

对象边框标注数据的弱监督图像语义分割*

徐树奎¹, 周浩²

(1. 中国电子科技集团第28研究所 博士后流动工作站, 江苏南京 210007;

2. 国防科技大学 系统工程学院, 湖南长沙 410073)

摘要:针对图像语义分割应用中像素级标注数据费时昂贵的问题,主要研究以对象边框标注数据为代表的弱监督模型下的图像语义分割方法。使用基于金字塔的密集采样全卷积网络提取图像的像素级特征,并用 GrabCut 算法转化对弱监督数据进行数据标记,通过将图像特征和标记数据进行联合训练,构建了基于金字塔密集采样全卷积网络的对象边框标注弱监督图像语义分割模型,并在公开数据集上进行了验证。实验结果表明,所构建的弱监督模型与 DET3-Proposed 模型、全矩形转化模型以及 Bbox-Seg 模型相比,达到了更好的分割效果。

关键词:图像语义分割;全卷积网络;弱监督模型;GrabCut 算法

中图分类号:TP391 **文献标志码:**A **文章编号:**1001-2486(2020)01-187-07

Image semantic segmentation of weakly supervised model with bounding box annotations data

XU Shukui¹, ZHOU Hao²

(1. Mobile Postdoctoral Center, The 28th Research Institute of China Electronic Technology Group Corporation, Nanjing 210007, China;

2. College of Systems Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract: For the pixel level marked data of image segmentation is time consuming and expensive, the image segmentation under the weak supervised model with bounding box annotations was studied. The high-resolution pixel features were extracted by a pyramid densely sampled fully convolution network, and then the weakly supervised data by the GrabCut algorithm was transformed. Finally, the image features and the labeled data were jointly trained. A model of weakly supervised image semantic segmentation based on fully convolution network was constructed and verified on public data set. Experiments show that the constructed weak supervised model has a better segmentation effect compared with DET3-Proposed model, expectation-maximization model and Bbox-Seg model.

Keywords: image semantic segmentation; fully convolution network; weakly supervised model; GrabCut algorithm

图像语义分割在特征提取和像素级标注方面一直存在困难和挑战,是一个多学科交叉融合的难题,而深度学习的出现在一定程度上缓解了这些问题^[1]。

图像语义分割通常需要使用全监督数据去进行模型训练,人工进行像素级标注需要消耗大量时间,相关公开数据集中像素级标注的数据非常少。例如,在 ImageNet 数据集^[2]中,有 1400 多万幅图像带有场景类别标签,50 多万幅图像带有边框标注数据,其中仅有 4460 幅图像带有像素级标注数据。为解决图像语义分割像素级标注费时费力的问题,本文主要研究具有边框标记的弱监督图像数据,结合密集采样的金字塔全卷积网络进

行图像语义分割学习。使用深度学习技术可对图像中的目标进行检测和识别,能够快速识别并框出图像数据中的目标对象,在此基础上,将具有目标标识的图像作为训练数据进行像素级的图像语义分割,这就是对象边框标注数据的弱监督图像语义分割^[3]。

对象边框标注无法精确描述目标对象的边界,仅能大致标识出目标对象在图像中的位置,是对目标对象的弱标注,如图 1 所示。Lin 等^[4]的研究表明,对图像进行对象边框标注消耗的时间和花费仅为像素级标注的 1/15。

传统的弱监督图像语义分割中,GrabCut 算法^[5]通过边界框来训练高斯混合模型的参数,区

* 收稿日期:2018-11-01

基金项目:国家自然科学基金资助项目(61671459)

作者简介:徐树奎(1982—),男,山东郓城人,博士,工程师,E-mail:xskgfd@163.com

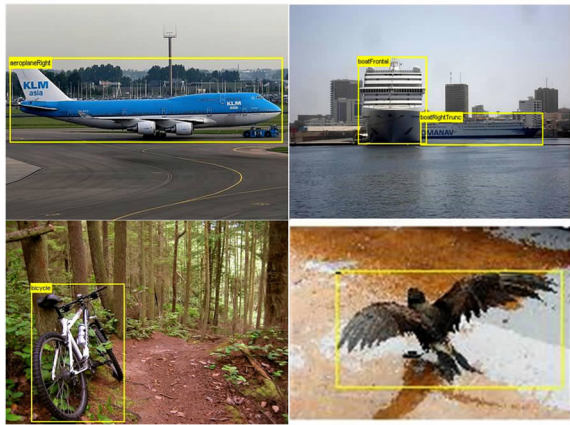


图 1 对象边框标注的弱标注样本示例

Fig. 1 An example graph of the weak data of bounding box annotations

分前景和背景,进而完成图像分割任务。Pandey^[6]利用边界框通过隐变量结构支持向量机(Support Vector Machine, SVM)框架来定位目标对象的像素位置。Vezhnevets^[7]主要通过研究基于对连接相邻图像区域的最小生成树的森林求平均值和基于常规 8 邻域马尔可夫随机场的两个空间模型来实现弱监督。Verbeek 等^[8]通过创建用于恢复训练图像的像素标签多图像模型进行图像级对象标注的弱监督图像语义分割。梁恩辉等^[9]从海上红外图像的热红外特性出发,提出了基于能量图的分割方法。

近几年,深度卷积神经网络在图像语义分割、对象检测等方面的研究取得了一定的进展。陈鸿翔^[10]设计了一种卷积和反卷积层神经网络相结合的深度神经网络结构,能够预测像素所属的语义类别。李春阳^[11]在道路分割识别领域引入深度学习方法中的深度信念网络(Deep Belief Network, DBN)模型进行图像小窗口特征提取,运用图像分割的思想实现道路区域的提取。刘丹等^[12]以超像素为基本处理单元,使用卷积神经网络对图像语义分割上下文的关系进行建模,结合多尺度技术和卷积神经网络设计了一种面向图像语义分割的多尺度卷积神经网络(Convolutional Neural Network, CNN)模型。Arbeláez 等^[13]提出一种基于深度神经网络自下而上的多尺度组合分组图像语义分割方法。结合快速分割算法和高性能分层分割器,使用分组策略,通过探索组合空间,实现多尺度区域的图像分割。Long 等^[14]采用卷积层代替全连接层,形成一个全卷积网络,通过重复利用卷积层的输出提高效率,并通过高低层结合的方式提高像素级特征提取的精确性。

CRFasRNN^[15]将 DenseCRF 真正融入全卷积网络中,形成了一种类似于端对端的循环神经网络(Recurrent Neural Network, RNN)模型,并使用平均场近似的方法来进行,把每一步描述成一层类似卷积的计算。弱监督的图像语义分割的研究者^[16-17]相对较少,Vezhnevets 等^[18]定义了一个结构化模型的参数族,每个模型都以不同的方式对视觉线索进行加权,并给定一个最大预期协议模型选择原则来进行评估。Xu 等^[19]将弱监督图像语义分割形式化为潜在的结构化预测框架中的多实例学习,从而对图形模型中类的存在与否以及对超像素的语义标签的分配进行编码。Papandreou 等^[20]通过全卷积网络进行特征提取和期望最大化算法训练模型结合,取得了较好的效果,但其在对象边框标注的图像语义分割模型中仍存在一定的问題。对于弱监督模型来说,像素的特征提取尤为重要,所以本文在上述模型的基础上引入金字塔模型,在二维密集采样的基础上再进行多尺度特征的提取,同时将全卷积网络与传统的图像语义分割的 GrabCut 算法结合,能够有效地增强分类的准确性。

1 图像特征提取

由于全连接网络层在分类深度卷积网络中会将二维图像转换为一维数据,导致图像空间信息丢失,因此,本文在构建弱监督模型深度网络结构时,将全连接层转换为卷积层,以符合图像语义分割的二维特征要求。另外,通过对最大池化层的密集采样和卷积层的“孔”采样,使输出特征具有高分辨率。

1.1 最大化池化层的密集采样

深度卷积神经网络的池化和下采样会导致数据特征大量损失,无法满足图像语义分割应用要求。本文通过调整池化层滑动步长参数以及在数据周围补零,实现对特征进行密集采样,达到既不破坏池化层特性也尽可能保留特征分辨率的目的。具体的方法如图 2 所示,假设原始特征图为 5×5 ,并且有一个单位的零填充,池化核大小为 3×3 ,滑动步长为 2,则池化操作之后的特征图大小为 3×3 ,特征分辨率降低了 64%,如图 2(a)所示。

若是在池化操作过程中将原本的滑动步长 2 改为 1,则在同样的池化核变换下,特征图大小仍然为 5×5 ,这样通过密集的池化层操作之后特征图的分辨率大小并没有发生变化,如图 2(b)所示。上述参数设置能够保持最大化池化层的本质

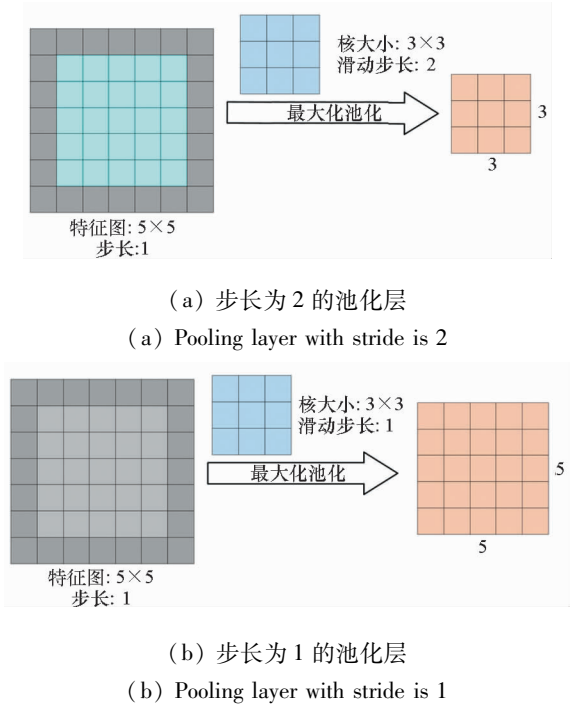


图 2 两种池化层的对比

Fig. 2 Comparison of two pooling layers

特性,保留主要特征,忽略弱性特征和干扰特征,进而学习全局的抽象特征。另外,滑动步长和增加填充还能在选取最有表达力特征的同时兼顾特征的全面性。

1.2 卷积层的“孔”采样

卷积操作也会使特征分辨率缩小,因此也需要对卷积层进行设计,使卷积层在不改变数据结构的情况下,不缩减特征分辨率。通过在非零滤波器中插入“孔”实现卷积层的上采样是一种经常在信号处理中用到的方法,最开始是用于对非下采样小波变换的有效计算^[21]。一些研究者^[22]将这种方法及其变种引入神经网络实现上下文信息结合。本文设计了“孔”采样和双线性差值相结合的方法,保留图像特征分辨率。

将原始滤波器的上采样因子设为 2,通过带“孔”的卷积核进行卷积操作,可以获得图像所有位置的特征响应。上采样会导致卷积核的尺寸增加,由于只需要考虑非零滤波器值,所以卷积核参数的数量和每个位置的操作数量是不变的。如图 3(a) 所示,假设原始特征图为 7×7 ,卷积核大小为 3×3 ,滑动步长为 1,则卷积操作之后的特征图大小为 5×5 。而在图 3(b) 中,若是在卷积操作过程中对原始特征图添加两个单位的零填充,并把卷积核大小扩充为 5×5 ,扩充的部分值为 0,则卷积操作之后的特征图大小仍然为 7×7 ,这样通过孔卷积层操作之后特征图的分辨率大小并没

有发生变化。

上述两种密集采样和“孔”采样方法虽然能够输出与图像原始大小一样的特征图,但是这种做法会导致计算量大,时间效率低,代价难以接受。采用密集采样的池化层、带孔的卷积层与双线性插值相结合的折中方法,使得网络最终的输出特征分辨率与输入数据大小一致。参考 Chen 等^[23]的实验结果,通过密集采样和“孔”采样将原网络的分辨率提升 4 倍,再通过双线性插值获得与输入图像大小一致的特征图。

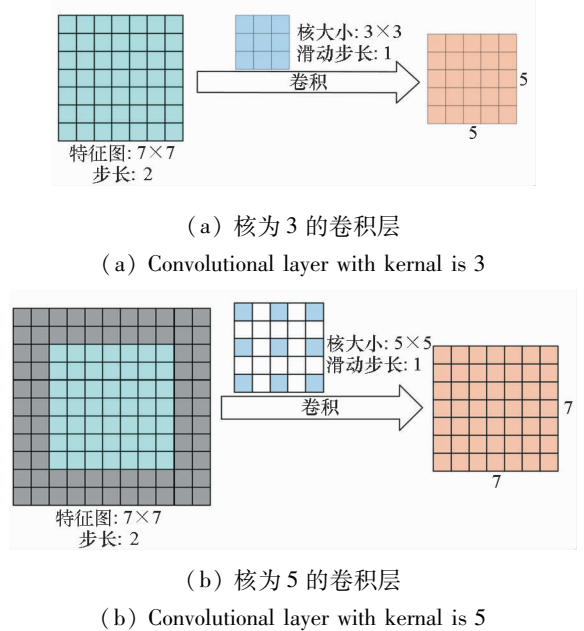


图 3 两种卷积层的对比

Fig. 3 Comparison of two convolution layers

1.3 多尺度金字塔池化全卷积网络

一般全卷积网络仅从特征角度出发对图像进行语义分割,难以对上下文关系建模,忽略了现实中的合理性,在对图像进行语义分割时存在错误的关系匹配问题。例如,在河面上将船归类于汽车,这种语义分割结果很明显是错误的。为了能够提取更准确的多尺度弱监督语义分割特征,需要对上下文关系进行强化和建模。本模型将采用金字塔池化的方法,在图像的特征图上进行不同尺度的特征抽取,根据不同尺度的特征对上下文关系进行学习。具体模型如图 4 所示。将三种不同尺度的金字塔池化模块用于全卷积网络中的池化层,其大小分别是 2×2 、 3×3 和 6×6 ,金字塔池模块中不同级别的输出包含不同大小的功能图。每一个特征图中的所有位置都会进行这三种尺度的金字塔池化,其中 2×2 大小的池化和原来全卷积网络中的池化设置相同,并且使用最大化池化,而其他两种尺度的池化是为了获得每一个

像素周围更大范围的上下文关系信息,因此需要采用平均池化且不丢失特征信息。通过将密集采样的池化操作和金字塔池化模块相结合,可以提取精细特征的同时提取上下文关系信息。

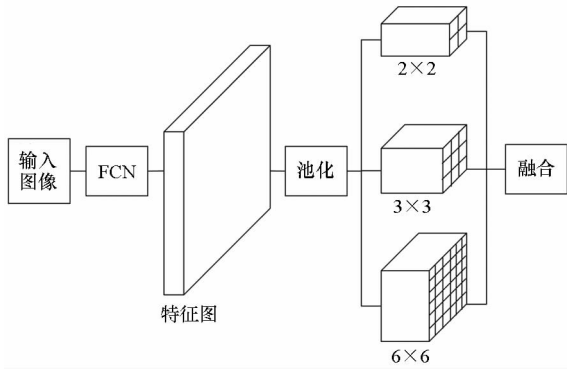


图 4 金字塔池化模型结构

Fig. 4 Pyramid pool model structure

2 基于 GrabCut 算法的像素标注训练模型

从对象边框标注图像中,如图 5(a)所示,可以获得图像像素值和对象的矩形框位置,这对图像语义分割是不够的,通常采用矩形转化方法将边框标注图像转换为像素级标注,如图 5(b)所示。从图中可以看出,矩形转化方法无法很好地区分矩形内部标注对象和背景图像,导致标注模型训练准确性降低。GrabCut 算法通过边界框来训练高斯混合模型的参数,区分前景和背景,进而获得像素级标注数据,如图 5(c)所示,是当前图像分割中常用的方法。图 5(d)为标签的真实值。

GrabCut 算法使用 RGB 色彩模型,用 k 个高斯分量的全协方差高斯混合模型来对图像中的像素建模,获得前景和背景像素。具体方法是:

图像的 Gibbs 能量定义为:

$$E(\alpha, k, \theta, z) = U(\alpha, k, \theta, z) + V(\alpha, z) \quad (1)$$

其中: α 表示像素是否为目标像素, $\alpha_i = 1$ 表示 i 像素是目标像素, $\alpha_i = 0$ 表示 i 像素不是目标像素; $k = \{k_1, \dots, k_n\}, k_n \in \{1, \dots, K\}$ 表示第 n 个像素对应的高斯分量; θ 表示高斯混合模型参数; z 表示具体的像素值。

色彩混合高斯模型 U ,使用像素归属于某类目标或背景的概率的负对数,表示将像素分配给某类目标或背景的惩罚大小:

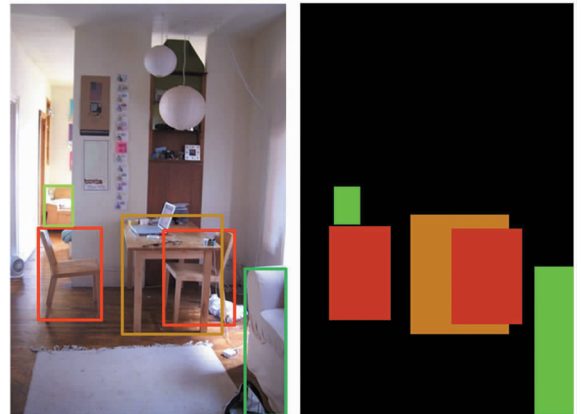
$$U(\alpha, k, \theta, z) = \sum_n \left\{ -\lg \pi(\alpha_n, k_n) + \frac{1}{2} \lg \det \Sigma(\alpha_n, k_n) \right\} + \sum_n \left\{ \frac{1}{2} [z_n - \mu(\alpha_n, k_n)]^T \Sigma(\alpha_n, k_n)^{-1} [z_n - \mu(\alpha_n, k_n)] \right\} \quad (2)$$

式中, $\pi(\cdot)$ 为混合加权系数, μ 为高斯分量的均值向量, Σ 为协方差矩阵。

边界能量项 V 表示两个相邻像素之间不连续情况下的惩罚大小:

$$V(\alpha, z) = \gamma \sum_{(m,n) \in C} [\alpha_m \neq \alpha_n] - \beta \|z_m - z_n\|^2 \quad (3)$$

其中, γ 和 β 为参数。用两个相邻像素的欧氏距离来衡量它们是否属于同一类。



(a) 边框标记数据
(a) Bounding box data

(b) 矩形转化数据
(b) Rectangle data



(c) GrabCut 转化数据
(c) GrabCut data

(d) 真实值
(d) Ground truth

图 5 对象边框标注数据转化像素级标注示例
Fig. 5 A sample graph of transforming bounding box annotations into pixel annotations

GrabCut 算法使用迭代过程求解能量函数最小值,进而获得最优参数。算法求解过程如图 6 所示,共分 7 步:第 1 步,对图像中像素进行初始化,分为框内目标像素和框外背景像素;第 2 步,初始化框内像素和框外像素的 α 值;第 3 步,计算均值和协方差,确定高斯混合模型参数;第 4 步,计算混合高斯模型分量,将目标像素值代入每个目标的高斯混合模型分量中,取最大分量概率作

为此像素的高斯分量,即 $k_n = \arg \min_{k_n} D_n(\alpha_n, k_n, \theta, z_n)$;第 5 步,学习三种类型的参数,计算各类样本的均值和协方差以及由像素所占的数量比值所确定的权重 $\theta = \arg \min_{\theta} U(\alpha, k, \theta, z)$, 获得高斯混合模型的三类参数;第 6 步,通过最小割算法和 Gibbs 能量项 $\min_{\{\alpha_n; n \in T_U\}} \min_k E(\alpha, k, \theta, z)$ 进行分割;第 7 步,重复迭代前 4 ~ 6 步,直至算法收敛才结束。

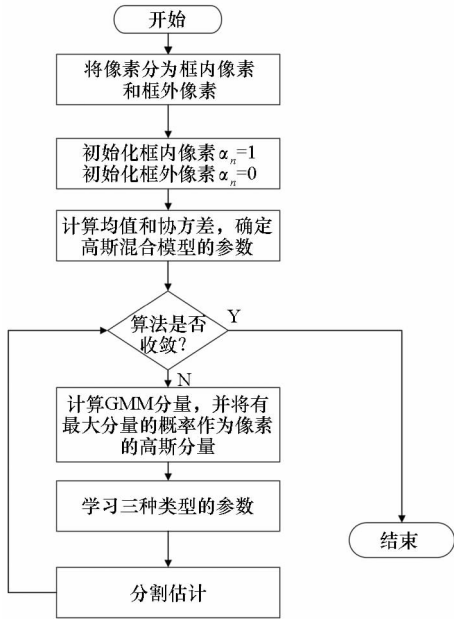


图 6 GrabCut 算法转化流程图

Fig. 6 Flow chart of GrabCut algorithm transformation

采用 GrabCut 算法转化得到的标注结果如图 5(c) 所示,这种方法能够得到较好的分割结果,但是在目标和背景较为相似的区域,则无法得到准确的结果。为了解决这个问题,本文结合全卷积网络进行训练和特征提取,提升图像分割准确性,模型的框架如图 7 所示。

3 实验及结果分析

3.1 实验数据及参数设置

本文使用公开的 PASCAL VOC2012 数据集作为实验数据,包括训练集和验证集两部分,含 20 个目标对象类和 1 个背景类。这 20 个目标对象类包括:人、鸟、猫、牛、羊、狗、马、飞机、自行车、船、公共汽车、小轿车、摩托车、火车、瓶子、椅子、餐桌、盆栽植物、沙发、电视。采用平均交叉联合率 (Intersection-Over-Union, IOU) 作为检测结果的主要评测指标。

本文使用 VGG-16 ImageNet 作为全卷积网络结构框架模型初始参数,感受野设为 128×128 。全卷积网络的参数通过小批量随机梯度下降方法

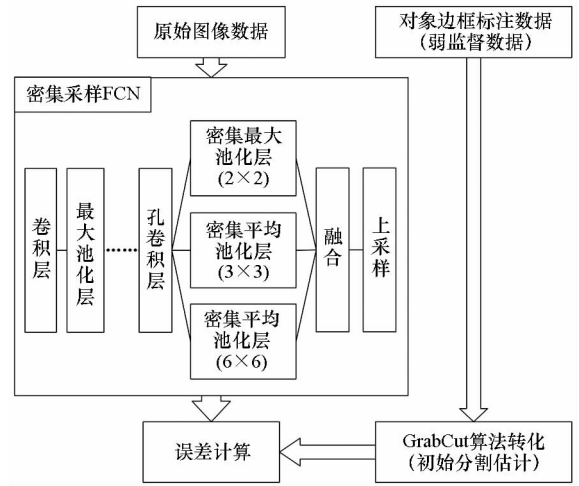


图 7 本文模型结构

Fig. 7 Structure of the proposed model

来更新,批量设为每次 20 ~ 30 幅图像,初始学习率设为 0.001,迭代数次之后更新学习率为上次的 0.1,冲量遗忘因子设为 0.9,衰减权重设为 0.000 5。使用 4 路 GPU 深度学习工作站进行模型训练,具体配置为: i7-5930k + 64 GB + 4 * NVIDIA TITAN X。

3.2 实验分析

本文模型实验将与 DET3-Proposed 模型^[24]、全矩形转化模型^[20]以及 Bbox-Seg 模型^[20]在 PASCAL VOC2012 公开数据集上进行比较,其比较结果如表 1 所示。

DET3-Proposed 模型是采用传统的方法进行对象边框标注的弱监督图像分割,通过一组投票器模型组合对像素进行分割,而其余几种模型都是基于神经网络进行特征提取和像素分类,通过表格中的数据可知,传统方法的 DET3-Proposed 模型在 PASCAL VOC2012 数据集上的 IOU 仅为 48.0%,而其余神经网络模型的 IOU 都显著地高于 DET3-Proposed 模型。

表 1 三种方法的 IOU 对比

Tab. 1 IOU comparison for three methods

模型	训练数量	验证集平均 IOU/%
DET3-Proposed		48.0
全矩形转化	10 582	52.5
Bbox-Seg	10 582	60.6
无金字塔 FCN-GrabCut[本文]	12 031	62.0
金字塔 FCN-GrabCut[本文]	12 031	64.7

对比全矩形转化模型、Bbox-Seg 模型以及本文两种模型,全矩形转化方法由于无法区分方框内的像素分类,在验证集上的 IOU 仅为 52.5%,而 Bbox-Seg 模型和基于无金字塔的 GrabCut 算法转化的像素级标注数据在验证集上的 IOU 分别为 60.6% 和 62.0%。这表明从对象标注数据转化得到的像素级标注数据越准确,分割的效果越好,选择不同的预处理转化模型对分割结果影响较大。另外,从表格中可以看出,本文的 GrabCut 算法模型比 Bbox-Seg 模型也要高出 4.1%。因此,综合对比可以发现,在对象边框标注数据的弱监督图像语义分割应用中,通过简单的方法对对象边框标注数据进行像素级转化可以取得较好的分割效果。对无金字塔 FCN-GrabCut 模型和金字塔 FCN-GrabCut 模型比较发现,通过金字塔池化对图像信息上下文关系进行提取和建模,能够更好地对物体间的关系进行正确分类,从而提高图像语义分割的准确率,在验证集上的 IOU 从 62% 提高到了 64.7%。

对象边框标注的弱监督图像语义分割模型的部分结果如图 8 所示。通过对比图 8 中的 Bbox-Seg 模型和本文模型,可以发现本文模型在对象边缘和对象分类的准确度上都有了明显的提升和改进。从图 8 中第二行的结果可以看出,Bbox-Seg 模型对于马的边缘分割效果很差。而本文模型相比起来能够更加精细地对马的腿部等细小部位进行更加准确的分割,同时也证明了本文模型能够通过密集采样的池化层提取更加精细的特征。从图 8 中第三行可以发现,Bbox-Seg 模型还存在很大的错误类别判断,相比起来,本文模型能够综合上下文特征对对象的类别进行更加准确的判别。

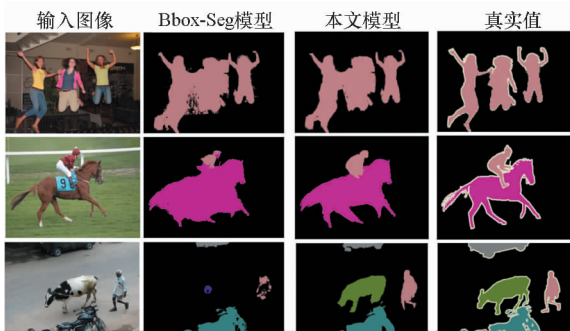


图 8 本文模型与 Bbox-Seg 模型的分割结果对比
Fig. 8 Comparison of segmentation result between the proposed model and Bbox-seg model

3.3 联合测试

为了进一步测试模型的性能,通过结合 PASCAL VOC2012 和 MS-COCO 数据集来测试本

文模型,使用来自 MS-COCO 数据集的训练集作为额外的训练数据对模型进行训练,本文模型和其他主流模型的结果如表 2 所示。

表 2 模型在 PASCAL VOC2012 和 MS-COCO 联合数据训练的测试结果

Tab. 2 Results for these methods trained on PASCAL VOC2012 and MS-COCO datasets

模型	训练数量	验证集平均 IOU/%
WSSL ^[20]	133 K (semi)	70.4
BoxSup ^[25]	133 K (semi)	71.0
$M \cap G +$ ^[26]	110 K (box)	69.9
金字塔 FCN-GrabCut [本文]	110 K (box)	70.3

在表 2 中,将结果与 PASCAL VOC2012 + MS-COCO 数据集上的半监督模型^[20,25]进行了比较,其中一些图像是像素级注释,而一些仅具有边界框注释:133 K (semi) = PASCAL VOC2012 10 K (像素级标注) + MS-COCO 123 K (边框标注),以及 110 K (box) = PASCAL VOC2012 10 K (边框标注) + MS-COCO 100 K (边框标注)。本文的弱监督方法(没有任何完全监督图像)达到 70.3%,与 WSSL 和 BoxSup 这两种半监督方法只差 0.1% 和 0.7%。这是因为 WSSL 和 BoxSup 这两种半监督方法的训练过程中包含一部分的像素级标注数据,而本文模型训练过程中不包含任何像素级标注数据。通过表 2 可以看出,本文弱监督模型基本能够达到其他半监督模型的效果。在没有任何像素级训练数据的模型下,本文模型比 $M \cap G +$ 提高了 0.4% (即 69.9% 对 70.3%)。

4 结论

针对像素级标注数据费时昂贵的缺点,本文研究了以对象边框标注数据为代表的弱监督模型下的图像语义分割方法。根据对象边框标注的数据中包含的类别大概位置信息进行初步分割,然后通过金字塔密集采样的全卷积网络对图像进行特征提取和更加精细的像素级分割。本文模型与 DET3-Proposed 模型、全矩形转化模型以及 Bbox-Seg 模型相比,在 PASCAL VOC2012 数据集上取得了更好的弱监督分割效果;同时本文还通过对比无金字塔密集采样 FCN-GrabCut 模型和金字塔密集采样 FCN-GrabCut 模型在数据集上的分割结果,证明了金字塔密集采样 FCN-GrabCut 模型能够提取和对图像中上下文的关系进行建模,分割

效果有明显的提高。

参考文献 (References)

- [1] 罗希平, 田捷, 诸葛婴, 等. 图像分割方法综述[J]. 模式识别与人工智能, 1999, 12(3): 300-312.
LUO Xiping, TIAN Jie, ZHUGE Ying, et al. A survey on image segmentation methods [J]. Pattern Recognition and Artificial Intelligence, 1999, 12(3): 300-312. (in Chinese)
- [2] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2009.
- [3] 周浩. 基于卷积神经网络和条件随机场的图像分割研究[D]. 长沙: 国防科技大学, 2016.
ZHOU Hao. Image segmentation based on convolution neural network and conditional random field [D]. Changsha: National University of Defense Technology, 2016. (in Chinese)
- [4] Lin T Y, Maire M, Belongie S, et al. Microsoft COCO: common objects in context [C]// Proceedings of European Conference on Computer Vision, 2014: 740-755.
- [5] Rother C, Kolmogorov V, Blake A. "GrabCut": interactive foreground extraction using iterated graph cuts [J]. ACM Transactions on Graphics, 2004, 23(3): 309-314.
- [6] Pandey M, Lazebnik S. Scene recognition and weakly supervised object localization with deformable part-based models [C]// Proceedings of IEEE International Conference on Computer Vision, 2011.
- [7] Vezhnevets A, Ferrari V, Buhmann J M. Weakly supervised semantic segmentation with a multi-image model [C]// Proceedings of IEEE International Conference on Computer Vision, 2011.
- [8] Verbeek J, Triggs B. Region classification with Markov field aspect models [C] // Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2007.
- [9] 梁恩辉, 周安然, 裴继红, 等. 基于能量图的海上红外图像目标分割方法[J]. 指挥信息系统与技术, 2018, 9(2): 79-84.
LIANG Enhui, ZHOU Anran, PEI Jihong, et al. Maritime infrared image target segmentation method based on energy map [J]. Command Information System and Technology, 2018, 9(2): 79-84. (in Chinese)
- [10] 陈鸿翔. 基于卷积神经网络的图像语义分割[D]. 杭州: 浙江大学, 2016.
CHEN Hongxiang. Semantic segmentation based on convolutional neural networks [D]. Hangzhou: Zhejiang University, 2016. (in Chinese)
- [11] 李春阳. 基于机器学习的道路区域识别方法研究[D]. 杭州: 浙江大学, 2016.
LI Chunyang. Study on road area detection based on machine learning methods [D]. Hangzhou: Zhejiang University, 2016. (in Chinese)
- [12] 刘丹, 刘学军, 王美珍. 一种多尺度 CNN 的图像语义分割算法[J]. 遥感信息, 2017, 32(1): 57-64.
LIU Dan, LIU Xuejun, WANG Meizhen. Semantic segmentation with multi-scale convolutional neural network [J]. Remote Sensing Information, 2017, 32(1): 57-64. (in Chinese)
- [13] Arbeláez P, Pont-Tuset J, Barron J, et al. Multiscale combinatorial grouping [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 328-335.
- [14] 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.
- [15] 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.
- [16] Zhou H, Han A, Yang H D, et al. Edge gradient feature and long distance dependency for image semantic segmentation [J]. IET Computer Vision, 2018, 13(1): 53-60.
- [17] Zhou H, Lei J, Wang F L, et al. Image semantic segmentation with finer edges and complete parts from bounding box annotations [J]. Journal of Electronic Imaging, 2019, 28(2): 023010.
- [18] Vezhnevets A, Ferrari V, Buhmann J M. Weakly supervised structured output learning for semantic segmentation [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2012: 845-852.
- [19] Xu J, Schwing A G, Urtasun R. Tell me what you see and I will show you where it is [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 3190-3197.
- [20] Papandreou G, Chen L C, Murphy K P, et al. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation [C]// Proceedings of the IEEE International Conference on Computer Vision, 2015: 1742-1750.
- [21] Holschneider M, Kronland-Martinet R, Morlet J, et al. A real-time algorithm for signal analysis with the help of the wavelet transform [J]. Wavelets: Time-Frequency Methods and Phase Space, 1989: 286-297.
- [22] Papandreou G, Kokkinos I, Savalle P A. Modeling local and global deformations in deep learning: epitomic convolution, multiple instance learning, and sliding window detection [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2015.
- [23] Chen L C, Papandreou G, Kokkinos I, et al. Semantic image segmentation with deep convolutional nets and fully connected CRFs [C]// Proceedings of Computer Vision and Pattern Recognition, 2014.
- [24] Xia W, Domokos C, Dong J, et al. Semantic segmentation without annotating segments [C]// Proceedings of the IEEE International Conference on Computer Vision, 2013: 2176-2183.
- [25] Dai J F, He K M, Sun J. Boxsup: exploiting bounding boxes to supervise convolutional networks for semantic segmentation [C]// Proceedings of the IEEE International Conference on Computer Vision, 2015, 1: 1635-1643.
- [26] Khoreva A, Benenson R, Hosang J, et al. Simple does it: weakly supervised instance and semantic segmentation [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2017, 1: 1665-1674.