

门户 商业智能 深度学习 查看内容

何恺明团队推出Mask^X R-CNN，将实例分割扩展到3000类

2017-12-4 10:32 | 发布者: 炼数成金\_小数 | 查看: 13177 | 评论: 0 | 来自: AI科技大本营

**摘要:** 今年10月，何恺明的论文“Mask R-CNN”摘下ICCV 2017的最佳论文奖（Best Paper Award），如今，何恺明团队在Mask R-CNN的基础上更进一步，推出了（以下称Mask^X R-CNN）。这篇论文的第一作者是伯克利大学的在读博士 ...

网络 数据库 工具 模型 基础

Learning to Segment Every Thing

Ronghang Hu<sup>1,2,\*</sup> Piotr Dollár<sup>2</sup> Kaiming He<sup>2</sup> Trevor Darrell<sup>1</sup> Ross Girshick<sup>2</sup>  
<sup>1</sup>BAIR, UC Berkeley <sup>2</sup>Facebook AI Research (FAIR)

今年10月，何恺明的论文“Mask R-CNN”摘下ICCV 2017的较佳论文奖（Best Paper Award），如今，何恺明团队在Mask R-CNN的基础上更进一步，推出了（以下称Mask^X R-CNN）。

这篇论文的第一作者是伯克利大学的在读博士生胡戎航（清华大学毕业），标题非常霸气，叫是“Learning to Segment Every Thing”。从标题上可以看出，这是一篇在实例分割问题(instance segmentation)中研究扩展分割物体类别数量的论文。

由于现有的目标实例分割（object instance segmentation）方法要求所有训练实例都必须标记有分割掩码（segmentation mask），使得注释新类别的成本十分昂贵，而且还将实例分割模型限制在约100个详细注释的类别。本论文提出了一种全新的偏监督（partially supervised）训练方式以及一个新的权重传递函数（weight transfer function），用大量的类别（所有类别都标有边界框注释（box annotations），但只有一小部分类别标有掩码注释）训练实例分割模型。

论文作者表示，他们成功使用Visual Genome数据库中的边界框注释以及COCO数据库中80个类别的掩码注释，训练Mask R-CNN检测并分割3000个视觉概念。此外，该论文还首次探究了如何让实例分割模型可以全面地理解视觉世界。

以下是论文简介，enjoy！

介绍

热门频道

大数据	商业智能
量化投资	科学探索
	创业



**校长招入室弟子**

炼数成金个性化教育产品“专才计划”  
上线，导师制私人订制尊贵服务

即将开课

- 机器学习及其matlab实现—从基础到实践
- 大数据的Linux基础（第19期）
- 高可用架构设计与实践（第九期）
- Python突击—从入门到精通到项目实战（第
- Watson Analytics数据分析应用实战公开课
- Oracle 12C RAC集群原理与管理实战（第11
- 区块链技术从入门到精通（第四期）
- Python自然语言分析（第七期）
- Scala从基础到开发实战（第八期）
- 左飞的机器学习十八般算法武艺详解（第二
- OpenCV计算机视觉产品实战（第二期）
- 【免费公开课】Julia快速数据分析（第二
- 老板说术之玩转数据展示（第十期）
- MySQL性能优化最佳实践（第三期）
- 金融市场基础（第五期）
- Zookeeper分布式系统开发实战（第七期）
- python web框架企业实战详解（第14期）
- 知识图谱实战（第一期）
- 【免费公开课】验模神器AMOS（第四期）
- zabbix企业级实践（第四期）
- Spark大数据平台应用实战（第三期）
- 机器学习（第21期）
- 深度学习PostgreSQL（第七期）
- 软件架构必备基础（第二期）
- ROS机器人操作系统实战（第二期）
- 【免费公开课】企业级Hadoop大数据平台
- Oracle DBA从小白到入职实战应用（第六
- Python金融投资分析实践（第六期）
- 大数据算法导论（第14期）
- Go语言实战编程（第三期）
- 深度学习框架Tensorflow学习与应用（第五
- 深入探索甲骨文云计算（第一期）
- 【免费公开课】Java经验谈（第三期）
- R语言数据分析、展现与实例（第30期）
- 深入浅出Spring（第五期）

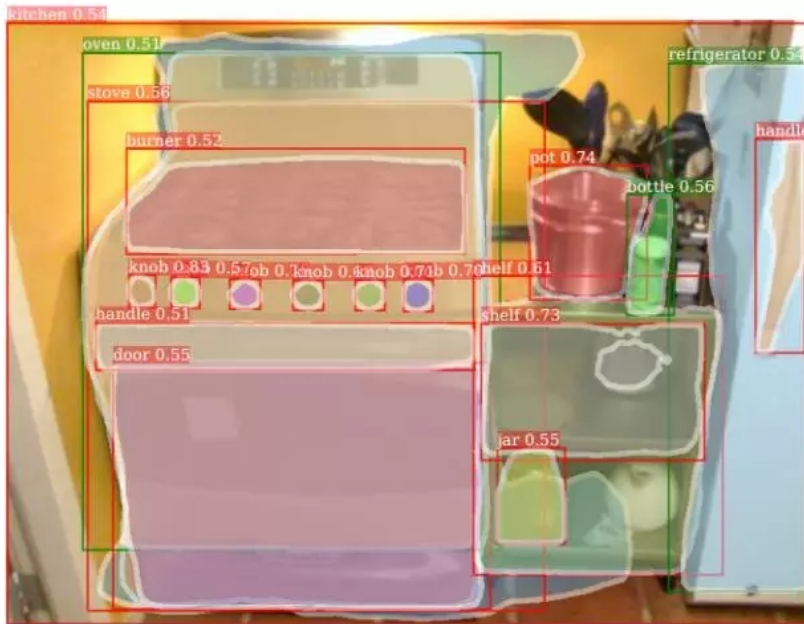


图1. 我们通过偏监督方法来探索如何训练实例分割模型：在训练时，一个类的子集（绿色框）具有实例掩码的注释；剩余的类（红色框）只有边界框注释。该图显示了我们的模型在Visual Genome数据集上训练了3000个类后的输出，这个过程中仅使用了COCO数据集中80个类别的掩码注释。

目标检测器已经变得非常准确了，并拥有了很重要的新功能。其中最令人兴奋的功能是能够为每个检测到的对象预测前景分割掩码，这个任务我们称之为实例分割。在实际应用中，传统的实例分割系统往往只能对包含约100个对象类别的数据集起作用，而这只是大千世界中的沧海一粟。

造成这种现象的一个主要原因是，现有最领先的实例分割算法都需要强监督学习，而这样的监督学习有很大的限制，并且采集新类别图片的代价也是十分昂贵的。相比之下，带边界框注释的图片则会更丰富和也更便宜。这就引出了一个问题：在不是所有类别都标有完整实例分割注释的前提下，是否有可能训练出高质量的实例分割模型为此，本文介绍了一种新的偏监督实例分割任务，并提出了一种新的迁移学习的方法来完成它。

我们制定的基于偏监督学习的实例分割任务如下：

- （1）给定一组感兴趣的类别和一个有实例掩码注释的小的子集，而其他类别只有边界框注释；
- （2）实例分割算法可以利用这个数据来拟合一个模型，该模型可以分割所感兴趣的集合中的所有对象类别的实例。由于训练数据是完整注释数据（带掩码的示例）和弱注释数据（仅带框的示例）的混合，因此我们将该任务称为偏监督任务。

本文所提出的偏监督学习样例流程的主要好处是它允许我们通过利用两种类型的现有数据集来构建一个大规模的实例分割模型：那些在大量的类上使用边界框注释的数据集，比如Visual Genome，以及那些在少数类别上使用实例掩码注释的，例如COCO数据集。正如我们接下来将要展示的那样，这使得我们能够较先进的实例分割方法扩展到数千个类别，这对于在现实世界中部署实例分割是非常重要的。

为了解决偏监督的实例分割问题，我们提出了一种基于Mask R-CNN的新型迁移学习的方法。Mask R-CNN非常适合我们的任务，因为它将实例分割问题分解为了目标的边界框检测和掩码预测两个子任务。这些子任务是由专门的网络“头部（heads）”共同训练的。我们的方法背后的直觉是，一旦训练完成了，边界框头部（the bounding

- Python数据分析案例实战（第六期）
- DevSecOps安全交付应用实战（第三期）
- 机器读心术之神经网络与深度学习（第九期）
- Spark企业级大数据项目实战（第一期）
- 【免费公开课】R七种武器之网络爬虫
- Python数据分析（第11期）
- Node.js项目实战：从编写代码到服务器部署
- 金融的人工智能革命（第二期）
- 黄美灵的Spark ML机器学习实战（第二期）
- 开源计算机视觉库OpenCV从入门到应用
- 人工智能前沿系列之生成式对抗网络（第五期）
- 深入浅出设计模式（第一期）
- 【免费公开课】数据陷阱解读（第26期）
- 深入浅出Oracle（第七期）
- 深度学习框架Keras学习与应用（第二期）
- Excel数据分析师突击——从入门到精通到项目
- Python机器学习Kaggle案例实战（第四期）
- 大数据必知的java基础（第八期）
- 机器读心术之文本挖掘与自然语言处理（第一期）
- django打造电商项目（第一期）
- 让服务飞起来：实时计算及其应用（第二期）
- ELK Stack及Solr企业级搜索引擎实战（第七期）
- Python机器学习（第五期）



#### 热门文章

- AWS推出MXNet模型服务器
- TensorFlow的多平台基准测试
- 深度学习与神经科学相遇（三）[译]
- 深度学习与神经科学相遇（二）[译]
- 深度学习与神经科学相遇（一）[译]
- 卷积神经网络的复杂度分析
- 使用TensorFlow和TensorBoard从零开始构建
- 谷歌推出开源工具DeepVariant，用深度学习
- 那么多GAN哪个好？谷歌大脑泼来冷水：都
- 如何构建高可读性和高可重用的TensorFlow



box head ) 参数编码嵌入到每个对象类别，使该类别的视觉信息转移到偏监督掩码头部 ( the partially supervised mask head ) 参数上。

为了让这个直觉具象化，我们设计了一个参数化的权重传递函数，该函数被训练成根据图片类别的边界框检测参数来预测类别的实例分割参数。权重传递函数可以在Mask R-CNN中使用带有掩码注释的类作为监督学习的数据来进行端到端的训练。在推理时，权重传递函数用于预测每个类别的实例分割参数，从而使模型能够分割所有目标的类别，包括在训练时没有掩码注释的目标类别。

我们在两种不同的设置环境中评估了我们的方法。首先，为了在数据集上建立包含高质量的注释和评估指标的定量的结果，我们使用了COCO数据集来模拟偏监督的实例分割任务。具体地说，我们将COCO数据集所有的类别划分为带有掩码注释的子集和一个只提供实例分割系统边界框注释的子集。由于COCO数据集仅涉及少量（80类）的语义分离很好的类，因此定量评估的结果是准确可靠的。实验结果表明，我们的方法得到了比该任务基准线高很多的结果，在没有采用训练用的掩码的情况下，掩码的AP相对增幅高达40%。

在第二种设置中，我们使用包含3000个类别的Visual Genome（VG）数据集进行了一次大规模的实例分割模型训练。VG数据集包含大量目标类别的边界框注释，但是由于许多类别在语义上重叠（例如，近义词）并且注释并不详尽，使得模型的较精确度和召回率难以衡量。而且，VG数据集不是用实例掩码来标注的。作为替代，我们采用VG数据集来提供大规模实例分割模型的定性输出。我们模型的输出如图1和5所示：

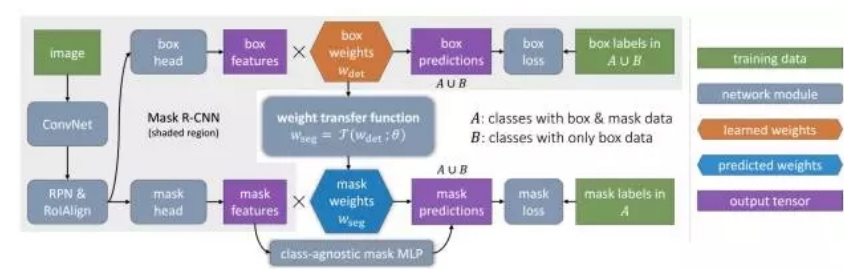


图2.我们提出的Mask<sup>X</sup> R-CNN方法的详细说明。Mask<sup>X</sup> R-CNN不是直接学习掩码预测参数  $w_{seg}$ ，而是使用学习权重传递函数  $T$  从其对应的检测参数  $w_{det}$  中预测出类别的分割参数。在训练阶段，函数  $T$  只需要数据集  $A$  中每个类别的掩码数据，它就可以在测试阶段对数据集  $A \cup B$ （并集）中的所有类进行参数学习。我们还使用了一个互补的全连接的多层感知器（MLP）来增加了掩码头部（mask head）的内容。

学习分割一切物体

假设集合  $C$  为一组对象类别（例如 ‘things’ ），我们要用这些类别来训练一个实例分割模型。大多数现有方法假设  $C$  中所有的训练实例都标有掩码注释。我们放宽了这个要求，只假设  $C = A \cup B$ ，也就是说：集合  $A$  中的类别实例都标有掩码注释，集合  $B$  中的类别实例只有边界框注释。由于集合  $B$  中的类别只带有关于目标任务（实例分割）的弱标签，我们将使用组合强标签（strong labels）和弱标签（weak labels）的类别来训练模型的问题称为偏监督学习问题。

注意：我们可以轻易地将实例的掩码注释转换为边界框注释，因此我们假设  $A$  中的类别也带有边界框注释。由于Mask RCNN这样的实例分割模型都带有一个边界框检测器和一个掩码预测器，我们提出的

利用权重传递函数预测掩码

我们的方法建立在Mask R-CNN上，因为Mask R-CNN实例分割模型不仅结构简单，而且可以实现非常优秀的结果。简单来说，我们可以将Mask R-CNN看作为添加有一个掩码预测分支（小型全卷积网络）的Faster R-CNN边界框检测模型。在预测阶段，



模型用掩码分支处理每个检测对象，为每个对象预测一个实例级别的前景分割掩码。在训练阶段，并行训练掩码分支和Faster R-CNN中的标准边界框检测器。

在Mask R-CNN中，边界框分支的最后一层以及掩码分支的最后一层均包含对每个类别执行边界框分类和实例掩码预测任务时所用的类别参数。我们选择的方法是：使用一个通用的权重传递函数，根据某一类别的边界框参数预测它的掩码参数，这个函数可以作为模型的组部分与模型一起进行训练；而不是分别学习某一类别的边界框参数和掩码参数。

给定一类别  $c$ ，假设  $w_{\text{det}}^c$  为类别  $c$  在边界框检测器最后一层上的目标检测权重， $w_{\text{seg}}^c$  为类别  $c$  在mask分支上的mask权重。我们使用一个通用的权重预测函数  $\mathcal{T}(\cdot)$  将  $w_{\text{seg}}^c$  参数化，而不是将  $w_{\text{seg}}^c$  直接作为参数。

$$w_{\text{seg}}^c = \mathcal{T}(w_{\text{det}}^c; \theta), \quad (1)$$

1. 其中  $\theta$  为类别不可知的学习参数。

同一传递函数  $\mathcal{T}(\cdot)$  可应用于任何类别  $c$ ，因此选择的  $\theta$  值应使  $\mathcal{T}(\cdot)$  可以泛化到训练期间掩码未被观察到的任何类别。我们预计这种泛化是可能实现的，因为检测权重  $w_{\text{det}}^c$  可以被视为基于外观的类别视觉嵌入。

传递函数  $\mathcal{T}(\cdot)$  可以作为一个小型的全卷积神经网络。图2展示了权重传递函数与Mask R-CNN结合形成Mask^X R-CNN的过程。注意：边界框识别器包含两种类型的检测权重：RoI分类权重  $w_{\text{cls}}^c$  以及边界框回归权重（regression weights） $w_{\text{box}}^c$ 。

在试验时我们可以使用一种类型的检测权重（即：

$$w_{\text{det}}^c = w_{\text{cls}}^c \text{ OR } w_{\text{det}}^c = w_{\text{box}}^c,$$

也可以使用两类权重的级联（concatenation）（即： $w_{\text{det}}^c = [w_{\text{cls}}^c, w_{\text{box}}^c]$ ）。

基准：类别不可知的掩码预测

DeepMask证明了：训练深度学习模型执行类别不可知掩码预测任务（不考虑类别而预测目标掩码）是不可能实现的。对于掩码质量稍微损失的Mask R-CNN而言也是这样。在其他试验中，如果类别不可知模型经过训练后可以预测COCO类别中一个类别子集的掩码，那么这些预测值在预测阶段（inference time）就可以泛化到其他60个COCO类别上。依据这些结论，我们用带有一个不可知FCN掩码预测器的Mask R-CNN作为基准。事实证明，这是一个非常优秀的基准。接下来，我们提出了一个可以用于改进该基准和权重传递函数的扩展。

扩展：FCN+MLP 掩码预测器

两种类型的掩码检测器都可用于Mask RCNN：

（1）FCN预测器使用一个全卷积网络预测  $M \times M$  掩码；

（2）MLP预测器使用一个多层感知器预测掩码，这个多层感知器是由全连接层构成的，类似于DeepMask。在Mask R-CNN中，FCN预测器获得的掩码平均精度（AP）。但是，这两种预测器可以互补。直观看来，MLP 掩码预测器可以更好地捕捉物体的“主要特征”，而FCN 掩码预测器可以更好地捕捉物体的细节（例如：物体的边界）。根据这一观察，我们通过将基准类别不可知FCN预测器和权重传递函数（使用的是一个FCN预测器）与类别不可知MLP 掩码预测器作出的预测整合在一起，以此来改进前两者的表现。试验结果表明，这个扩展可以改进基准和权重传递方法。

当对K个类别的class-agnostic 和class-specific 掩码预测进行整合时，这两个预测值都被添加到最终的 $K \times M \times M$ 输出中，其中class-agnostic掩码预测值 ( $1 \times M \times M$ ) 被分割了K次并添加到每个类别中。然后， $K \times M \times M$  掩码预测值经一个sigmoid单元处理后转化为每个类别的掩码概率值，其大小被调整为实际的边界框大小，并作为边界框最终的实例掩码。在训练期间，我们对 $K \times M \times M$  掩码概率应用了二元交叉熵损失函数。

在COCO数据集上进行试验

method	voc $\rightarrow$ non-voc		non-voc $\rightarrow$ voc	
	AP on B	AP on A	AP on B	AP on A
transfer w/ randn	15.4	35.2	19.9	31.1
transfer w/ GloVe [30]	17.3	35.2	21.9	31.1
transfer w/ cls	18.1	35.1	25.2	31.1
transfer w/ box	19.8	35.2	25.7	31.1
transfer w/ cls+box	<b>20.2</b>	35.2	<b>26.0</b>	31.2
class-agnostic (baseline)	14.2	34.4	21.5	30.7
fully supervised (oracle)	30.7	35.0	35.0	30.7

(a) Ablation on input to T .

method	voc $\rightarrow$ non-voc		non-voc $\rightarrow$ voc	
	AP on B	AP on A	AP on B	AP on A
transfer w/ 1-layer, none	19.2	35.2	25.3	31.2
transfer w/ 2-layer, ReLU	19.7	35.3	25.1	31.1
transfer w/ 2-layer, LeakyReLU	<b>20.2</b>	35.2	<b>26.0</b>	31.2
transfer w/ 3-layer, ReLU	19.3	35.2	<b>26.0</b>	31.1
transfer w/ 3-layer, LeakyReLU	18.9	35.2	25.5	31.1

(b) Ablation on the structure of T .

method	voc $\rightarrow$ non-voc		non-voc $\rightarrow$ voc	
	AP on B	AP on A	AP on B	AP on A
class-agnostic	14.2	34.4	21.5	30.7
class-agnostic+MLP	17.1	35.1	22.8	31.3
transfer	20.2	35.2	26.0	31.2
transfer+MLP	<b>21.3</b>	35.4	<b>26.6</b>	31.4

(c) Impact of the MLP mask branch.

method	training	stop grad	voc $\rightarrow$ non-voc		non-voc $\rightarrow$ voc	
		on $w_{det}$	AP on B	AP on A	AP on B	AP on A
class-agnostic	sw	n/a	14.2	34.4	21.5	30.7
transfer	sw	n/a	20.2	35.2	26.0	31.2
class-agnostic	e2e	n/a	19.2	36.8	23.9	32.5
transfer	e2e		20.2	37.7	24.8	33.2
transfer	e2e	✓	<b>22.2</b>	37.6	<b>27.6</b>	33.1

(d) Ablation on the training strategy.

表1. 方法的简化测试。(a,b,c)中的结果基于分阶段训练，我们在(d)中研究了端对端训练的影响。我们还用COCO数据集val2017评估了掩码的AP值，该数据集包含20个PASCAL VOC类别 (voc) 和60个其他类别 (非voc)。用强监督数据集A训练模型得出的结果用灰色字体表示。

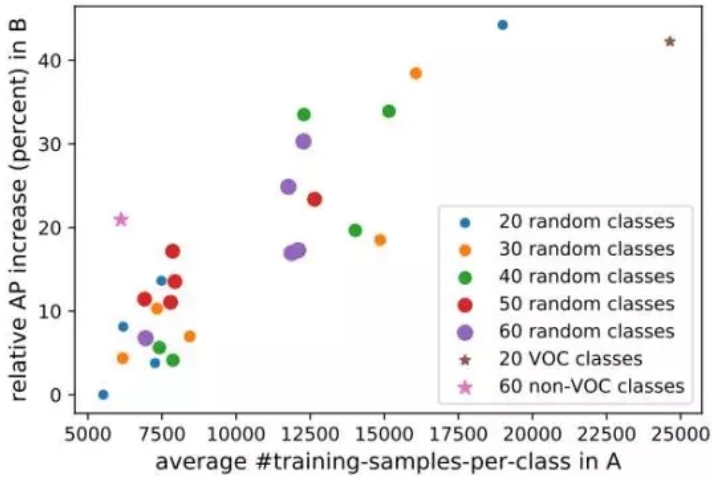


图3. 每个点代表我们的方法在COCO数据集随机划分的类别集A和类别集B上的表现。我们类别集A的类别数设定为20至60，然后绘制出类别集B中类别（无掩码注释）的掩码AP对比类别集A中每个类别的掩码注释数的平均值的变化图。

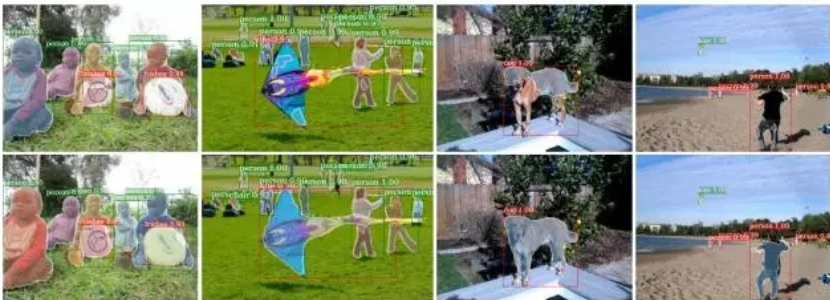


图4.类别不可知基准方法（第一行）与Mask^X R-CNN方法（第二行）的掩码预测对比。绿色方框代表类别集A中的类别，红色方框代表类别B中的类别。左侧两栏为A = {voc}，右侧两栏为A = { non-voc }。

backbone	method	voc → non-voc: test on B = {non-voc}						non-voc → voc: test on B = {voc}					
		AP	AP <sub>50</sub>	AP <sub>75</sub>	AP <sub>S</sub>	AP <sub>M</sub>	AP <sub>L</sub>	AP	AP <sub>50</sub>	AP <sub>75</sub>	AP <sub>S</sub>	AP <sub>M</sub>	AP <sub>L</sub>
R-50-FPN	class-agnostic (baseline)	19.2	36.4	18.4	11.5	23.3	24.4	23.9	42.9	23.5	11.6	24.3	33.7
	Mask <sup>X</sup> R-CNN (ours)	<b>23.7</b>	<b>43.1</b>	<b>23.5</b>	<b>12.4</b>	<b>27.6</b>	<b>32.9</b>	<b>28.9</b>	<b>52.2</b>	<b>28.6</b>	<b>12.1</b>	<b>29.0</b>	<b>40.6</b>
	fully supervised (oracle)	33.0	53.7	35.0	15.1	37.0	49.9	37.5	63.1	38.9	15.1	36.0	53.1
R-101-FPN	class-agnostic (baseline)	18.5	34.8	18.1	11.3	23.4	21.7	24.7	43.5	24.9	11.4	25.7	35.1
	Mask <sup>X</sup> R-CNN (ours)	<b>23.8</b>	<b>42.9</b>	<b>23.5</b>	<b>12.7</b>	<b>28.1</b>	<b>33.5</b>	<b>29.5</b>	<b>52.4</b>	<b>29.7</b>	<b>13.4</b>	<b>30.2</b>	<b>41.0</b>
	fully supervised (oracle)	34.4	55.2	36.3	15.5	39.0	52.6	39.1	64.5	41.4	16.3	38.1	55.1

表2. Mask^X R-CNN的端对端训练。如表1所示，我们用 ‘cls+box, 2-layer, LeakyReLU’ 作为Mask^X R-CNN的实现结构，然后添加了MLP 掩码分支（ ‘transfer+MLP’ ），按照相同的评估方案进行评估。我们还报告了AP50 和 AP75（以0.5 和 0.75为IoU临界值分别估计出平均准确度），小型（APS）、中型（APM）的大型（APL）物体的AP（平均准确度）。在ResNet-50-FPN 和 ResNet-101-FPN 基础网络无掩码训练数据的前提下，使用类别集B中的类别进行评估，得出的结论是：我们方法的表现远远优于基准。

表2对比了完整的Mask^X R-CNN的方法。它的表现大幅超越基准方法的表现（掩码平均准确度相对增加了20%多）。除了ResNet-50-FPN，我们还将ResNet-101 FPN作为基础网络进行了试验，如表2下半部分所示。我们在ResNet-50-FPN上发现的趋势也出现在 ResNet-101-FPN上，这表明基础网络并不会影响试验结果。图4给出了类别不可知基准方法和我们的方法的掩码预测示例。

大规模的实例分割

到目前为止，我们已经实验过了一个我们真正目标的模拟版本：即用宽阔的视觉理解来训练大规模的实例分割模型。我们相信这个目标代表了一个激动人心的视觉识别研究的新方向，要完成它，可能需从偏监督中学习到某种形式。为了实现这一目标，我们使

用Visual Genome ( VG ) 数据集的边界框和COCO数据集的实例掩码[22]两种数据，用偏监督任务的学习方式训练了一个大规模的Mask<sup>X</sup> R-CNN模型。VG数据集包含了108077张图像，以及超过7000类的用目标边界框注释（但不包括掩码）的同义词集。

为了训练我们的模型，我们选择了3000个最常见的同义词数据集作为我们的数据集A和数据集B来实现实例分割的任务，该数据集AB覆盖了COCO数据集中的所有80个类别。由于VG数据集图像与COCO数据集重叠较大，因此在用VG数据集训练时，我们将所有不在COCO val2017数据集中的图像作为训练集，并将其余的VG图像作为验证集。我们将VG数据集中所有与COCO数据集重叠的80个类别作为我们的带掩码的数据集A，VG数据集中其余的2920个类别作为我们的数据集B，因为它们只有边界框的注释。

训练。我们使用阶段式训练策略来训练我们的大规模Mask<sup>X</sup> R-CNN模型。

具体来说，我们使用ResNet-101-FPN网络结构作为我们的骨干网络，按照4.1节中所提到的超参数训练了一个Faster R-CNN模型来检测VG数据集中的3000个类别。然后，在第二阶段，我们用'cls + box, 2-layer, LeakyReLU'这几个结构构造了权重传递函数T，接着用函数T和类别未知的MLP掩码预测（即，'transfer + MLP'）组成了掩码头部(the mask head)。训练掩码头部的数据集是用包含80个类别的COCO数据集(数据集A)的子集，这些子集采用了COCO数据集train2017中拆分出来的掩码注释。

定性的结果。掩码AP（平均精确度）很难在VG数据集上进行计算，因为它只包含了目标边界框的注释。因此，我们使用我们的权重传递函数将结果可视化，以了解模型在A和B一起组成的数据集中所有3000个类别上训练完的性能。图5显示了在验证集上一些掩码预测的例子，可以看出，在那些与COCO数据集类别不重叠的VG数据集类别（红框中显示的是数据集B）上，我们的模型预测了比较合理的掩码。

这个可视化的结果显示了我们大规模实例分割模型的几个有趣的特性。首先，它已经学会了检测一些抽象的概念，如阴影和路径。这些往往是难以分割的。其次，能够简单地从小数据集获得前3000个同义词，一些概念比“事物(thing)”更像“东西(stuff)”。例如，该模型对孤立的树进行了合理的分割，但当检测到的“树”更像森林时，该模型往往会分割失败。最后，检测器在分割整个物体和部分物体（例如电车的窗户或冰箱的手柄）时做的是比较合理的。与在COCO数据集80个类别上训练的检测器相比，这些结果说明了现在训练的实例分割系统有很好的潜力去识别和分割数千个概念。





图5. Mask^X R-CNN模型在数据集Visual Genome上进行3000个类的掩码预测的示例。绿色框是与COCO（带掩码训练数据的数据集A）重叠的80个类，而红色框是剩余的2920个不在COCO数据集中的类（没有掩码训练数据的数据集B）。可以看出，我们的模型在数据集B中的许多类上生成了一个比较合理的掩码预测。

结论

本文针对大规模实例分割的问题，构造了一个偏监督的学习样例流程。其中只有一部分类在训练时具有实例掩码数据，而其余部分具有边界框注释。我们提出了一种新的迁移学习的方法，其中是用训练好的权重转移函数来预测：如何根据学习的参数来检测每个类别的边界框。在COCO数据集上的实验结果表明，在没有掩码训练数据的情况下，我们的方法大大提高了掩码预测的泛化能力。通过采用该方法，我们在Visual Genome数据集中建立了超过3000个类的大规模实例分割模型。我们得到的结果非常鼓舞人心，这足以说明我们开创了一个令人兴奋的新的实例分割的研究方向。研究人员还提到了一个非常具有挑战性的问题，在没有监督学习的情况下，如何将实例分割扩展到数千个类别。应该可以有很多方式来改进此方法去解决这样的问题。

论文链接：<https://arxiv.org/pdf/1711.10370.pdf>

欢迎加入本站公开兴趣群

商业智能与数据分析群

兴趣范围包括各种让数据产生价值的办法，实际应用案例分享与讨论，分析工具，ETL工具，数据仓库，数据挖掘工具，报表系统等全方位知识

QQ群：81035754



邀请

分享

收藏

分享到新浪微博

上一篇：最新StarGAN对抗生成网络实现多领域图像变换  
下一篇：如何构建高可读性和高可重用的 TensorFlow 模型

相关阅读

• OpenCV计算机视觉产品实战



• 《开启智慧眼-深度玩转计算机视觉与机器认

• 知识图谱实战

最新评论

评论

你猜我昵称叫啥\_

还可以输入140字



顺便说点什么吧.....

表情 ☒ 同步到微博

0条评论

还没有人评论过，赶快抢沙发吧！

关于我们

[关于我们](#)

[讲师招募](#)

[校园大使](#)

[版权声明](#)

新手指南

[教育模式](#)

[选课流程](#)

[学费返还](#)

[奖学金激励](#)

企业合作

[企业会员](#)

[内训合作](#)

[媒体合作](#)

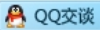
[代理合作](#)

联系我们

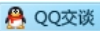
电话：4008-010-006

邮箱：kefu@dataguru.cn

客服QQ:



售后QQ:



订阅号



小程序

[关于我们](#) [手机版](#) [友情链接](#) [站点统计](#) [文本模式](#) [小游戏](#)

版权所有 广州市皓岚信息技术有限公司 合作伙伴 中山大学海量数据与云计算研究中心 粤ICP备08028958号

CopyRight 2011-2015 dataguru.cn All Right Reserved.