

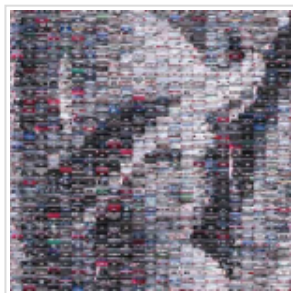
## cv\_family\_z的博客

目录视图

摘要视图

RSS 订阅

## 个人资料



cv\_family\_z



访问：370027次

积分：5398

等级：BLOG &gt; 6

排名：第5219名

原创：175篇 转载：3篇

译文：2篇 评论：197条

## 文章搜索

异步赠书：9月重磅新书升级，本本经典

程序员9月书讯

每周荐书：ES6、虚拟现实、物联网（评论送书）

## 图像分割“Instance-aware Semantic Segmentation via Multi-task Network Cas

2017-04-18 16:18

636人阅读

评论

分类：目标检测（41） 图像分割（13）

版权声明：本文为博主原创文章，未经博主允许不得转载。

模型包含三个网络：实例分辨，掩码计算，目标归类。三个网络形成级联的结构。运行时间上，使用图片需360ms。在MS COCO 2015分割比赛中获取第一名。

将instance-aware语义分割分成三个子任务：1.实例分辨，使用类别无关的bbox回归网络，对每个实例的像素级掩码；3.目标分类，预测每个掩码级别实例的类别。论文提出的多任务学习时级联结构，下一阶段依靠

关闭

## 文章分类

目标检测 (42)  
 模式识别 (3)  
 深度学习 (91)  
 目标识别 (7)  
 车型识别 (6)  
 行人检测 (13)  
 人脸识别 (20)  
 模型测试 (2)  
 3D-识别 (3)  
 行人检索 (12)  
 开源代码 (1)  
 行人属性 (5)  
 ZJ (78)  
 人脸检测 (3)  
 CVPR 2016 (37)  
 代码调试 (2)  
 目标跟踪 (4)  
 ICML2016 (2)  
 图像分割 (14)  
 ECCV2016 (3)  
 CNN网络压缩 (5)  
 杂项 (2)  
 车辆检索 (1)  
 车辆计数 (2)

## 文章存档

2017年09月 (6)

上一阶段的输出，如下所示：

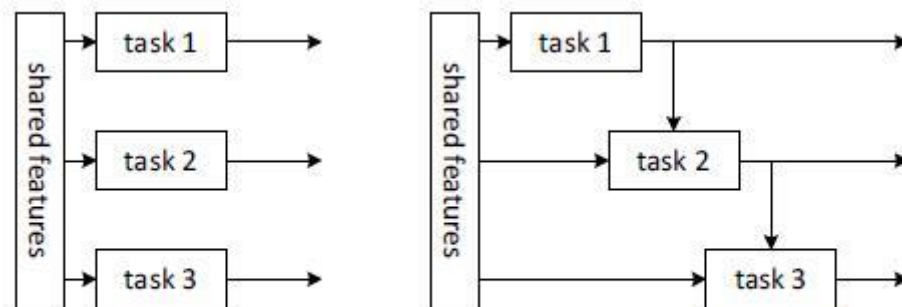


Figure 1. Illustrations of common multi-task learning (left) and our multi-task cascade (right).

## 多任务网络级联

三个阶段共享卷积特征，每个阶段包含一个损失项，但后面一阶段的损失依靠上一阶段的损失。

## 1. Box-级实例回归

网络结构与损失函数与RPN类似，在共享特征后，衔接一个3\*3的卷积层降维，然后接两个1\*1的卷积层，分别用于bbox回归和是否目标分类。RPN损失为：

$$L_1 = L_1(B(\Theta))$$

其中 $\Theta$ 表示所有需要优化的网络参数， $B$ 是此层网络的输出，即bbox。

## 2. Mask-级实例回归

给定阶段1的bbox，使用RoI池化提取固定长度的特征，之后衔接两个全连接层，第一个fc层将维度降为256，第二个fc层回归pixel-wise掩码，有 $m^2$ 个输出，对应相应大小的掩码。第二层的损失函数为：

$$L_2 = L_2(M(\Theta) | B(\Theta))$$

## 3. 实例分类

只保留掩码部分对应的特征：

$$F_i^{Mask}(\Theta) = F_i^{RoI}(\Theta) \cdot M_i(\Theta)$$

衔接两个4096-d的全连接层，损失函数为：

$$L_3 = L_3(C(\Theta) | B(\Theta), M(\Theta))$$

关闭

2017年08月 (6)

2017年07月 (3)

2017年06月 (3)

2017年05月 (4)

展开

## 阅读排行

开源代码文献

SSD: Single Shot MultiB (15492)

SSD: Single Shot MultiB (13547)

SSD: Single Shot MultiB (13214)

论文提要“You Only Look (12565)

SPPNet (10875)

Deep Residual Learning (8490)

Inception-v3:“Rethinking (7385)

目标检测--PVANET: Dee (6629)

论文提要“Fast Feature P (6153)

Faster R-CNN (5678)

## 评论排行

论文提要“You Only Look (25)

人脸检测“A Fast and Acc (20)

CompCars模型测试 (19)

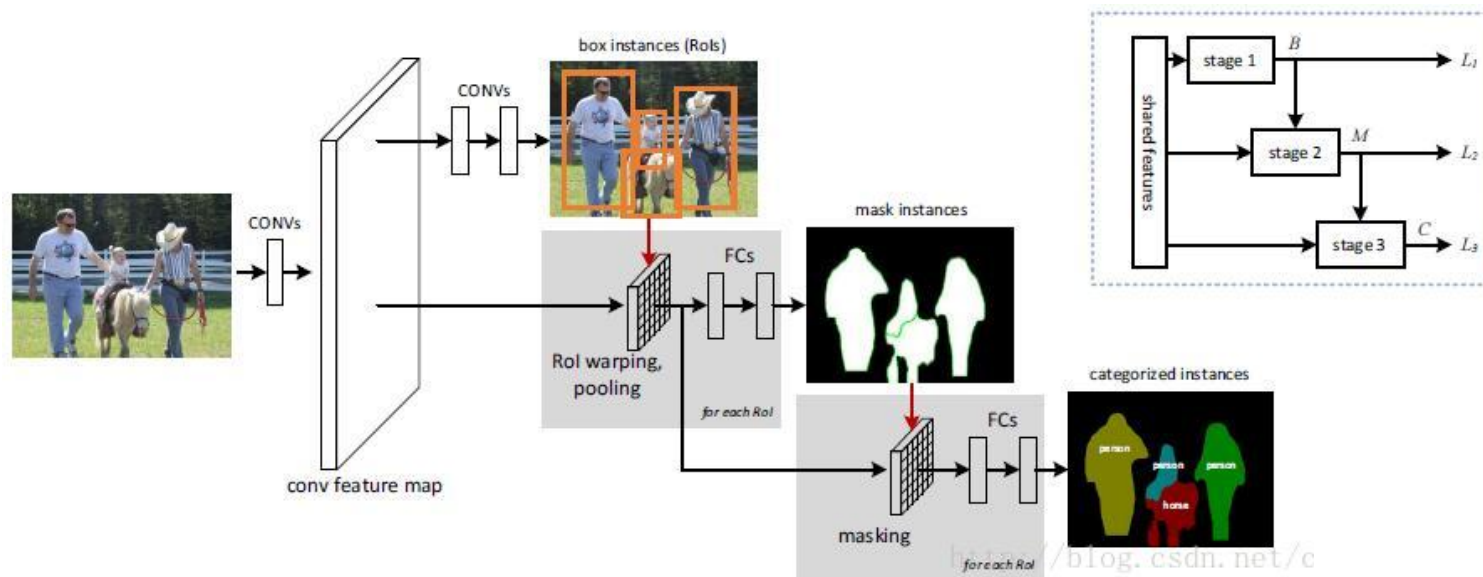
车辆检测“DAVE: A Unie (9)

论文提要“Fast Feature P (9)

车型识别“A Large-Scale (9)

论文提要“Hypercolumns (7)

整体网络结构如下图所示：



## 网络训练

1. 级联结构的损失函数为：

$$L(\Theta) = L_1(B(\Theta)) + L_2(M(\Theta) | B(\Theta)) + L_3(C(\Theta) | B(\Theta), M(\Theta)),$$

## 1. RoI Warping 层

目的是生成每个box更具分辨性的特征，该层裁切一个特征图区域，使用插值法将其warp到目标尺寸。给定预测出的bbox和全图的卷积特征图，RoI warp层在box内差值，输出固定大小的特征。

$$\mathcal{F}_i^{RoI}(\Theta) = G(B_i(\Theta))\mathcal{F}(\Theta)$$

增加RoI warp层是期望获取更高分辨率的特征，作用与空间转换网络（STN）类似，作者认为特征插值能够带来更具分辨力的特征。

## 更多层级联

第3级级联的输出可以作为新的proposals，再将第2级和第3级连接到其后，可以提升性能，示意图如下所示：

- 人脸识别 - A Discriminati (6)
- Beyond Local Search: Tr (5)
- A Lightened CNN for De (4)

### 推荐文章

- \* CSDN新版博客feed流内测用户征集令
- \* Android检查更新下载安装
- \* 动手打造史上最简单的Recycleview 侧滑菜单
- \* TCP网络通讯如何解决分包粘包问题
- \* SDCC 2017之大数据技术实战线上峰会
- \* 快速集成一个视频直播功能

### 最新评论

A Lightened CNN for Deep Face  
wyc2015fq: @liuxin000619:我也跑了，没有作者说的在cpu上那么快的速度，你那边什么情况能交流下不。

Multi-Task Learning with Low Ra  
linolzhang: 开通了知乎专栏，以文会友，欢迎大家投稿！  
<https://zhuanlan.zhihu.com/re-...>

Shallow and Deep Convolutional  
zhangyujun8175: 请问你有这篇文章的Deep model 吗，如果有可以发我一份吗游戏  
1095967026@qq.co...

MobileNets: Efficient Convolutio  
RjunL: 你好，请问用caffe实现需要修改caffe.proto中的内容吗？我试着去运行这个网络，但是总是报...

人脸识别 -Do We Really Need to

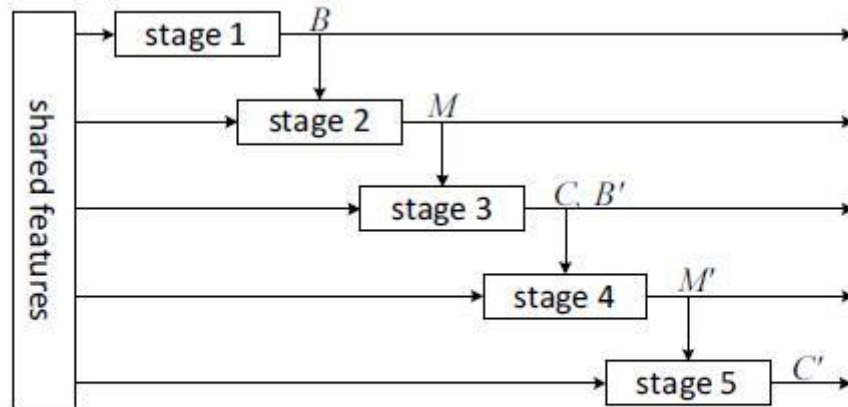


Figure 3. A 5-stage cascade. On stage 3, bounding boxes updated by the box regression layer are used as the input to stage 4.

### 实验结果

#### 1. 与其他实例分割方法的结果比较

method	mAP <sup>r</sup> @0.5 (%)	mAP <sup>r</sup> @0.7 (%)	time/img (s)
O <sup>2</sup> P [2]	25.2	-	-
SDS (AlexNet) [13]	49.7	25.3	48
Hypercolumn [14]	60.0	40.4	>80
CFM [7]	60.7	39.6	32
MNC [ours]	<b>63.5</b>	<b>41.5</b>	<b>0.36</b>

Table 2. Comparisons of instance-aware semantic segmentation on the PASCAL VOC 2012 validation set. The testing time per image (including all steps) is evaluated in a single Nvidia K40 GPU, except that the MCG [1] proposal time is evaluated on a CPU. MCG is used by [13, 14, 7] and its running time is about 30s. The running time of [14] is our estimation based on the description from the paper. The pre-trained model is VGG-16 for [14, 7] and ours. O<sup>2</sup>P is not based on deep CNNs, and its result is reported by [13].

关闭



husthy: 请问楼主有没有看过他们ResNet-101? 里面有个average\_face.bin, 完全不知道里面...

物体跟踪-Fully-Convolutional Si...  
qq\_20611159: 你好, 请问你跑通在github上下载的代码了吗?

MobileNets: Efficient Convolutio...  
qq\_34726032: 好的, 谢谢, 已明白, model直接训练得到的

图像分割“Fully Convolutional Ins...  
cv\_family\_z: @sjtukng1118:对于ROI中的某个像素, 1) detection:whether it b...

图像分割“Fully Convolutional Ins...  
姜淘淘: 博主, 你好! 请教一个问题: 文章2.2. Joint Mask Prediction and Clas...

MobileNets: Efficient Convolutio...  
cv\_family\_z: @qq\_34726032:1. 原始的卷积输入通道M, 输出通道N, 卷积和特征图组合是一步完成的; 2.d...

2. 与Fast/Faster RCNN检测结果比较

system	training data	mAP <sup>b</sup> (%)
R-CNN [10]	VOC 12	62.4
Fast R-CNN [9]	VOC 12	65.7
Fast R-CNN [9]	VOC 07++12	68.4
Faster R-CNN [26]	VOC 12	67.0
Faster R-CNN [26]	VOC 07++12	70.4
MNC [ours]	VOC 12	70.9
MNC <sub>box</sub> [ours]	VOC 12	73.5
MNC <sub>box</sub> [ours] <sup>†</sup>	VOC 07++12	75.9

Table 4. Evaluation of (box-level) object detection mAP on the PASCAL VOC 2012 test set. “12” denotes VOC 2012 trainval, and “07++12” denotes VOC 2007 trainval+test and 2012 trainval. The pre-trained model is VGG-16 for all methods.<sup>†</sup>: <http://host.robots.ox.ac.uk:8080/anonymous/NUWDYX.html>

顶 0 踩 0

上一篇 目标检测“A-Fast-RCNN: Hard Positive Generation via Adversary for Object Detection”

下一篇 MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications

相关文章推荐

- 《Instance-aware Semantic Segmentation via Mu...
- Presto的服务治理与架构在京东的实践与应用--王...
- MSRA instance-aware semantic segmentation的...
- 深入掌握Kubernetes应用实践--王渊命
- 论文解读-<Instance-aware Semantic Segmentatio...
- Python基础知识汇总
- 图像分割“LIP: Self-supervised Structure-sensitive...
- Android核心技术详解
- 图像分割“Not All Pixels Are Equal: Difficulty-Awar...
- Retrofit 从入门封装到源码解析
- 人工智能之自动驾驶系列（一）：概要
- 自然语言处理工具Word2Vec
- 图像分割“Understanding Convolution for Semanti...
- 图像分割“Efficient Deep Models for Monocular Ro...
- 目标检测“Perceptual Generative Adversarial Netw...
- 图像分割“LinkNet: Exploiting Encoder Represe...

### 查看评论

1楼 [御宅暴君](#) 2017-05-10 14:43发表



提问！Mask 层怎么光只凭 RoI 出的 14x14 分辨率 feature map 就能 mask 出 pixel-level 级别的 mask 结果。

Re: [cv\\_family\\_z](#) 2017-05-12 10:22发表



回复御宅暴君：主要的解释在论文3.2节，ROI后接两个全连接层，第二个全连接层有28\*28个输出，对每个输出与Ground Truth进行二值逻辑回归，得到pixel-wise的mask。

您还没有登录,请[\[登录\]](#)或[\[注册\]](#)

\* 以上用户言论只代表其个人观点，不代表CSDN网站的观点或立场

关闭

[公司简介](#) | [招贤纳士](#) | [广告服务](#) | [联系方式](#) | [版权声明](#) | [法律顾问](#) | [问题报告](#) | [合作伙伴](#) | [论坛反馈](#)

[网站客服](#)   [杂志客服](#)   [微博客服](#)   [webmaster@csdn.net](mailto:webmaster@csdn.net)   400-660-0108 | 北京创新乐知信息技术有限公司 版权所有 | 江苏知之为计算机有限公司 | 江苏乐知网络技术有限公司

京 ICP 证 09002463 号 | Copyright © 1999-2017, CSDN.NET, All Rights Reserved

