标签, 推断图像中的每一个像素是否属于对象类, 并约束模型为分类正确的像素分配更高的权重。BoxSup<sup>[6]</sup>则是使用物体边界框作为弱监督信号, 其基本思想是, 首先使用区域建议方法生成候选分割掩模, 分割网络在这些候选掩模的监督下进行训练, 更新的网络反过来改进了用于训练的候选分割掩模, 整个过程是迭代的。虽然分割掩模在刚开始的时候是粗糙的, 但它们随着训练的进行逐渐得到改进, 然后为训练分割网络提供有用的信息。基于半监督学习的图像语义分割方法大多综合使用少量的像素级标注的训练图像和大量的弱标注图像。例如文献<sup>[7]</sup>构建了一个由分类网络和分割网络组成的框架, 其中分类网络作为整个框架的特征提取器, 首先用大量的具有图像类别信息的数据进行训练, 然后在分类网络固定的情况下, 使用少量具有像素级标注的图像优化分割网络。然而, 以上这些方法并没有取得很好的性能, 因为它们无法从弱监督信号或少量像素级标注图像中恢复详细的目标边界。

生成对抗网络 (Generative Adversarial Networks, GAN)<sup>[8]</sup>自被提出以来就受到了各界的广泛关注, 并被尝试应用到很多领域, 例如图像到图像生成<sup>[9]</sup>、图像语义分割等<sup>[10-12]</sup>, 并取得了很好的性能提升。典型的 GAN 模型包括生成器、鉴别器两部分, 这两部分通过对抗的方式进行训练。在训练过程中, 生成器的目标是将采样自某固定分布的样本变换成近似服从目标数据分布的样本, 而鉴别器的目标是准确判别出样本来自生成数据还是目标数据。通过反向传播算法, 生成器、鉴别

器的性能都得以提升, 最终达到鉴别器无法准确区分数据来源的状态, 此时可以认为生成器生成的数据已近似服从目标数据分布。文献<sup>[10]</sup>是第一个将生成对抗网络应用于图像语义分割的工作, 该方法为后来的方法奠定了基本框架, 即将主流语义分割网络作为生成对抗网络的生成器。通过对抗训练, 该方法可以在不需要增加复杂的后期处理的情况下, 鼓励分割网络生成更加接近目标数据的样本。然而, 它并不是适用于所有数据集的通用方法, 相比于基准方法也没有明显的性能提升。文献<sup>[12]</sup>从另一角度思考问题, 提出一种半监督图像语义分割方法, 并创造性的提出了全卷积鉴别器。该算法将鉴别器视为分割网络, 并用生成器生成的样本扩展数据集。然而, 该方法并没有取得很好的效果, 因为生成的样本不足够接近真实样本, 不能帮助训练分割网络。文献<sup>[11]</sup>是目前性能最优的半监督语义分割方法, 该方法沿用了文献<sup>[10]</sup>中提出的基本框架, 即分割网络作为生成器, 同时采用了全卷积方式的鉴别器, 该鉴别器不仅能够为输入数据提供像素级的鉴别 (鉴别每一个像素点采样自生成图像还是真实标签), 还能进一步为无标签数据生成置信图, 并将置信图作为额外的监督信号, 来实现半监督学习。同时, 文献<sup>[11]</sup>也存在着一些问题, 其中最为显著的就是分割网络过于简单, 直接通过一次上采样操作将特征图放大到输入图像大小, 在分割过程没有充分利用浅层网络中包含的丰富的图像细节信息, 但是, 将不同尺度的特征信息相融合有助于进一步提升网络分割性能<sup>[4]</sup>。

基于以上介绍和分析, 本文在文献<sup>[11]</sup>的基础上, 提出一种基于生成对抗网络的半监督图像语义分割方法, 其中编码器-解码器作为生成对抗网络生成器, 鉴别器依然采用全卷积的方式。不同于文献<sup>[11]</sup>使用的直接上采样, 本文提出的解码器模块在逐步上采样逐渐增大特征分辨率的同时, 通过整合来自浅层精细的外观信息恢复目标分割细节。此外, 全卷积鉴别器通过发现无标签数据分割结果中的可信区域, 并将可信区域作为额外的监督信号来实现半监督学习。实验结果显示, 相较于其他半监督图像语义分割方法, 本文提出的算法有明显的分割性能提升。

## 2 网络结构与训练过程

本文提出的网络结构包括分割网络 (生成器)、鉴别网络两部分, 如图 1 所示。分割网络将大小为  $H \times W \times 3$

的 RGB 图像作为输入, 输出大小为  $H \times W \times C$  的分割图, 其中  $C$  为语义类别. 鉴别网络将生成的分割结果或图像真实标签作为输入, 输出大小为  $H \times W \times 1$  的空间概率图, 也称置信图.

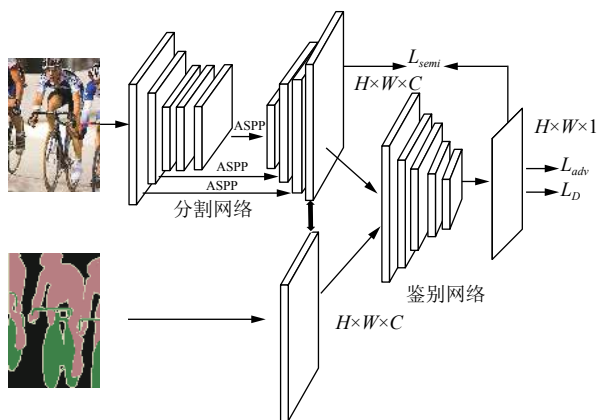


图1 模型整体框架

### 2.1 多尺度分类器 ASPP

ASPP 是 DeepLab-v2 提出的用于捕获目标不同尺度上下文信息的重要模块, 如图 2 所示. ASPP 中存在 4 个并行的空洞卷积操作, 其卷积核大小都为  $3 \times 3$ , 步长也都为 1, 每个空洞卷积的膨胀率  $d$  分别为 6, 12, 18, 24, 相应的输入特征填充率  $p$  分别为 6, 12, 18, 24, 由式 (1) 可知.

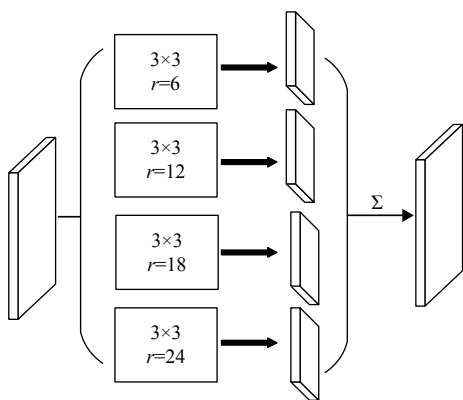


图2 分类器 ASPP 示意图

$$W_{out} = \frac{W_{in} + 2p - [k + (k - 1) \times (d - 1)]}{s} + 1 \quad (1)$$

这种参数设置的空洞卷积具有相同大小的输出, 且与输入特征的分辨率也相同, 所以经 ASPP 处理并不会改变特征分辨率. 在本文中, 我们对原始 ASPP 做了如下简化工作, 去除了不同采样率的空间卷积之后

的  $1 \times 1$  卷积, 取而代之的是, 在进行不同采样率的空间卷积时将通道数设置为语义类别数, 并将得到的不同尺度的特征相加, 作为最终分类结果. 最后, 经双线性插值上采样和沿通道方向的 Softmax 操作获取最后的分割输出.

### 2.2 基于编码器-解码器的分割网络

编码器-解码器模块由 2 部分组成: 编码器和解码器. 其中编码器逐渐减小特征图大小并捕获更高层次的语义信息; 解码器逐渐恢复对象细节信息和空间维度. 整个结构利用来自编码器模块的多尺度特征, 并从解码器模块恢复空间分辨率.

基于生成对抗网络的语义分割方法中分割网络可以是任意主流图像语义分割模型, 例如 FCN、DeepLab-v2 等, 本文采用的是基于 ResNet-101<sup>[13]</sup> 的 DeepLab-v2 模型, 并将其改造成编码-解码结构, 如图 3 所示. 编码器模块经 ResNet-101 中一个卷积、一个最大池化和 4 个残差块处理后, 逐渐降低特征分辨率为输入图像的  $1/8$  大小, 同时, 不同深度的卷积层编码了不同尺度的特征信息. 解码器模块首先将第四个残差块输出的特征图经多尺度分类器 ASPP 处理, 然后进行 2 倍的双线性上采样变为输入图像的  $1/4$ , 同时将编码器中输入图像  $1/4$  大小的特征图, 也即为第一个残差块的输出, 同样经过分类器 ASPP 处理, 将两者相加进行第二次 2 倍的双线性上采样, 此时输出的特征图为输入图像的  $1/2$ . 同样地, 将编码器中输入图像  $1/2$  大小的特征图, 也即为 ResNet-101 中第一个卷积层的输出, 经过分类器 ASPP 处理, 并于之前求得的输入图像  $1/2$  大小的特征图相加, 最后再经过一次 2 倍的双线性上采样即可恢复至输入图像大小. 在整个解码过程中, 始终保持特征图的通道数为图像语义类别数 (例如对于数据集 PASCAL VOC 2012, 解码过程中通道数为 21), 这将有效减少压缩特征数的成本. 解码器模块经过两次加和、三次双线性上采样逐渐增大特征分辨率, 并在此过程中不断融合具有相同分辨率的浅层语义信息, 从而优化物体分割细节.

### 2.3 鉴别网络

这里鉴别器沿用了文献[11]采用全卷积结构. 整个鉴别器由 5 个卷积核为  $4 \times 4$ 、步长为 2 的卷积层组成, 每个卷积层的通道数依次为 64, 128, 256, 512, 1. 此外, 为了将输出特征图放缩到原图大小, 本文在整个鉴别器的最后加了一个上采样层.



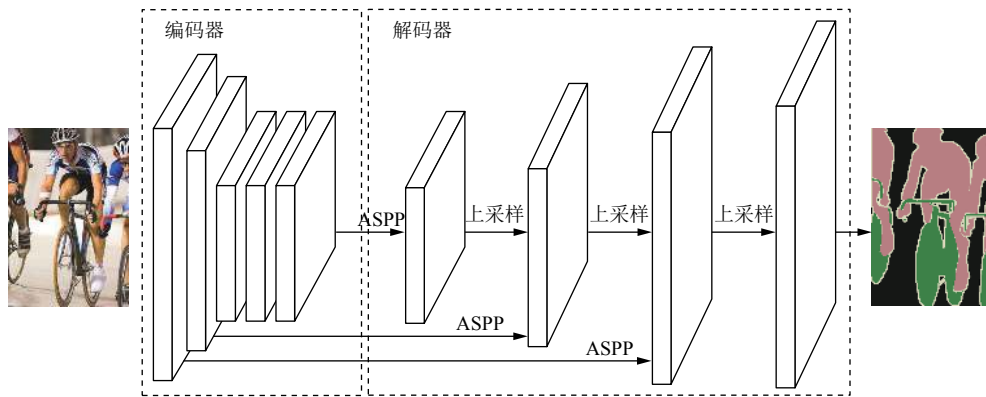


图3 分割网络示意图

### 2.4 模型训练过程和损失函数

本文使用以下符号表示法:  $X_n$  表示大小为  $H \times W \times 3$  的待分割 RGB 图像,  $Y_n$  表示大小为  $H \times W \times C$  的图像真实标签, 其中  $C$  为类别数,  $G(\cdot)$  表示分割网络, 其输出维度同样为  $H \times W \times C$ ,  $D(\cdot)$  表示鉴别网络, 其输出维度为  $H \times W \times 1$ .

(1) 鉴别网络: 本文通过最小化二分类交叉熵损失函数  $L_D$  来训练鉴别网络.

$$L_D = - \left( \sum_{i=1}^{H \times W} \log(1 - D(G(X_n))^{(i)}) + \log(D(Y_n)^{(i)}) \right) \quad (2)$$

其中,  $i$  为位置标号. 需要指出的是, 为了保持分割网络和鉴别网络的平衡这里只有有标签数据训练鉴别网络.

(2) 分割网络: 本文通过最小化多任务交叉熵损失函数  $L_{seg}$  来训练分割网络, 它是标准多分类交叉熵损失和对抗损失的加权和. 由式 (3) 算得:

$$L_{seg} = L_{mce} + \lambda_{adv} L_{adv} + \lambda_{semi} L_{semi} \quad (3)$$

式 (3) 中的第一项  $L_{mce}$  为多分类交叉熵损失, 它是应用于图像语义分割任务的标准损失函数,  $L_{adv}$  是对抗损失,  $L_{semi}$  为半监督学习损失.  $\lambda_{adv}$  和  $\lambda_{semi}$  是为了最小化整个损失函数设置的两个超参数.

首先考虑用有标签数据训练分割网络的情况. 给定一张待分割图像  $X_n$ , 分割网络生成的分割图为  $G(X_n)$ , 则多分类交叉熵  $L_{mce}$  可由式 (4) 算得:

$$L_{mce} = - \sum_{i=1}^{H \times W} \sum_{c=1}^C Y_n^{(i,c)} \log(G(X_n)^{(i,c)}) \quad (4)$$

对抗损失函数由式 (5) 算得:

$$L_{adv} = - \sum_{i=1}^{H \times W} \log(D(G(X_n))^{(i)}) \quad (5)$$

通过对抗损失, 整个网络模型将鼓励分割网络生成对于鉴别网络而言难以和图像真实标签区分开的分

割图, 从而达到优化目标分割细节的目的.

当用无标签数据训练分割网络时, 我们考虑在半监督条件下的对抗训练. 对于无标签数据, 由式 (4) 和式 (5) 可知, 由于缺少图像真实标签  $Y_n$ , 无法直接计算多分类交叉熵损失  $L_{mce}$ , 但对抗损失  $L_{adv}$  只涉及到分割网络、鉴别网络和输入图像  $X_n$ , 所以依然有效. 为实现半监督学习任务, 我们利用已经过少量有标签数据训练的鉴别网络和大量的无标签数据, 以一种自学习的方式来进一步训练分割网络. 主要思想是, 整个网络模型经少量有标签数据训练之后, 鉴别器已基本具备对生成的分割结果进行鉴别的能力, 由于分割网络的分割性能还没有达到足够好的状态, 为无标签数据生成的分割图中存在一定的被严重错误分割的图像区域, 如果这些区域一起参与到进一步训练分割网络中, 将会严重影响分割网络性能. 这里用全卷积鉴别器为无标签数据生成置信图, 用来推断生成的分割结果中哪些区域足够接近图像真实标签分布, 然后将置信图二值化, 以突出可信区域, 最后将分割结果中的可信区域作为无标签图像的伪标签, 实行半监督学习. 则半监督损失  $L_{semi}$  可由式 (6) 算得:

$$L_{semi} = - \sum_{i=1}^{H \times W} \sum_{c=1}^C I(D(G(X_n))^{(i)} > T_{semi}) Y'_n \log(G(X_n)^{(i)}) \quad (6)$$

其中,  $I(\cdot)$  为指示函数,  $Y'_n$  为由分割网络生成的分割图. 在训练期间可以发现伪标签  $Y'_n$  和指示函数的值都是常数, 因此可以将式 (6) 简单的看作是带掩码的多分类交叉熵损失函数.

### 3 实验与分析

为评估模型性能, 本文在图像语义分割数据集

PASCAL VOC 2012<sup>[14]</sup>和 Cityscapes<sup>[15]</sup>上进行了一系列实验. 整个模型实现在具有 11G 内存的单个 GPU k40 上, 为训练分割网络, 本文采用 SGD 优化器, 初始学习率为 $2.5 \times 10^{-4}$ , 并按多项式衰减, 衰减率为 0.9, 动量为 0.9, 权重衰减率为 $10^{-4}$ . 对于鉴别器网络, 本文使用 Adam 优化器, 学习率为 $10^{-4}$ . 使用平均交并比 (mIoU) 作为测试指标.

PASCAL VOC 2012 包含 20 种常见对象类和背景类. 本文使用的是由文献[16]提供额外标注的增强 PASCAL VOC 2012 数据集, 共有 10 582 张训练图像, 1449 张验证图像. Cityscapes 包含 19 种对象类和背景类, 共有 2975 张训练图像, 500 张验证图像和 1525 张测试图像. 表 1、表 2 分别给出了模型在数据集 PASCAL VOC 2012 和 Cityscapes 验证集上的评估结果, 并与其他半监督图像语义分割方法进行了比较, 其中基准方法指的是没有加入对抗训练的分割网络, 本文基准方法是指编码器-解码器结构的分割网络. 为营造半监督学习的环境, 本文从 10 582 张有标签训练集中随机采样 1/8, 1/4, 1/2 数据, 保留其标签作为有标签数据, 其余图像用作无标签数据. 实验表明, 本文提出的方法在两个数据集上都取得更优的分割性能, 其中在 Cityscapes 验证集上更是比文献[11]高出 2.4%–4%. 同时对抗训练的加入, 使得模型平均分别在两个数据集取得了 3.6%、3.2% 的性能提升.

表 1 PASCAL VOC 2012 验证集上半监督方法性能比较

算法	mIOU(%)		
	1/8	1/4	1/2
基准方法 <sup>[11]</sup>	66.0	68.3	69.8
文献 <sup>[11]</sup>	69.5	72.1	73.8
基准方法	67.9	69.6	71.1
本方法	71.3	73.4	74.7

表 2 Cityscapes 验证集上半监督方法性能比较

算法	mIOU(%)		
	1/8	1/4	1/2
基准方法 <sup>[11]</sup>	55.5	59.9	64.1
文献 <sup>[11]</sup>	58.8	62.3	65.7
基准方法	58.2	63.4	65.2
本方法	61.3	66.3	68.8

虽然本文提出的算法是用于解决半监督语义分割问题的, 但依然可以用全部有标签数据参与训练并验证模型有效性. 表 3、表 4 分别展示的是在数据集 PASCAL VOC 2012 和 Cityscapes 上, 用全 Z 部有标签数据训练

模型的评估结果, 并与主流的图像语义分割模型做了对比. 实验结果显示, 本文所提出的方法与目前主流图像语义分割方法相比具有一定优势. 需要指出的是, 基于 ResNet-101 的 DeepLab-v2 取得的性能略优于本文所提出方法. 这主要是因为, 为了减少内存占用量, 本文并没有使用 DeepLab-v2 用到的多尺度输入, 并对分类器 ASPP 做了一定的简化, 并且在 DeepLab-v2 中也提到, 将同一图像缩放成不同尺度输入到深度卷积神经, 并将得到的不同粒度的特征聚合, 可以取得更好的性能提升. 图 4、图 5 分别给出了本文提出算法在数据集 PASCAL VOC 2012 和 Cityscapes 上用 1/8 有标签数据训练之后, 在验证集上的分割效果展示, 从左到右依次为输入图像、图像真实标签和本方法分割结果.

表 3 PASCAL VOC 2012 验证集上实验结果对比

算法	mIoU(%)
FCN-8s <sup>[2]</sup>	67.2
Dilation10 <sup>[3]</sup>	73.9
DeepLab-v2-VGG16 <sup>[4]</sup>	71.57
DeepLab-v2-ResNet-101 <sup>[4]</sup>	76.35
文献 <sup>[10]</sup>	72.0
文献 <sup>[11]</sup>	74.9
基准方法	73.7
本方法	75.8

表 4 Cityscapes 验证集上实验结果对比

算法	mIoU(%)
FCN-8s <sup>[2]</sup>	65.3
Dilation10 <sup>[3]</sup>	67.1
DeepLab-v2-ResNet-101 <sup>[4]</sup>	70.4
文献 <sup>[11]</sup>	67.7
基准方法	67.3
本方法	70.1



图 4 在 PASCAL VOC 2012 验证集上分割效果展示



图5 在 Cityscapes 验证集上分割效果展示

#### 4 总结与展望

本文基于 DeepLab-v2 提出了一种简洁、高效的编码器-解码器分割网络,并将其与生成对抗网络相结合,应用到半监督图像语义分割任务中。解码器模块逐步上采样逐渐增大特征分辨率的同时,不断融合编码器浅层语义信息,从而优化目标分割细节。在 PASCAL VOC 2012 和 Cityscapes 数据集上的实验表明,该方法性能明显优于其他半监督图像语义分割方法。目前该方法在融合浅层特征时只是简单的将这些特征加和,但深层特征更抽象,表达能力也更强。接下来将考虑不同层次特征融合时的权重,以期进一步改善半监督图像语义分割性能。

#### 参考文献

- 1 郑菲, 孟朝晖, 郭闯世. 基于反卷积特征学习的图像语义分割算法. 计算机系统应用, 2019, 28(1): 147-155. [doi: 10.15888/j.cnki.csa.006716]
- 2 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 3431-3440.
- 3 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv: 1511.07122, 2016.
- 4 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]
- 5 Pinheiro PO, Collobert R. Weakly supervised semantic segmentation with convolutional networks. arXiv: 1411.6228v1, 2015.
- 6 Dai JF, He KM, Sun J. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1635-1643.
- 7 Hong S, Noh H, Han B. Decoupled deep neural network for semi-supervised semantic segmentation. Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2015. 1495-1503.
- 8 Goodfellow I, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, QC, Canada. 2014. 2672-2680.
- 9 Isola P, Zhu JY, Zhou TH, *et al.* Image-to-image translation with conditional adversarial networks. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 1125-1134.
- 10 Luc P, Couprie C, Chintala S, *et al.* Semantic segmentation using adversarial networks. arXiv: 1611.08408, 2016.
- 11 Hung WC, Tsai YH, Liou YT, *et al.* Adversarial learning for semi-supervised semantic segmentation. arXiv: 1802.07934, 2018.
- 12 Souly N, Spampinato C, Shah M. Semi supervised semantic segmentation using generative adversarial network. 2007 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 5688-5696.
- 13 He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770-778.
- 14 Everingham M, Van Gool L, Williams CKI, *et al.* The pascal visual object classes (VOC) challenge. International Journal of Computer Vision, 2010, 88(2): 303-338. [doi: 10.1007/s11263-009-0275-4]
- 15 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, NV, USA. 2016. 3213-3223.
- 16 Hariharan B, Arbeláez P, Bourdev L, *et al.* Semantic contours from inverse detectors. IEEE 2011 International Conference on Computer Vision. Barcelona, Spain. 2011. 991-998.