

DOI: 10.13245/j.hust.181205

# 基于环视相机的无人驾驶汽车实例分割方法

邓璇元 杨明 王春香 王冰

(上海交通大学 a 自动化系; b 系统控制与信息处理教育部重点实验室, 上海 200240)

**摘要** 针对环视鱼眼图像中目标几何畸变大导致建模难的问题, 提出一种基于可变形卷积网络的实例分割方法, 主要是在 Mask R-CNN 框架的基础上引入可变形卷积和可变形 RoI Pooling(候选区域池化)来提升网络对几何畸变的建模能力. 针对神经网络训练数据缺乏、易过拟合的问题, 提出了基于多任务学习的训练方法. 首先将现有大规模普通图像数据集转换为鱼眼数据集来弥补训练数据不足的问题, 然后采用多任务学习的训练方法将转换的图像和真实图像放在同一个框架中训练以提高网络的泛化能力. 用该方法在真实的环视鱼眼图像上做测试, 结果表明: 相对于原始 Mask R-CNN 的方法平均精度提升了 3.1%, 证明了该方法在真实交通环境中的有效性.

**关键词** 图像处理; 无人驾驶; 环境感知; 实例分割; 可变形卷积网络; 多任务学习; 环视相机

中图分类号 TP391.4 文献标志码 A 文章编号 1671-4512(2018)12-0024-06

## Surround view cameras based instance segmentation method for autonomous vehicles

Deng Liuyuan Yang Ming Wang Chunxiang Wang Bing

(a Department of Automation; b Key Laboratory of System Control and Information Processing, Ministry of Education of China, Shanghai Jiao Tong University, Shanghai 200240, China)

**Abstract** Aimed at handling the problem of modeling large geometric distortions caused by surround view cameras, a deformable convolutional networks based instance segmentation method was proposed. The method introduced the deformable convolution and deformable RoI (region of interest) Pooling into the framework of Mask R-CNN. Aimed at handling the problems of insufficient data to train the deep neural networks and overfitting, a multi-task learning based training method was proposed. First, an existing large-scale dataset of conventional images was transformed to a fisheye-style dataset to compensate the lack, and then a multi-task learning method was adopted to train the transformed images and real-world images in a united architecture to improve the generalization ability. The proposed method was tested on the real-world fisheye images. It shows an improvement of 3.1% over the original Mask R-CNN method, which demonstrates the effectiveness of the proposed method in real-world traffic environments.

**Key words** image processing; autonomous driving; environmental perception; instance segmentation; deformable convolutional networks; multi-task learning; surround view cameras

无人驾驶汽车是室外轮式移动机器人在交通领域的重要应用<sup>[1]</sup>. 无人车须要广泛地感知车辆周围环境, 识别交通环境中的车辆、行人等道路使用者, 来安全地导航与控制. 目前, 由四个 180°鱼眼相机组成的环视相机已经广泛应用于消费级汽车中, 用

来为驾驶员提供车辆周围 360°的视野, 消除驾驶员观察的盲区. 在无人驾驶中, 可以通过环视相机检测周围的道路使用者, 为无人车提供 360°的感知结果. 普通的目标检测方法都是用边界框来表示检测的结果, 这种结果表达方式不能精确地表达出目标

收稿日期 2018-06-01.

作者简介 邓璇元(1990-), 男, 博士研究生, E-mail: lydeng@sjtu.edu.cn.

基金项目 国家自然科学基金资助项目(U1764264, 61873165); 上海汽车工业科技发展基金资助项目(1733, 1807).

在图像中的位置. 特别是在环视的鱼眼图像中, 当目标离车辆比较近时, 用边界框来表达会有极大误差. 本研究探讨环视鱼眼图像中的交通目标实例分割, 力求不仅检测出目标, 还将目标边缘清晰地分割出来.

得益于卷积神经网络(CNN)在视觉图像处理上的发展, 近年来目标检测和语义分割都取得了巨大的进步. 而实例分割更具有挑战性, 要同时完成检测和分割的任务, 目前主要沿着两个方向来研究实例分割. 第一个方向是以检测为基础, 先得到掩码建议再进行分割. 早期的方法<sup>[2-3]</sup>用 MCG<sup>[3]</sup>或选择性搜索<sup>[4]</sup>处理图像来获得掩码建议, 但速度很慢. MNC<sup>[5]</sup>通过用 CNN 实现掩码建议和掩码分割, 来提高推理的速度. 文献[6]提出了一个通用的目标实例分割框架 Mask R-CNN, 该方法是在 Faster R-CNN<sup>[7]</sup>目标检测的基础上扩展了对每个感兴趣区域(RoI)进行掩码分割的分支. 该分支可以与分类和边界框回归并行执行, Mask R-CNN 是当前这一类最好的方法之一. 另一个方向是在分割的基础上计算得到目标实例, 如 DWT<sup>[8]</sup>, SGN<sup>[9]</sup>, DeepMask<sup>[10]</sup>和 FCIS<sup>[11]</sup>. 其中 FCIS 是在 InstanceFCN<sup>[12]</sup>的启发下实现的第一个全卷积端对端的实例分割解决方案. FCIS 在精度和效率上都取得了不错的性能, 但当对相互重叠的目标进行实例分割时容易出现误检.

基于 CNN 技术, 对普通图像的实例分割精度已经取得很大的提升, 但目前还没有针对鱼眼图像的实例分割方法. 在鱼眼图像中, 目标存在复杂的形变, 该形变与目标相对相机的距离和角度都有关. 而普通的 CNN 在本质上对几何形变的建模能力有限, 主要因其采用的是固定的几何结构, 例如固定的卷积核采样位置(一般是  $3 \times 3$  的固定栅格)和固定的池化层(一般是  $2 \times 2$  的固定栅格). 如何让网络有效地对这些复杂几何形变建模是本研究的难点之一. 本研究提出基于可变形卷积网络的目标实例分割方法, 引入可变形卷积和可变形 RoI Pooling<sup>[13]</sup>来提升网络对几何形变的建模能力.

另一方面, 目前最先进的目标实例分割都是基于有监督的深度学习方法. 这类方法须要用大量的带标注的数据去训练, 来提高网络的泛化能力. 本研究将基于深度学习的目标实例分割方法用在真实交通环境里的环视鱼眼图像上, 但是目前并不存在公开的大规模的带标注的鱼眼实例分割数据集, 人工标注大规模的实例分割数据集往往费时费力. 为解决网络训练的问题, 本研究首先将现有的大规模

的普通图像实例分割数据集转换成鱼眼数据集来弥补训练数据的不足, 再结合人工标注的小规模的环视鱼眼图像, 采用基于多任务学习的训练框架来完成训练, 最终实现真实交通环境下, 环视鱼眼图像中准确的交通目标实例分割.

## 1 基于可变形卷积网络的目标实例分割方法

目标的实例分割不仅要检测出目标的位置, 识别出目标的类别, 还须要将每个目标实例分割出来. 为了同时完成检测和分割的任务, 本研究采用 Mask R-CNN<sup>[6]</sup>作为基础框架. 为提升网络对鱼眼图像中目标几何形变的建模能力, 在 Mask R-CNN 中引入可变形卷积和可变形 RoI Pooling<sup>[13]</sup>.

### 1.1 Mask R-CNN 实例分割框架

Mask R-CNN 是在 Faster R-CNN 目标检测的基础上扩展了 FCN<sup>[14]</sup>分割网络作为预测分割掩码的分支. 整个网络可分为特征提取、区域建议、区域分类、边框回归和掩码分割. 首先用合适的特征模型对图像进行有效的高层特征表示, 然后对特征层中的候选目标区域提出建议, 再用分类器对候选区域进行类别判断, 并通过回归得到精确的目标边界框, 最后对目标区域进行分割, 得到最终目标实例.

本研究采用 ResNet-50<sup>[15]</sup>并结合 FPN<sup>[16]</sup>作为特征模型提取图像的高层次抽象表示, 其中 FPN<sup>[16]</sup>采用自顶向下和横向连接的架构, 能够很好地利用高层特征丰富的语义信息和底层特征准确的细节信息. 区域建议的生成使用 RPN(区域建议网络)<sup>[7]</sup>, RPN 能快速地从特征层中生成候选目标边界框, 最后再通过 RoI Pooling(候选区域池化)模块<sup>[7]</sup>从候选框中提取特征, 对每个候选区域进行分类、边界框回归和掩码分割. 在训练过程中, 分类、边界框回归和掩码分割三个任务是同时进行的, 整个网络可进行端对端的训练.

### 1.2 基于可变形卷积网络的目标实例分割

普通的卷积采用的是固定的采样栅格, 如对于一个  $3 \times 3$  的卷积核, 采样栅格定义为

$$\mathbf{R} = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\},$$

对每一个输出位置  $\mathbf{p}_0$  有

$$\mathbf{y}(\mathbf{p}_0) = \sum_{\mathbf{p}_n \in \mathbf{R}} \mathbf{w}(\mathbf{p}_n) \cdot \mathbf{x}(\mathbf{p}_0 + \mathbf{p}_n),$$

式中:  $\mathbf{p}_n$  为采样栅格中的第  $n$  个位置;  $\mathbf{w}$  为卷积核的权值;  $\mathbf{x}$  为输入的特征层.  $3 \times 3$  可变形卷积在普通卷积的基础上, 对每一个采样位置都引入了一个

可以学习的位置偏移量  $\Delta p_n$ ，定义为

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot x(p_0 + p_n + \Delta p_n),$$

$\Delta p_n$  是从输入特征层中学习到的，使得对于每一个输出位置  $p_0$ ，都可以通过学习得到任意形状的采样栅格。

RoI Pooling 是用来从任意大小感兴趣区域中提取固定大小特征的方法，其采样形式也是固定的采样栅格。假设 RoI 在上角位置为  $c_0$ ，RoI Pooling 将此 RoI 划分为  $k \times k$  的区间，对第  $(i, j)$  个区间输出

$$y(i, j) = \sum_{c \in \text{bin}(i, j)} x(c_0 + c) / n_{ij},$$

式中  $n_{ij}$  为第  $(i, j)$  个区间内像素的总数。与可变形卷积原理一样，可变形 RoI Pooling 是在普通 RoI Pooling 的基础上，对每一个采样位置添加了一个可学习的偏移矢量  $\Delta c_{ij}$ ，使得对不同形状的目标可以自适应地学习到不同形状的采样栅格，其定义为

$$y(i, j) = \sum_{c \in \text{bin}(i, j)} x(c_0 + c + \Delta c_{ij}) / n_{ij}.$$

可变形卷积和可变形 RoI Pooling 只在原有的模块基础上添加了少量的参数，可以用它们去替换

网络中的对应模块，同样可以进行端对端的训练。对这两个可变形模块的更详细说明参见文献[13]。

本研究利用可变形卷积和可变形 RoI Pooling 的特性来对鱼眼图像中目标的复杂形变建模，增强 Mask R-CNN 框架为可变形卷积网络，在特征提取和 RoI Pooling 的过程中分别引入可变形卷积和可变形 RoI Pooling。具体地，在 ResNet-50 的最后三个阶段中加入可变形卷积：在最后一个阶段的每个块里的卷积层中，以及在倒数第二个和第三个阶段的最后一个块中的  $3 \times 3$  卷积层中使用可变卷积。将 RoI Pooling 直接替换为对应的可变形模块，可以帮助分类和边界框回归，但是并不利于掩码的分割。因为在可变形 RoI Pooling 中，采样位置是根据输入的特征层自适应学习的，这将改变输入图像和输出结果之间的空间对应关系。

本研究采用两种不同的方法从感兴趣区域提取特征，对分类和边界框回归使用可变形 RoI Pooling，对掩码分割使用 RoIAlign<sup>[6]</sup>，RoIAlign 相对于普通的 RoI Pooling 可以减少量化误差。总的框架如图 1 所示。

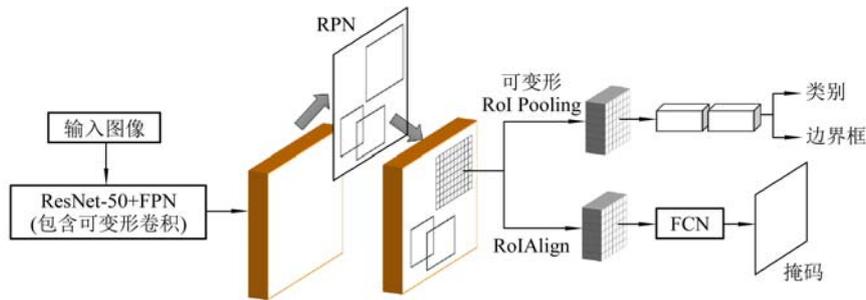


图 1 基于可变形卷积网络的实例分割框架

## 2 基于多任务学习的训练方法

### 2.1 大规模训练数据集的生成

本研究用文献[17]提出的方法将普通图像转换成鱼眼图像，核心思想是根据鱼眼相机的成像模型，构建普通图像平面到鱼眼图像平面的映射函数，从而将普通图像中的场景映射到鱼眼图像中去。鱼眼采用最常用的等距投影模型，即

$$r = f\theta, \tag{1}$$

式中： $r$  为图像中的点到光轴中心的距离； $f$  为相机的焦距； $\theta$  为入射光线与相机光轴之间的夹角，即入射角。而理想的透视模型针孔成像模型为

$$r = f \tan \theta. \tag{2}$$

根据式(1)和(2)，可以得到从鱼眼到普通图像的映射

函数

$$r_c = f_0 \tan(r_f / f_0), \tag{3}$$

式中： $r_c$  和  $r_f$  分别为鱼眼图像和普通图像中的点到相机光轴中心的距离； $f_0$  为可调节的基准焦距。通过式(3)可以把普通图像转换成鱼眼图像。

### 2.2 多任务学习

利用转换的大规模数据集来帮助提高真实图像下的实例分割性能，最简单的方式是用大规模数据集来训练网络模型，再根据迁移学习的思想，用真实图像来微调这个模型。但是当标注的真实图像规模比较小时，网络很容易过拟合。而本研究采用多任务的学习框架，将真实图像和转换的图像上的实例分割构建为两个相关的子任务，在同一个框架下一起训练，以提高网络在真实场景中的泛化能力。

图 2 所示是基于多任务学习的训练框架，图中： $L_{cls\_r}$ ， $L_{box\_r}$  和  $L_{mask\_r}$  为真实图像上对应的分类、

回归和分割的损失;  $L_{cls\_t}$ ,  $L_{box\_t}$  和  $L_{mask\_t}$  为转换图像上的损失. 两个任务在学习过程中, 除了最后一个与类别相关的作为分类器的全连接或卷积层以外, 其他的层都共享权值. 通过共享网络的权值来学习两个任务的共享表示, 可以将一个任务的知识迁移到另一个任务中. 在多任务学习中仍有两个难点: 一是两个任务图像采集的环境不同, 图像的统计分布有差异, 直接共享网络会增加学习共享权值的难度, 不能学习到最佳的共享特征表示; 二是如何平衡两个任务的损失权重, 若真实图像的权重太大, 则容易过拟合, 若权重太小则在真实图像上的性能会不够好.

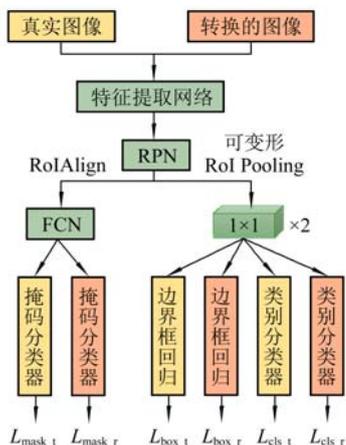


图 2 基于多任务学习的训练框架

对于第一个难点, 通过尽量将转换图像和真实图像的统计分布独立开的方式, 来学习更好的共享权值. 本研究利用 AdaBN<sup>[18]</sup> 的思想和 Batch Normalization (BN) 层<sup>[19]</sup> 分别计算两个任务的统计分布, 对于数量为  $m$  的一小批的输入样本  $\mathbf{X} = \{x_1, x_2, \dots, x_m\}$ , BN 层将对每一个维样本计算均值  $E[x]$  和方差  $\text{Var}[x]$ , 进行归一化后再进行重构变换得到新的输出  $y_i$ , 公式如下:

$$E[x] = \frac{1}{m} \sum_{i=1}^m x_i; \text{Var}[x] = \frac{1}{m} \sum_{i=1}^m (x_i - E[x])^2;$$

$$\hat{x}_i = \frac{x_i - E[x]}{\sqrt{\text{Var}[x] + \epsilon}}; y_i = \gamma \hat{x}_i + \beta.$$

在训练过程中将转换的图像和真实的图像数据分开, 分别计算 BN 层中的均值  $E[x]$  和方差  $\text{Var}[x]$ . 在每个卷积层前面都添加了 BN 层, 原本的 FPN 分支中并不包含 BN 层, 为提高性能也在每一个卷积层前添加了 BN 层. 通过这种方式, 将两个不同分布的数据集放在了同一个框架下学习.

对于第二个难点, 本研究采用不同的损失权重比例来训练特征提取的网络层 (ResNet-50+FPN) 和与任务相关的分类器网络层. 任务包括检测(分类和

边界框回归)和掩码分割. 定义训练过程中网络总的损失为  $L = (1 - \alpha)(L_{cls\_r} + L_{box\_r} + L_{mask\_r}) + \alpha(L_{cls\_t} + L_{box\_t} + L_{mask\_t})$ , 式中  $\alpha$  为任务损失比例因子, 用来平衡真实图像和转换的图像带来的损失. 通过改变  $\alpha$  来调整两个任务的权重比例, 例如, 当  $\alpha = 0.5$  时, 两个任务的损失权重相同.  $\alpha$  在训练过程中类似一个正则项因子, 若  $\alpha$  太小, 则模型容易过拟合; 若  $\alpha$  过大, 则网络更专注于学习转换的图像, 模型容易欠拟合. 为了使网络取得更好的泛化性能, 本研究在训练网络前端时采用较大的  $\alpha$ ; 在固定前端的权值后, 用较小的  $\alpha$  来训练任务相关部分网络.

### 3 实验

#### 3.1 训练数据

为了实现真实交通环境下的环视鱼眼图像实例分割, 用安装了环视相机的车辆在真实的道路上采集鱼眼图像, 并手动标注了 600 张鱼眼图像, 其中 400 张用于训练, 200 张用于测试. 共标注了 8 个实例类别, 有汽车、卡车、公交车、摩托车、自行车、行人和骑车的人. 但是对于深度学习网络, 400 张图像用于训练是不够的. 图 3 是训练样本的示例, 左边一列是真实交通环境下的环视鱼眼图像, 右边一列是 Fisheye Cityscapes 中的图像. 本研究用前面提到的方法把 Cityscapes 数据集<sup>[20]</sup> 转换成鱼眼数据



图 3 训练样本示例

库, 记为 Fisheye Cityscapes. Fisheye Cityscapes 有 2 975 张用于训练, 500 张用于测试. 从图 3 中可以看出真实的鱼眼图像和转换的鱼眼图像中的目标都发生了较大的形变.

### 3.2 训练细节

实验在 Intel Xeon E5-2620 CPU 和两块 Nvidia 1080Ti GPU 的平台上进行, 采用 MXNet 搭建和训练网络, 用 Mask R-CNN 作为基础框架. 使用 ResNet-50 和 FPN 作为特征提取的部分, 在 ImageNet 上预训练的权值可以用来初始化这部分网络. 可变形卷积和可变形 RoI pooling 与文献[13]中的初始化方法相同, 其他的网络层用 xavier 方法初始化.

采用四步交替的训练方法<sup>[7]</sup>. 第一步, 训练 RPN, 用 ImageNet 预训练的权值来初始化; 第二步, 用第一步 RPN 生成的区域建议来训练检测和分割网络, 也是用 ImageNet 预训练的权值来初始化; 第三步, 再次训练 RPN, 这次是用第二步训练的网络来初始化, 训练过程中固定特征提取部分(ResNet-50+FPN)的权值; 第四步, 用第三步的 RPN 生成的区域建议再次训练检测和分割网络, 用第三步所用的权值来初始化, 也固定特征提取部分的权值. 这样第三步和第四步特征提取部分用的是同样的权值, 最后将特征提取、RPN、检测和分割网络的训练权值合成在一个统一的网络里.

当进行多任务训练时, 在训练第一步和第二步中使用任务损失比例因子  $\alpha = 0.5$ , 第三步和第四步中任务损失因子  $\alpha = 0.3$ .

### 3.3 实验结果

将原始的 Mask R-CNN 方法作为基准方法与本文提出的方法进行比较. 基准方法使用的是 RoIAlign 来提取感兴趣区域的特征, 采用平均精度 (AP)<sup>[20]</sup>作为度量指标. 首先, 在 Fisheye Cityscapes 数据集上比较基准方法与本文的基于可变形卷积网络的实例分割方法的性能. 如表 1 所示, 在 Fisheye Cityscapes 数据集上的测试结果都不太高, 说明鱼眼图像比普通图像的难度要大, 本文方法比基准方法有了 0.8% 的 AP 提升. 然后, 本研究结合大规模的 Fisheye Cityscapes 数据集和小规模的人工标注的真实环视鱼眼图像来训练网络, 在真实环境的图像上比较基准方法与本文方法. 如表 2 所示, 本文的方法较基准方法有了 3.1% 的 AP 提升. 图 4 展示

表 1 Fisheye Cityscapes 数据集上测试结果 %

方法	AP	AP <sub>50</sub>
基准方法	19.5	39.7
本文方法	20.3	40.1

了用本文方法处理的实例分割结果, 真实的鱼眼图像的分辨率是 512×864, 一张图像处理需要的时间是 1.1 s.

表 2 真实环视鱼眼图像上测试结果比较 %

方法	AP	AP <sub>50</sub>
基准方法	24.5	53.8
本文方法	27.6	56.8

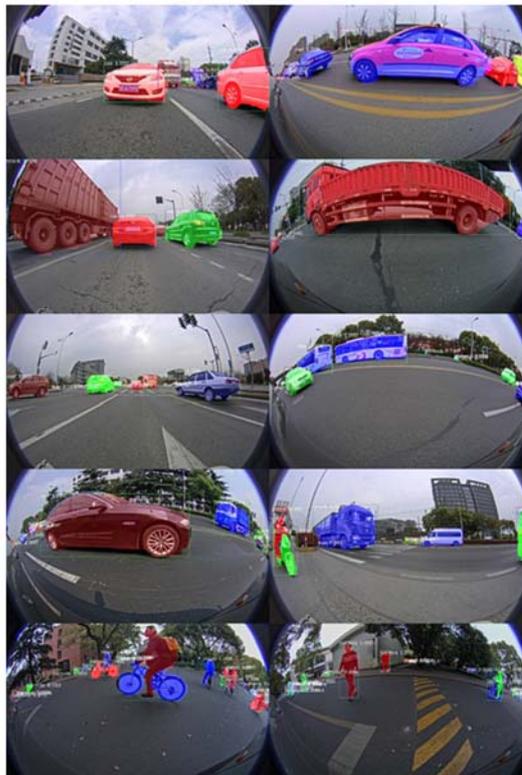


图 4 真实环视鱼眼图像上的测试结果

## 4 结语

本研究提出了基于环视相机的实例分割方法, 对环视鱼眼图像中的交通目标进行实例分割. 通过引入可变形卷积和可变形 RoI Pooling 来增强网络对几何形变的建模能力, 提出一种基于可变形卷积网络的实例分割方法. 转换现有的大规模的数据集来弥补训练数据的不足, 采用多任务学习的训练方法来提高网络在真实环视鱼眼图像中的泛化能力. 在真实环视鱼眼图像下的实验表明: 本文提出方法相比于以前的方法有了明显的性能提升, 可以进行环视鱼眼图像实例分割. 尽管本文方法已经取得了一定的性能提升, 但是实例分割的整体性能以及实时性还是偏低, 仍然有很大的提升空间.

### 参 考 文 献

[1] 杨明. 无人自动驾驶车辆研究综述与展望[J]. 哈尔滨

- 工业大学学报, 2006, 38(8): 1259-1262.
- [2] Hariharan B, Arbeláez P, Girshick R, et al. Simultaneous detection and segmentation[C]// Proc of European Conference on Computer Vision. Cham: Springer, 2014: 297-312.
- [3] Hariharan B, Arbeláez P, Girshick R, et al. Hypercolumns for object segmentation and fine-grained localization[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2015: 447-456.
- [4] Uijlings J R R, van de Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International Journal of Computer Vision, 2013, 104(2): 154-171.
- [5] Dai J, He K, Sun J. Instance-aware semantic segmentation via multi-task network cascades[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 3150-3158.
- [6] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]// Proc of 2017 IEEE International Conference on Computer Vision. New York: IEEE, 2017: 2980-2988.
- [7] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [8] Bai M, Urtasun R. Deep watershed transform for instance segmentation[C]// Proc of 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 2858-2866.
- [9] Liu S, Jia J, Fidler S, et al. Sgn: sequential grouping networks for instance segmentation[C]// Proc of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 3516-3524.
- [10] Pinheiro P O, Collobert R, Dollár P. Learning to segment object candidates[C]// Proc of Advances in Neural Information Processing Systems. New York: IEEE, 2015: 1990-1998.
- [11] Li Y, Qi H, Dai J, et al. Fully convolutional instance-aware semantic segmentation[C]// Proc of 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 4438-4446.
- [12] Dai J, He K, Li Y, et al. Instance-sensitive fully convolutional networks[C]// Proc of European Conference on Computer Vision. Cham: Springer, 2016: 534-549.
- [13] Dai J, Qi H, Xiong Y, et al. Deformable convolutional networks[C]// Proc of 2017 IEEE International Conference on Computer Vision. New York: IEEE, 2017: 764-773.
- [14] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 36(10): 1337-1342.
- [15] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770-778.
- [16] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]// Proc of 2017 IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 936-944.
- [17] Deng L, Yang M, Qian Y, et al. CNN based semantic segmentation for urban traffic scenes using fisheye camera[C]// Proc of 2017 IEEE Conference on Intelligent Vehicles Symposium(IV). New York: IEEE, 2017: 231-236.
- [18] Li Y H, Wang N Y, Shi J P, et al. Revisiting batch normalization for practical domain adaptation[J]. Pattern Recognition, 2016, 78(3): 109-117.
- [19] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]// Proc of IEEE International Conference on Machine Learning. New York: IEEE, 2015: 448-456.
- [20] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]// Proc of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 3213-3223.