

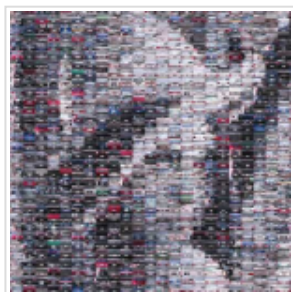
cv_family_z的博客

目录视图

摘要视图

RSS 订阅

个人资料



cv_family_z



访问：370024次

积分：5398

等级：BLOG > 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）

SSD: Single Shot MultiBox Detector 之再阅读

2016-07-14 10:58

13216人阅读

评论

分类：目标检测 (41) 深度学习 (90) ZJ (77)

版权声明：本文为博主原创文章，未经博主允许不得转载。

SSD: Single Shot MultiBox Detector SSD 一句话就是速度快，效果好！

第一版 8 Dec 2015，第二版是30 Mar 2016

主要改进是内容更加详实，实验更加丰富，尤其是和 Faster R-CNN 相比

比较明显。SSD把 候选区域提取步骤取消了。The fundamental improvement in speed comes from eliminating bounding box proposals and the subsequent pixel or feature resampling stage.

For 300×300 input, SSD achieves 72.1% mAP on VOC2007 test at 58 FPS on a Nvidia TitanX and for 500×500 input, SSD achieves 75.1% mAP, outperforming a comparable state of the art Faster R-CNN model.开源代码 <https://github.com/weiliu89/caffe/tree/ssd>

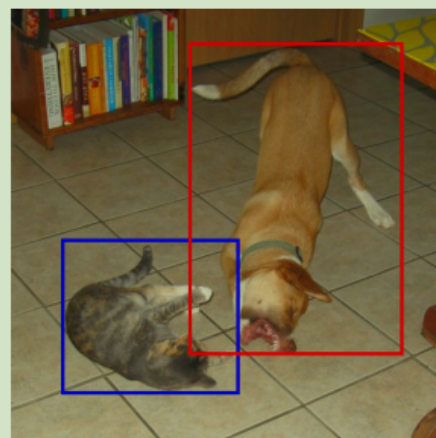
2 The Single Shot Detector (SSD)

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)



(a) Image with GT boxes

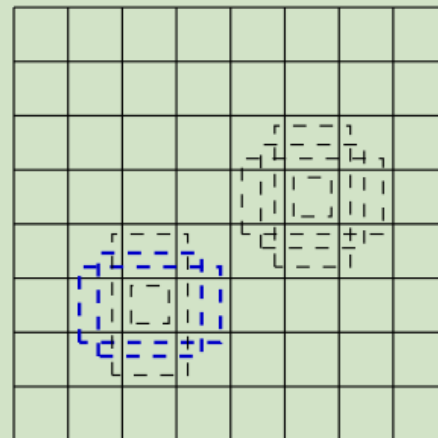
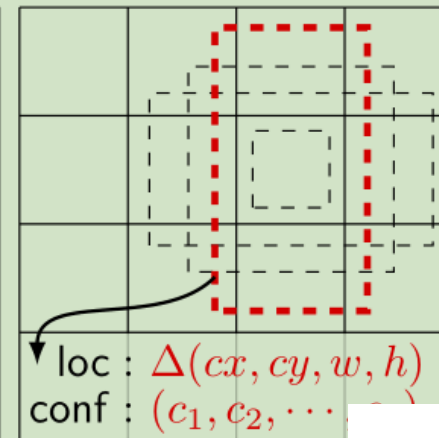
(b) 8×8 feature map(c) 4×4 feature

Fig. 1: SSD framework. (a) SSD only needs an input image and ground truth boxes for each object during training. In a convolutional fashion, we evaluate a small set of default boxes of different aspect ratios at each location in several feature maps of different scales (e.g. 8×8 and 4×4 in (b) and (c)). For each default box, we predict both the shape offsets and the confidences for all object categories $((c_1, c_2, \dots))$. At training time, we first match these default boxes to the ground truth boxes. For example, we have matched two default boxes with the cat. The matched boxes are treated as positives and the rest as negatives. The model loss is a weighted sum between localization loss (e.g. Smooth L1 [6]) and confidence loss (e.g. Softmax).

2.1 Model

SSD网络包括两个部分，前面的是基础网络，就是用于图像分类的标准网络，但是把涉及到分类的层全部裁掉，后面的网络是我们自己的设计的，主要实现以下目标：

Multi-scale feature maps for detection：我们加入卷积特征层，得到不同尺度的特征层，从而实现多尺度目标检测，用于不同尺度的目标预测的卷积模型是不同的。

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

源代码文献

(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 Uniec

(9)

论文提要"Fast Feature P

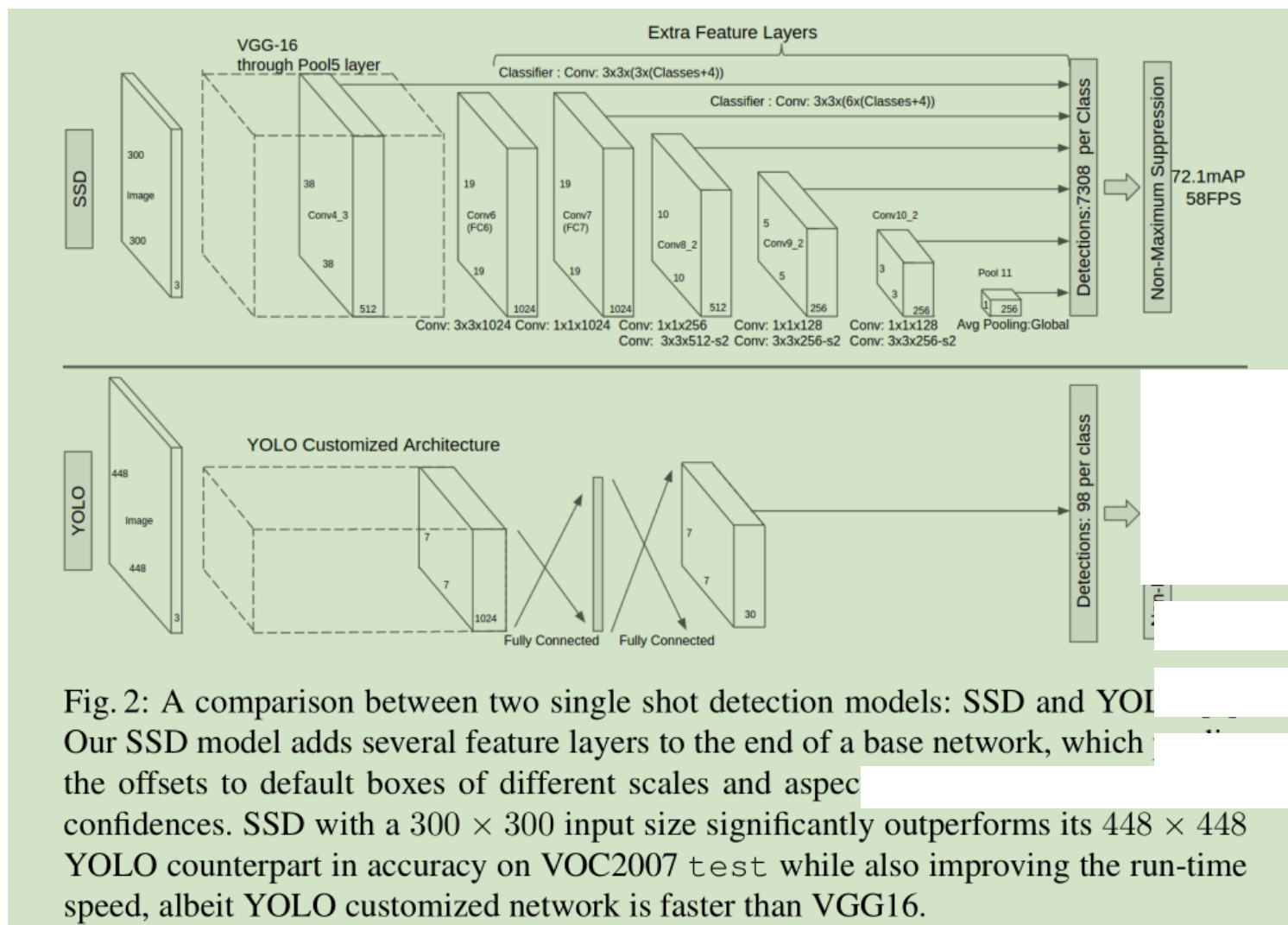
(9)

车型识别"A Large-Scale

(9)

论文提要"Hypercolumns

(7)



Convolutional predictors for detection

对于每个添加的特征层，我们使用一组卷积滤波器，可以得到一组固定数目的目标检测的预测。对于一个尺寸为 $m \times n$ ， p 通道的特征层，我们使用一个 $3 \times 3 \times p$ 的小核作为一个基础元素来预测一个可能检测的信息（类别信息，位置信息）

关闭

- 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,如果有可以发我以份吗游戏
1095967026@qq.co...
- MobileNets: Efficient Convolutior 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 tc

Default boxes and aspect ratios

在 Faster R-CNN中使用了 anchor boxes 实现不同大小和宽高比的物体提取,本文使用了类似的一组 default bounding boxes,和 Faster R-CNN 主要区别在于,我们是在不同尺度的特征层上进行这些default bounding boxes 检测运算的。

2.2 Training

训练SSD和训练一个使用候选区域及池化的标准检测器最大不同之处在于,真值信息需要被赋予一组固定集合检测输出中某一个特定输出。当这个赋值确定之后,损失函数和后向传播就可以被端到端的应用。

Matching strategy

在训练时,我们需要建立真值和 default boxes的对应关系。对于每个真值,我们选择宽高比、尺度的 default boxes 与之匹配,选择重合最大的 default boxe。这个和 origin [7] 是相似的。但是不同于 MultiBox,我们match default boxes to any ground truth with jaccard overlap higher than a threshold(0.5),这么做是为了简化学习问题

Training objective

SSD的训练目标函数是从 MultiBox 目标函数衍生出来的,但是被拓展到多类别问题。

$$L(x, c, l, g) = \frac{1}{N} (L_{conf}(x, c) + \alpha L_{loc}(x, l, g)) \quad (1)$$

Choosing scales and aspect ratios for default boxes

主要是利用了不同尺寸的特征层,在文献【10,11,12】中已经使用过,例如是 Hypercolumn。图1给出了一个示例,不同大小目标对应不同尺度

Hard negative mining

这里我们将正负样本比保持为 3:1

[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...

Data augmentation

为了使得模型适应各种情况, 我们做了训练数据扩展

3 Experimental Results

Base network 使用 VGG16

Method	mAP	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv
Fast [6]	70.0	77.0	78.1	69.3	59.4	38.3	81.6	78.6	86.7	42.8	78.8	68.9	84.7	82.0	76.6	69.9	31.8	70.1	74.8	80.4	70.4
Faster [2]	73.2	76.5	79.0	70.9	65.5	52.1	83.1	84.7	86.4	52.0	81.9	65.7	84.8	84.6	77.5	76.7	38.8	73.6	73.9	83.0	72.6
SSD300	72.1	75.2	79.8	70.5	62.5	41.3	81.1	80.8	86.4	51.5	74.3	72.3	83.5	84.6	80.6	74.5	46.0	71.4	73.8	83.0	69.1
SSD500	75.1	79.8	79.5	74.5	63.4	51.9	84.9	85.6	87.2	56.6	80.1	70.0	85.4	84.9	80.9	78.2	49.0	78.4	72.4	84.6	75.5

Table 1: **PASCAL VOC2007 test detection results.** Both Fast and Faster use input images whose minimum dimension is 600. The two SSD models have the same settings except that they have different input sizes (300×300 vs. 500×500). It is obvious that larger input size leads to better results.

	SSD300					
more data augmentation?	✓	✓	✓	✓	✓	✓
use conv4_3?	✓	✓	✓	✓	✓	✓
include $\{\frac{1}{2}, 2\}$ box?	✓	✓	✓	✓	✓	✓
include $\{\frac{1}{3}, 3\}$ box?	✓	✓	✓	✓	✓	✓
use atrous?	✓	✓	✓	✓	✓	✓
VOC2007 test mAP	65.4	68.1	69.2	71.2	71.4	72.1

Table 2: Effects of various design choices and components on SSD performance.

关闭

Method	<i>mAP</i>	FPS	# Boxes
Faster R-CNN [2](VGG16)	73.2	7	300
Faster R-CNN [2](ZF)	62.1	17	300
YOLO [5]	63.4	45	98
Fast YOLO [5]	52.7	155	98
SSD300	72.1	58	7308
SSD500	75.1	23	20097

Table 5: **Results on Pascal VOC2007 test.** SSD300 is the only real-time detection method that can achieve above 70% mAP. By using a larger input image, SSD500 outperforms all methods on accuracy while maintaining a close to real-time speed. The speed of SSD models is measured with batch size of 8.

顶 踩

1

1

关闭

上一篇 [Faster R-CNN 之再阅读](#)

下一篇 [CNN网络二值化--XNOR-Net : ImageNet Classification Using Binary Convolutional Neural Networks](#)

相关文章推荐

- [CNN网络二值化--XNOR-Net : ImageNet Classification Using Binary Convolutional Neural Networks](#)
- [SSD安装及训练自己的数据集](#)
- [Presto的服务治理与架构在京东的实践与应用--王...](#)
- [Retrofit 从入门封装到源码解析](#)
- [SSD的配置安装与测试](#)
- [Makefile.config for SSD](#)

- 深入掌握Kubernetes应用实践--王渊命
- 用ImageNet的数据集训练SSD(Single Shot MultiB...
- Python基础知识汇总
- SSD: Single Shot MultiBox Detector的安装配置和...
- Android核心技术详解
- 自然语言处理工具Word2Vec
- GTX1070+CUDA8.0+Ubuntu16.04+Caffe+SSD 深...
- ubuntu14.04配置SSD物体识别检测环境运行demo
- AlexNet论文翻译
- SSD: Single Shot MultiBox Detector 训练KITTI数...

查看评论

1楼 [baidu_38729758](#) 2017-05-10 14:45发表



作者辛苦了，不过自己看原文（包括看了您的译文后）还是没有明白训练部分的：“训练SSD和训练一个使用候选区域检测器最大不同之处在于，真值信息需要被赋予一组固定集合检测输出中某一个特定输出。当这个赋值确定之后，损失可以被端到端的应用。”该如何理解？
或许您可以解释的更具体通俗点吗？

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

[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](#)

关闭

[网站客服](#) [杂志客服](#) [微博客服](#) [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

