

知乎

首发于
深度学习大讲堂

258



分享

Towards Good Practices for Recognition & Detection

Qiaoyong Zhong, Chao Li, Yingying Zhang, Haiming Sun

Shicai Yang, Di Xie, Shiliang Pu

pushiliang@hikvision.com

Hikvision Research Institute

October 9, 2016

【高手之道】海康威视研究院ImageNet2016竞赛经验分享



程程

来自莫屯的小胖纸。

+ 关注她

李沐、罗若天、贾扬清、孔涛、王峰等 258 人赞了该文章

深度学习大讲堂致力于推送人工智能，深度学习方面的最新技术，产品以及活动。请关注我们的知乎专栏！

摘要

海康威视研究院独家授权，分享ImageNet2016竞赛Scene Classification第一名，Object Detection第二名，Object Localization第二名，Scene Parsing第七名背后的技术修炼之道。

下面为大家介绍海康威视研究院在本次ImageNet2016竞赛中的相关情况。

IMAGENET

ECCO
Common Objects in Context

Towards Good Practices for Recognition & Detection

Qiaoyong Zhong, Chao Li, Yingying Zhang, Haiming Sun

Shicai Yang, Di Xie, Shiliang Pu

258

11条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

知乎

首发于
深度学习大讲堂

October 9, 2016



258

团队成员和分工如下：



分享

Team Members

HIKVISION

Scene Classification:

- *Shicai Yang*

Scene Parsing:

- *Haiming Sun*
- *Di Xie*

DET + LOC:

- *Qiaoyong Zhong*
- *Chao Li*
- *Yingying Zhang*
- *Di Xie*

Summary of Our Submissions

HIKVISION

• Scene Classification

- 1st place, 0.0901 top5 error

• Scene Parsing

- 7th place, 0.53335 average mIoU & pixel accuracy

• Object Detection

- 2nd place, 0.653 mAP

• Object Localization

- 2nd place, 0.0874 localization error

Scene Classification

HIKVISION

• Data Augmentation

258

● 11条评论

分享

★ 收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

14
258
分享

知乎 首发于 深度学习大讲堂

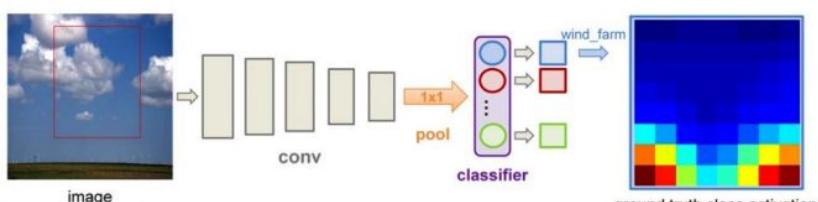
- Crop Sampling
 - scale jittering (from [3][4])
 - scale and aspect ratio augmentation (from [5])
 - random area ratio ($a = [0.08, 1]$)
 - random aspect ratio ($s = [3/4, 4/3]$)
 - crop size: $W' = \sqrt{W^*H^*a^*s}$; $H' = \sqrt{W^*H^*a/s}$
 - random offset to pick crop center, then crop and resize
- Supervised Data Augmentation

[1] <https://github.com/facebook/fb.resnet.torch/>
 [2] A. Krizhevsky, et al. ImageNet Classification with Deep Convolutional Neural Networks. NIPS, 2012.
 [3] K. Simonyan, et al. Very Deep Convolutional Networks for Large-Scale Image Recognition. ICLR, 2015.
 [4] K. He, et al. Deep Residual Learning for Image Recognition. CVPR, 2016.
 [5] C. Szegedy, et al. Going Deeper with Convolutions. CVPR, 2015.

数据增强对最后的识别性能和泛化能力都有着非常重要的作用。我们使用下面这些数据增强方法。
 第一，对颜色的数据增强，包括色彩的饱和度、亮度和对比度等方面，主要从Facebook的代码里改过来的。第二，PCA Jittering，最早是由Alex在他2012年赢得ImageNet竞赛的那篇NIPS中提出来的。我们首先按照RGB三个颜色通道计算了均值和标准差，对网络的输入数据进行规范化，随后我们在整个训练集上计算了协方差矩阵，进行特征分解，得到特征向量和特征值，用来做PCA Jittering。第三，在图像进行裁剪和缩放的时候，我们采用了随机的图像差值方式。第四，Crop Sampling，就是怎么从原始图像中进行缩放裁剪获得网络的输入。比较常用的有2种方法：一是使用Scale Jittering，VGG和ResNet模型的训练都用了这种方法。二是尺度和长宽比增强变换，最早是Google提出来训练他们的Inception网络的。我们对其进行了改进，提出Supervised Data Augmentation方法。

Scene Classification **HIKVISION**

- Supervised Data Augmentation (SDA)
 - train a model from scratch (coarse model)
 - use coarse model to generate ground truth class activation
 - randomly select a location based on prob. of target class
 - map this location to original image
 - randomly select a crop center near that location in original image
 - other steps are similar with the method in GoogLeNet paper



Inspired by: [6] B. Zhou, et al. Learning Deep Features for Discriminative Localization. CVPR, 2016.

尺度和长宽比增强变换有个缺点，随机去选Crop Center的时候，选到的区域有时候并不包括真实目标的区域。这意味着，有时候使用了错误的标签去训练模型。如图所示，左下角的图真值标签是风车农场，但实际上裁剪的区域是蓝天白云，其中并没有任何风车和农场的信息。我们在Bolei今年CVPR文章的启发下，提出了有监督的数据增强方法。我们首先按照通常方法训练一个模型，然后用这个模型去生成真值标签的Class Activation Map（或者说Heat Map），这个Map指示了目标物体出现在不同位置的概率。我们依据这个概率，在Map上随机选择一个位置，然后映射回原图，

14

11条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

知乎

首发于
深度学习大讲堂



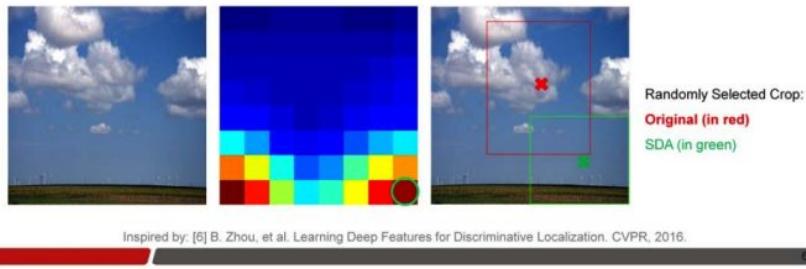
258



分享

• Supervised Data Augmentation (SDA)

- train a model from scratch (coarse model)
- use coarse model to generate ground truth class activation
- randomly select a location based on prob. of target class
- map this location to original image
- randomly select a crop center near that location in original image
- other steps are similar with the method in GoogLeNet paper



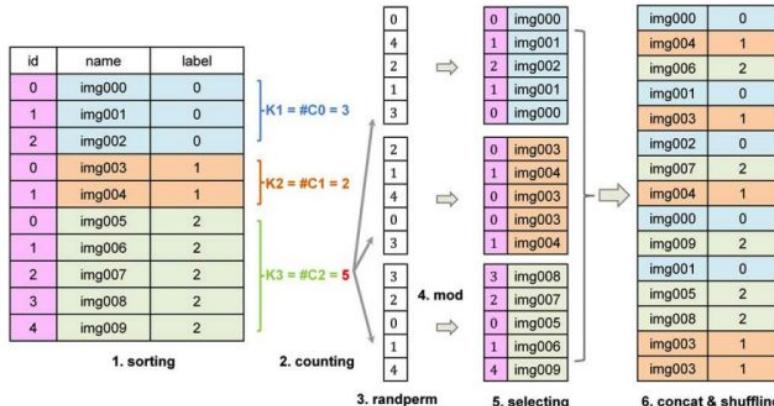
如图所示，对比原始的尺度和长宽比增强变换，我们方法的优点在于，我们根据目标物体出现在不同位置的概率信息，去选择不同的Crop区域，送进模型训练。通过引入这种有监督的信息，我们可以利用正确的信息来更好地训练模型，以提升识别准确率。（+0.5~0.7）

Scene Classification

HIKVISION

• Imbalanced Class Problem

- Balanced Sampling via *Label Shuffling*



[7] L. Shen, et al. Relay Backpropagation for Effective Learning of Deep Convolutional Neural Networks. ECCV, 2016.

场景数据集有800万样本，365个类别，各个类别的样本数非常不平衡，有很多类别的样本数达到了4万，也有很多类别的样本数还不到5000。这么大量的样本和非常不均匀的类别分布，给模型训练带来了难题。在去年冠军团队的Class-Aware Sampling方法的启发下，我们提出了Label Shuffling的类别平衡策略。在Class-Aware Sampling方法中，他们定义了2种列表，一是类别列表，一是每个类别的图像列表，对于365类的分类问题来说，就需要事先定义366个列表，很不方便。我们对此进行了改进，只需要原始的图像列表就可以完成同样的均匀采样任务。以图中的例子来说，步骤如下：首先对原始的图像列表，按照标签顺序进行排序；然后计算每个类别的样本数量，并得到样本最多的那个类别的样本数。根据这个最多的样本数，对每类随机都产生一个随机排

1 258

11条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

知乎

首发于
深度学习大讲堂

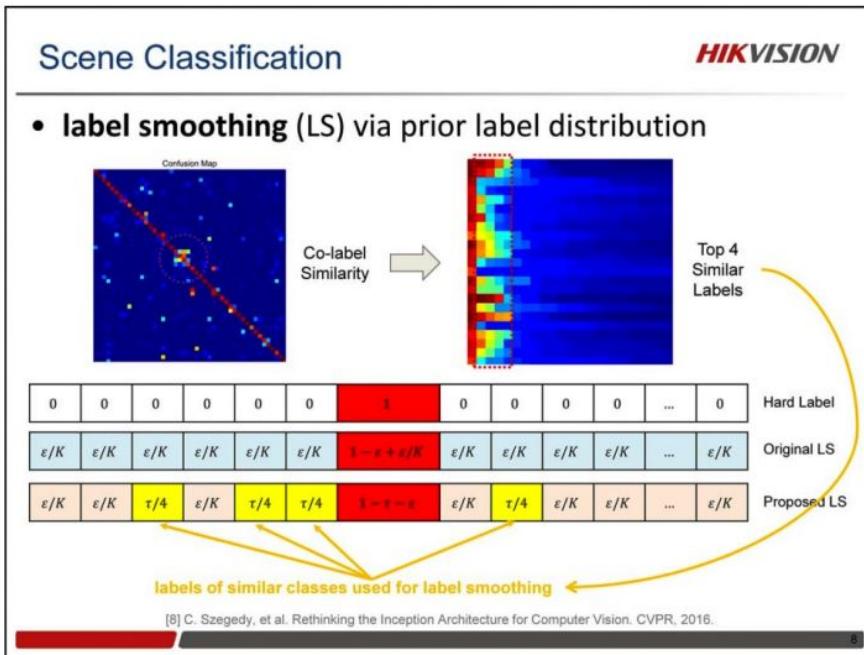
然后再重新做一遍这些步骤，得到一个新的列表，接着训练。Label Shuffling方法的优点在于，只需要原始图像列表，所有操作都是在内存中在线完成，非常易于实现。



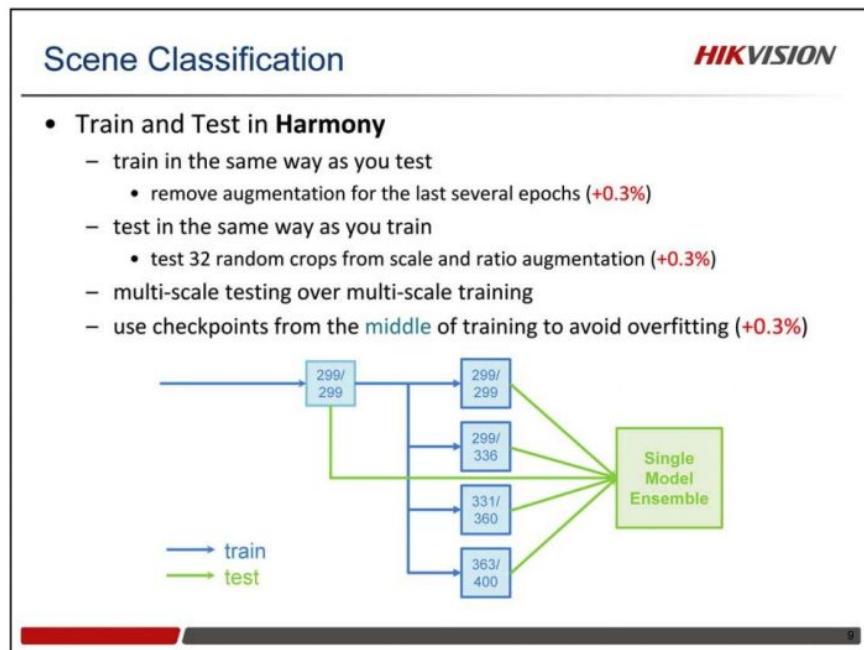
258



分享



我们使用的另外一个方法是Label Smoothing，是今年Google的CVPR论文中提出来的方法。根据我们的混淆矩阵（Confusion Matrix）的分析，发现存在很多跨标签的相似性问题，这可能是由于标签模糊性带来的。所以，我们对混淆矩阵进行排序，得到跟每个标签最相近的4个标签，用它们来定义标签的先验分布，将传统的one-hot标签，变成一个平滑过的soft标签。通过这种改进，我们发现可以从某种程度上降低过拟合问题。（+0.2~0.3）



258

11条评论

分享

收藏

下一篇
[Technical Review] ECCV16 Center Los...

知乎

首发于
深度学习大讲堂

用尺度和长宽比增强数据增强，在测试的时候也同样做这个变化，随机取32个crop来测试，也可以在最后的模型上提升一点性能。还有一条，就是多尺度的训练，多尺度的测试。另外，值得指出的是，使用训练过程的中间结果，加入做测试，可以一定程度上降低过拟合。



258



分享

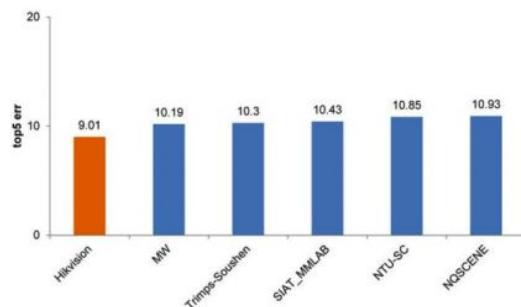
Scene Classification

HIKVISION

• Models

- Inception v3/Inception ResNet v2, and their variants
- Wider ResNet with 50/64 layers

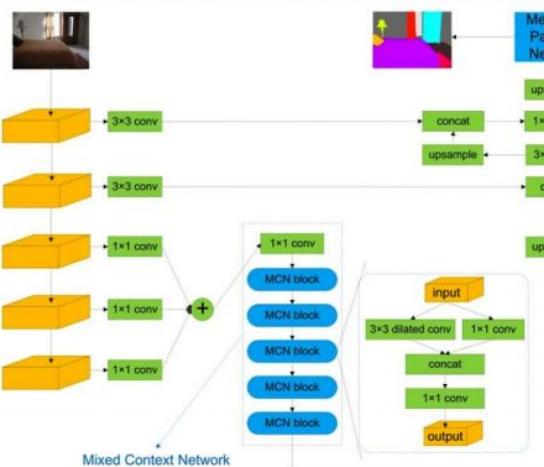
• Results



对于模型结构，没什么特别的改进，我们主要使用了Inception v3和Inception ResNet v2，以及他们加深加宽的版本。还用到了Wide ResNet。

Scene Parsing

HIKVISION



overall architecture for scene parsing

此次竞赛的语义分割任务非常具有挑战性。它一方面需要目标整体层面的信息，同时还需要每个像素的分类准确率。目前有很多语义分割的模型，但哪一种框架是最好的仍然是一个问题。我们设计了一个Mixed Context Network (MCN)，它由一系列Mixed Context Blocks (MCB) 堆叠而成。

258

11条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

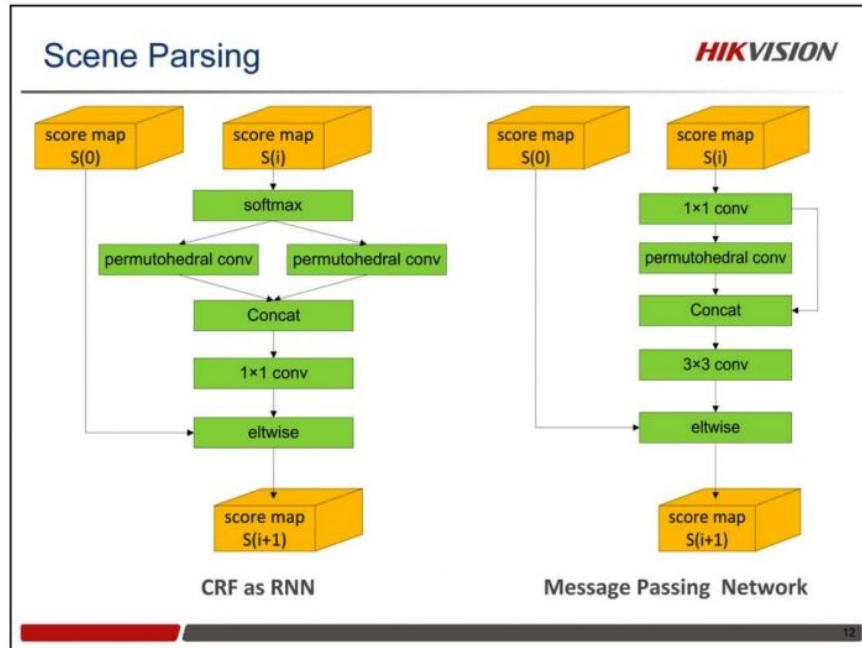
知乎

首发于
深度学习大讲堂

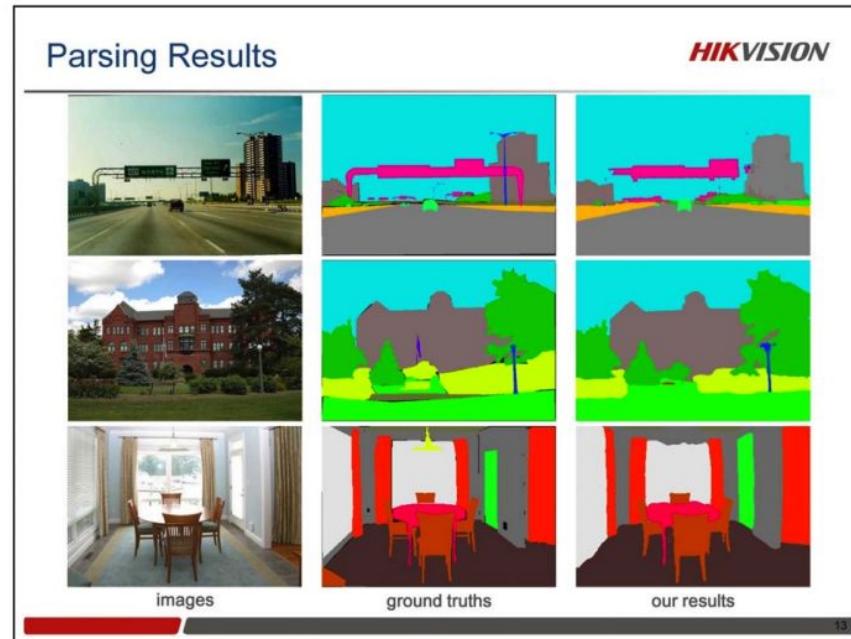
14

258

分享



CRF as RNN可以被加到网络的最后，与CNN一起联合训练。但是CRF as RNN比较耗费显存，尤其是在类别数比较大的时候（比如ADE20K有150类）。可以通过降低输入图像的分辨率来节省显存，但是这样做也会带来一些负面影响。为了解决这个问题，我们引入了一个新的比较省显存的模块，叫做MPN。在MPN中，我们首先将Score Map的通道从150降到了32，然后接了一个Permutohedral卷积层，用于做高维的高斯滤波。我们去掉了其中的平滑项，仅仅将1x1卷积层的特征和Permutohedral卷积层的特征连接，然后接一个3x3的卷积。实验证明，这样的结构也能较好地工作。



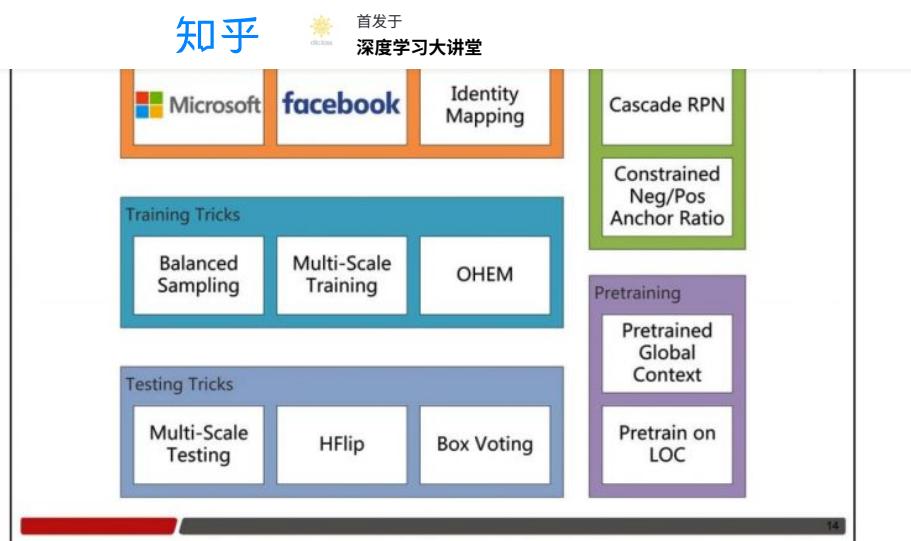
258

11条评论

分享

收藏

下一篇
[Technical Review] ECCV16 Center Los...

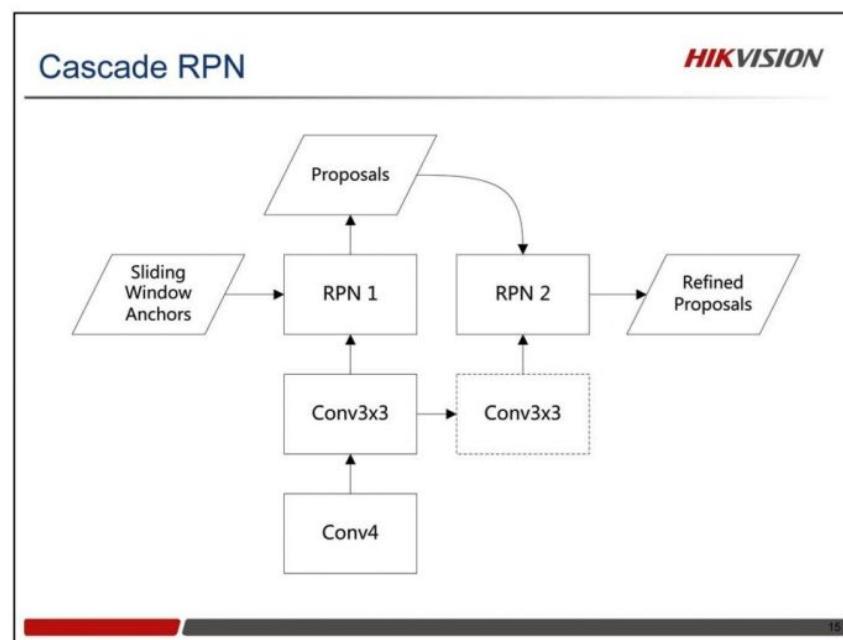


14

258

分享

我们的检测和定位，都是基于Faster-RCNN这个框架。图中我们列出了所有用到的技巧。有很多技巧在以前的文献中都可以找到，比如多尺度的训练和测试，难样本挖掘，水平翻转和Box Voting。但我们自己也做了很多新的改进，比如样本均衡，Cascade RPN，预训练的Global Context等。至于网络结构，我们仅仅用了三个ResNet-101模型。一个来自于MSRA，一个来自于Facebook，还有一个是我们自己训练的Identity Mapping版的ResNet-101。我们最好的单模型结果，是源自我们自己训练的Identity Mapping版的ResNet-101。



15

我们设计了一个轻量级的Cascade RPN。2个RPN顺序堆叠在一起。RPN 1使用滑窗Anchors，然后输出比较精确定位的Proposals。RPN 2使用RPN 1的Proposals作为Anchors。我们发现这个结构可以提升大中尺寸Proposals的定位精度，但不适合小的Proposals。所以在实际中，我们RPN1提取小的Proposals，RPN2提取大中尺寸的Proposals。注：Proposals尺寸的阈值是64 * 64。

Constrained NEG/POS Anchor Ratio

HIKVISION

258

11条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

知乎 首发于
深度学习大讲堂

- Expected N/P ratio: 1
- Real N/P ratio: usually **> 10**
- Our RPN
- Min batch size: 32
- Max N/P ratio: 1.5

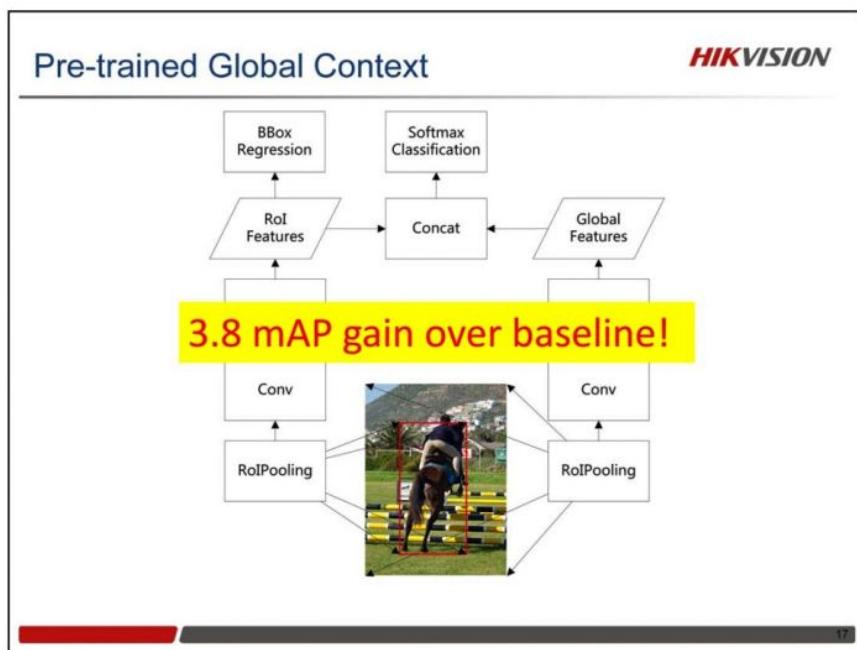
Recall@0.5	91.0	91.2	92.0	91.9
Recall@0.7	70.2	77.9	74.0	79.7
Average Recall	52.5	57.2	54.6	57.9

5.4 AR gain
9.5 Recall@0.7 gain

16

148
258
分享

另外一个改进就是限制正负Anchor的比例。在传统的RPN中，Batch Size通常是256，理想的正负Anchor比例是1。但是在实际使用中，这个比例往往会很大，通常会大于10。所以我们缩小了Batch Size，控制最大的比例为1.5。最小的Batch Size设置为32。对比实验表明，使用Cascade RPN和限制正负Anchor比例这两个策略，在ImageNet DET的验证集上，AR提升了5.4个点，Recall@0.7提升了9.5个点，而Recall@0.5只提升1个点。这说明Proposals的定位精度得到了显著的改善。



Global Context在去年Kaiming的论文中就已经提到，他们使用这个方法得到了1个点的mAP提升。我们也实现了自己的Global Context方法：除了在在ROI上做ROI Pooling之外，我们还对全局做了ROI Pooling来获得全局特征。这个全局特征仅仅被用来分类，不参加bbox回归。我们实验发现，Global Context如果使用随机初始化，其性能提升有限。当我们采用预训练的参数进行精调之后，发现mAP的性能可以提升3.8个百分点。

Pre-training on LOC HIKVISION

148
258
11条评论
分享
收藏
...
下一篇

[Technical Review] ECCV16 Center Los...

知乎

首发于
深度学习大讲堂



258



分享

0.5 mAP gain over baseline!

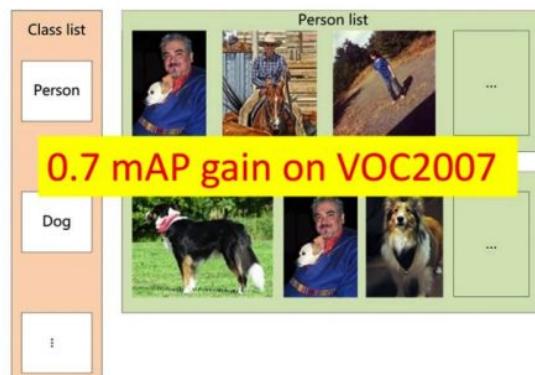
18

此外，我们发现，在1000类的LOC上预训练，然后再在DET数据上精调，可以得到额外0.5个点的mAP提升。

Balanced Sampling

HIKVISION

- Adapted from Shen et al. [7] for detection task



[7] L. Shen, et al. Relay Backpropagation for Effective Learning of Deep Convolutional Neural Networks. ECCV, 2016.

19

平衡采样是去年场景分类任务中所用到的一个技巧。我们也将它用来做检测任务。左侧是一个类别的列表，对于每个类别，我们又创建了一个图像列表。训练过程中，我们先从类别列表中选择一个类别，然后从这个类对应的图像列表中采样。和分类任务不同的是，检测任务中一张图像可能包含多个类别的目标。对于这种多标签的图像我们允许它们出现在多个类别的图像列表中。使用平衡采样技术，可以在VOC2007数据集上获得0.7的mAP提升。

Performance

HIKVISION

- ImageNet DET

	Team	mAP	Rank
Single Model	Hikvision	63.40	1
CUIImage		63.36	2

- PASCAL VOC2012

	Team	mAP	Rank
Hikvision		87.9	1
ResNet Baseline		83.8	2

258

11条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

The screenshot shows a Zhihu post with a table titled "IMAGENET CLS-LOC". The table has four columns: CLS, LOC, and Rank, with an additional header row for "LOC (Ensemble)". The data row shows values: 3.7 for CLS, 8.7 for LOC, and 2 for Rank. Below the table is a progress bar at 20% completion.

知乎 首发于
深度学习大讲堂

	CLS	LOC	Rank
LOC (Ensemble)	3.7	8.7	2

258 分享

集成上述所列的各项技术，我们的检测模型取得了SOTA的性能。在ImageNet DET任务中，我们以65.3的mAP获得了第二名。就单个模型而言，我们的模型能以少许优势排名第一。我们使用了相同的检测框架来完成ImageNet LOC任务。在最后的竞赛中，我们以8.7的定位误差排名第二。在PASCAL VOC 2012检测任务中，我们单模型获得了87.9的 mAP，超过了去年Kaiming的模型有4个百分点之多。

The slide is titled "Take Home Message" and features the Hikvision logo. It lists several key techniques:

- Scene Classification**
 - better utilize your data and model (SDA)
 - label smoothing via soft label
 - balanced sampling via label shuffling
 - train and test in harmony
- Scene Parsing**
 - Mixed Context Net & Message Passing Net
- Object Detection**
 - use identity mapping
 - cascade RPN
 - constrained NEG/POS anchor ratio
 - pre-trained global context
 - balanced sampling
- Object Localization**
 - LOC = CLS + DET

21

The slide is titled "Acknowledgments" and features the Hikvision logo. It lists the members of the HPC team:

- We would like to thank our HPC team:
 - Peng Wang
 - Jianfeng Peng
 - Xing Zheng
 - Zhiqiang Zhou
 - etc...

知乎

首发于
深度学习大讲堂

HIKVISION



258



分享

Thank you!

23

HIKVISION

该文章属于“深度学习大讲堂”原创，如需要转载，请联系[@果果是枚开心果.](#)

作者简介：

海康威视研究院 是海康威视最重要的核心部门，主要致力于基础技术和前沿技术的探索和创新，在视频编解码、视频图像处理、视频智能分析、云计算、大数据、云存储、人工智能等方面有深厚的技术积累，为海康威视核心产品和新兴业务拓展提供了有力的支撑，成为公司主营业务和创新业务发展的重要驱动力。研究院在KITTI、MOT、Pascal VOC、ImageNet等世界级人工智能竞赛中均获得过第一的好成绩。欢迎各位老师、学者、专家及其他业内人士，莅临杭州，参观交流。技术探讨、访问交流、求职招聘以及其他相关事宜，欢迎邮件联系谢迪博士：xiedi@hikvision.com。

原文链接：[【高手之道】海康威视研究院ImageNet2016竞赛经验分享](#)

欢迎大家关注我们的微信公众号，搜索微信名称：深度学习大讲堂



欢迎关注我们深度学习大讲堂的微信公众号。

编辑于 2016-10-27

深度学习 (Deep Learning)

258

11 条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

知乎

首发于
深度学习大讲堂

深度学习大讲堂

推送深度学习的最新消息，包括最新技术进展，使用以及活动，由中科视拓（SeetaT...）

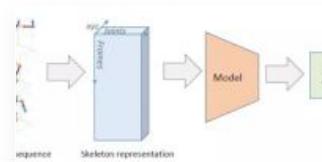
已关注

258

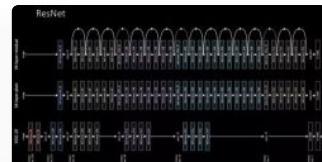


分享

推荐阅读



IJCAI 2018 | 海康威视Oral论文：
分层式共现网络，实现更好的动作识别和检测
机器之心 发表于机器之心



【仅需120万美元】24分钟训练
ImageNet，创世界纪录
新智元



阿里天池大数据竞赛心得：
50强付出与回报
面包君

258

11条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...

知乎

首发于
深度学习大讲堂

11条评论

切换为时间排序



写下你的评论...

258



分享

小土

1年前

谁能总结下ImageNet2016，所有top1、2的文章，列出个表来，谢谢！。。。

2

Beber

1年前

好吧一看是doctor的邮件地址就不敢投简历了

1

关闯生

1年前

海康威视是外资企业吗？

赞

EdisonGzq 回复 小土

1年前

看我的知乎，嘿嘿

赞 查看对话

普雅花郎 回复 关闯生

1年前

不是，是52所出来的，央企背景。

赞 查看对话

doonny

1年前

刷榜什么的，国人最喜欢了

赞

萧瑟

1年前

使用训练过程的中间结果，加入做测试，这个是指什么？

赞

StefanChou

8个月前

呃，深度学习本来是用来解决特征工程问题，然后发现算法和模型玩溜了之后，还是回到了特征工程上。

4

王二十 回复 StefanChou

7个月前

深有感觉，为了提高准确率也是拼了

赞 查看对话

思考中的哈士奇

6个月前

受到不少启发，挺好的，值得收藏

赞

何志

2个月前

有没有参考文献

赞

258

11条评论

分享

收藏

...

下一篇

[Technical Review] ECCV16 Center Los...