

2017-2018-2 学期学习汇报

汇报人：乔栋
实验室：110





中国地质大学
China University of Geosciences

艰苦朴素 求真务实

1952

汇报目录

- 一. 文献阅读情况
- 二. 实验情况
- 三. 研究现状分析总结
- 四. 创新思路
- 五. 存在的问题及困难



一 . 文献阅读情况

1、综述性论文阅读

泛读：14篇 精读：6篇

- [1] Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
- [2] Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE International Conference on Computer Vision. 2015.
- [3] Ren, Shaoqing, et al. "Faster R-CNN: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.
- [4] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." arXiv preprint arXiv:1506.02640 (2015).
- [5] Dai, Jifeng, et al. "R-FCN: Object Detection via Region-based Fully Convolutional Networks." arXiv preprint arXiv:1605.06409 (2016).
- [6] He, Gkioxari, et al. "Mask R-CNN" arXiv preprint arXiv:1703.06870 (2017).



一 . 文献阅读情况

2、方法及应用性论文

泛读：30篇 精读：5篇

- 目标检测和识别，10篇
- 行人重识别，3篇
- 人脸识别，2篇
- 姿态识别，2篇
- 遥感图像下小物体检测和识别，1篇
- 生成对抗网络，2篇

- Ruilong Li, Xin Dong, Zixi Cai, Dingcheng Yang, Haozhi Huang, Song-Hai Zhang, Paul L. Rosin, Shi-Min Hu, Pose2Seg: Human Instance Segmentation Without Detection, 2018
- Relation Networks for Object Detection, 2018
- Bharat Singh, Larry S. Davis, An Analysis of Scale Invariance in Object Detection - SNIP, 2018
- Min Bai, Raquel Urtasun, Deep Watershed Transform for Instance Segmentation, 2018
- Philip Colangelo, Nasibeh Nasiri, Asit Mishra, Eriko Nurvitadhi, Martin Margala, Kevin Nealis, Optimizing the Trade-off between Single-Stage and Two-Stage Object Detectors using Image Difficulty Prediction, 2018



中国地质大学
China University of Geosciences

艰苦朴素 求真务实

地大精神

《Pose2Seg: Human Instance Segmentation Without Detection》

方向: Instance Segmentation

出处: CVPR2018

论文摘要:

传统的图像实例分割算法是先进行物体检测，然后从检测的Bounding-box来分割物体边界。文中提出了一种全新的，基于pose的实例分割框架，该框架专门是基于人的pose来将人这个实例进行分割，而不再是基于proposal region detection。该算法框架可以获取和基于检测的算法几乎一样的准确度，并且可以更好的处理图像模糊问题，图像模糊问题应该是基于检测算法框架的最具挑战性的问题之一。



中國地質大學
China University of Geosciences

艱苦樸素 求真務實

地質大學

《Pose2Seg: Human Instance Segmentation Without Detection》

基本思路



- 1、检测出所有人的身体关键部位
- 2、连接关键部位，生成几个人体姿势（pose）



《Pose2Seg: Human Instance Segmentation Without Detection》

主要创新点

对齐模型 (Align module)：一般在实例检测中，都会有一个对齐模型，比如Mask-RCNN中的ROI-Align，该对齐模型用于裁剪用Bounding-box框出来的物体，并且调整尺寸到统一规格，以用于识别。因为本文基于Pose的算法很难找到bounding-box，所以，我们提出一种新的对齐模型：**AffineAlign**。

- ① 首先，我们从训练集中自动选择一些标准Pose作为模板
- ② 对于每个估计人型姿势，我们找到最相似的模板，然后计算放射变换矩阵 (Affine transformation matrix)。
- ③ 最后，将该矩阵应用与体征图像 (feature maps)，人型实例就以统一的规格中心对齐了。



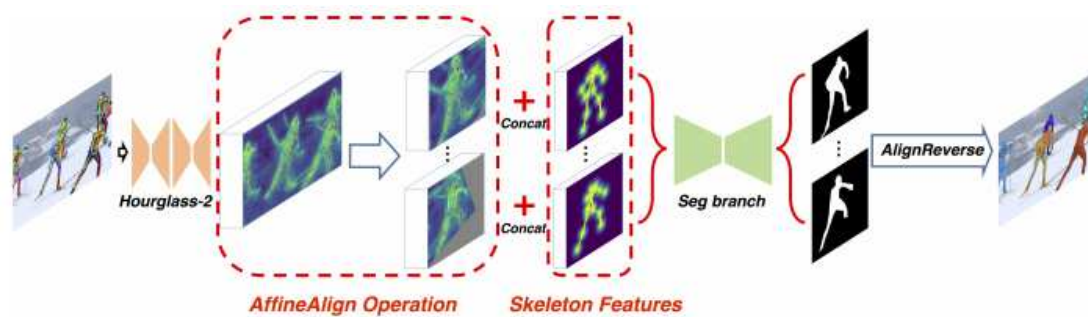
中国地质大学
China University of Geosciences

艰苦朴素 求真务实

1952

《Pose2Seg: Human Instance Segmentation Without Detection》

网络结构





《Pose2Seg: Human Instance Segmentation Without Detection》

总结与反思

- 1、在进行训练模型的时候，本文需要建立一个标准的姿势库以在测试的时候使用，但是文中只提出了标准姿势库是从训练样本中提取出来的，关于选取标准，并没有陈述清楚。
- 2、对齐模型是在物体识别中用于将Bounding-box的物体统一尺寸，以便于进行识别；本文最主要的创新点是提出了一种全新的对齐模型（Align module）：Affine Align；与传统的对齐模型不同的是：传统的对齐模型是通过切割、重新调整大小来进行统一尺寸；而该模型是通过矩阵仿射变换来进行统一尺寸。
- 3、本文所做的是人形实例分割，按照这种基本思路，对于同一类别的实物，比如：汽车、自行车，这些实物也是由一些固定的部件所组成的，考虑是否可以使用本文的思路。
- 4、传统的检测算法中，一般是使用最大抑制的方式来筛选Bounding-box，所以，当两物体重合度过高时，检测就会出现误差；本文也同时解决了该问题。



中国地质大学
China University of Geosciences

艰苦朴素 求真务实

110

《An Analysis of Scale Invariance in Object Detection》

方向: object detection

出处: CVPR2018

论文摘要:

本文在极端尺度变化的情况下进行检测和识别的不同算法进行了分析。基于分析，我们在具有相同尺寸的图像金字塔 (image pyramid) 训练和测试检测模型。因为小/大物体在相对更小/大的尺寸图像下是难以识别的，在这里我们介绍一种全新的训练范式 (training scheme)，叫做图像金字塔的尺度归一化 (Scale Normalization for Image Pyramids, SNIP)。



《An Analysis of Scale Invariance in Object Detection》

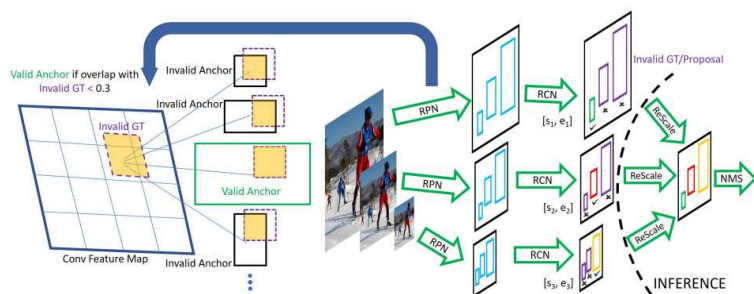
值得关注的是以下三组实验反映出的问题：

- ① 800(a11)和1400(a11)的对比：训练时使用不同大小的图训练，理论上如果使用更大图，小物体检测的性能应当有显著提升。但是实验表明这个提升非常小。文章中给出的解释是虽然1400的图训练会提升小物体的性能，但是会加大大物体训练的困难，所以此消彼长，并不会比较大提升。
- ② 1400(<80px)和1400(a11)的对比：既然大物体太难train了，可能对小物体造成干扰，是否去掉大物体可以提升小物体的性能呢？答案也是否定的，而且损失掉的大物体的语义会让结果变得更加糟糕。
- ③ MST：在Object Detection中，为了提升测试针对不同scale物体的性能，人们一般会使用Multi-scale training/testing这样的测试时融合的技巧来提升结果。与SNIP做法最大的区别就在于Multi-scale的做法扩充了不同scale样本的数目，但是仍然要求CNN去fit所有scale的物体。



《An Analysis of Scale Invariance in Object Detection》

网络结构



基本思想: SNIP是MST的升级版. 只有当这个物体的尺度与预训练数据集的尺度(通常224x224)接近时, 我们才把它用来做检测器的训练样本. 还基于一个假设, 即不同尺度的物体, 因为多尺度训练, 总有机会落在一个合理的尺度范围内. 只有这部分合理尺度的物体参与了训练, 剩余部分在BP的时候被忽略了。



中国地质大学
China University of Geosciences

艰苦朴素 求真务实

《Deep Watershed Transform for Instance Segmentation》

方向: Instance Segmentation

出处: CVPR2017

摘要:

现在大多数的实例分割都使用了比较复杂的原型，比如：条件随机场（Conditional Random Fields），循环神经网络（RNN），目标选取建议（object proposals），或者完全匹配模式（template matching schemes）。

而本文主要介绍了一种强大的端到端（end-to-end）卷积神经网络来解决此问题。

方法：研究结合了经典的watershed transform和现代深度学习（modern deep learning）来生成一张图片的能量地图（energy map），而实例分割就会以能量域（energy basins）的方式进行展现。然后，由单一的能量基线来生成一个切块，用与之相连接的成分来组成实例分割的结果。

结果：模型在challenging Cityscapes Instance Level Segmentation task中获得了优于state-of-the-art两倍的结果。



《Deep Watershed Transform for Instance Segmentation》

基本思路

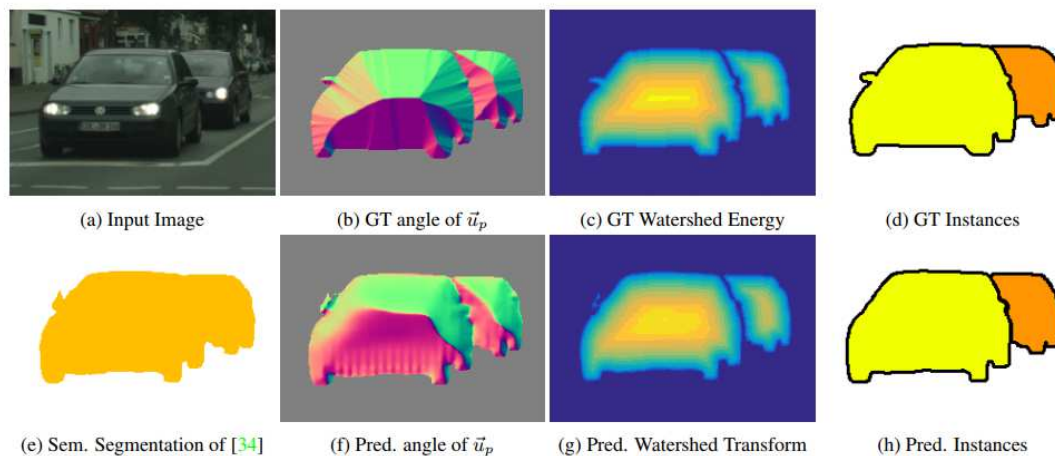
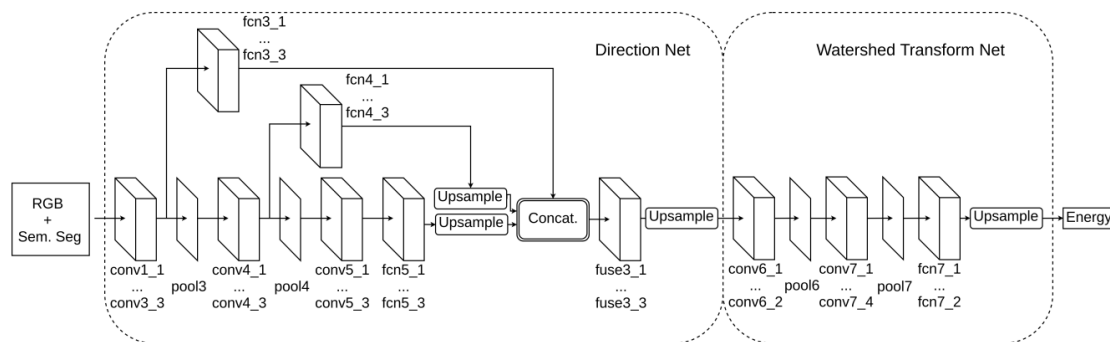


Figure 4: Our network takes the RGB image (a) and the semantic segmentation (e) as input, and predicts a unit vector at each foreground pixel pointing directly away from the nearest boundary (f). Based on this, we then predict a modified watershed transform energy (g), upon which we perform cut at a fixed threshold to yield the final predictions (h).



《Deep Watershed Transform for Instance Segmentation》

网络结构



1. 使用前向传播神经网络来学得分水岭变换的能量图 (the energy of watershed transform) ;
2. 该能量图满足: 每一个区域对应一个单一的instance; 所有的突起位于同一个高度线; 这样就能得到一张类似于等高线地图的能量图, 如上图中的(g);
3. 使用watershed transform算法完成instance segmentation。



中国地质大学
China University of Geosciences

艰苦朴素 求真务实

1952

《Deep Watershed Transform for Instance Segmentation》

总结与反思

本文的最大的难点在于如何将边界信息保存下来，以用于接下来 Watershed transform。在这里作者基于VGG-16的网络结构，设计了自己的Direction Net;

2. 思路：作者在使用距离计算的时候，使用的是点到边界的距离，且是距离边界最短的点，能否在训练的时候，使用instance的中间点（上下界和左右界的中间位置）参与距离运算。
3. 可以将该算法使用到目标检测领域。



中国地质大学
China University of Geosciences

艰苦朴素 求真务实

110

《Relation Networks for Object Detection》

方向: object detection

出处: CVPR2018

摘要:

关系模型化 (modeling relations) 是可以帮助进行物体识别 (object recognition) 的, 但是迄今为止, 并没有将其应用到深度学习领域。所有的物体检测系统仍然是分开独立的对物体进行识别的, 而在这过程中, 并未使用到他们之间的关系。

本文提出了一种物体关联模型 (object relation module)。该模型通过物体的 appearance feature 和几何特性之间的关系同时处理一组物体。



《Relation Networks for Object Detection》

网络结构

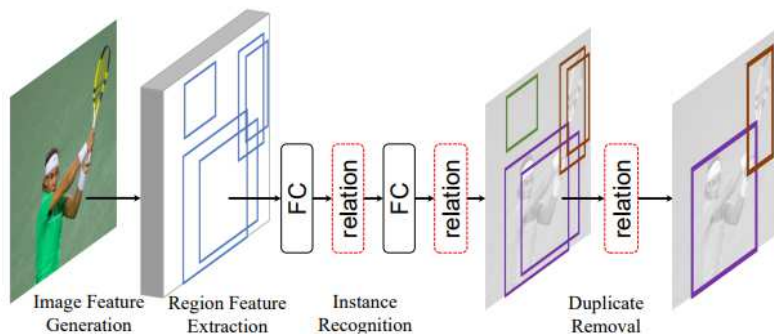


Figure 1. Current state-of-the-art object detectors are based on a four-step pipeline. Our object relation module (illustrated as red dashed boxes) can be conveniently adopted to improve both instance recognition and duplicate removal steps, *resulting in an end-to-end object detector*.

作者将关系模型当做了一个独立的部件，增加在了传统目标检测模型中。如图，作者在两个地方添加了关系模型：

1. Instance Recognition: the relation module enables joint reasoning of all objects and improves recognition accuracy;
2. Duplicate Removal: the traditional NMS method is replaced and improved by a lightweight relation network。



《Relation Networks for Object Detection》

总结与反思

- 1、目标检测新思路：不再是单独检测，而是利用物体之间的相互关系来实现同时多个目标的检测；
- 2、作者在使用关系模型的时候，一次性将所有检测目标加入计算。但是，一幅图像中并不一定所有物体都具有关联性，而大部分情况是只有距离相近的物体才有可能具有关联性，所以，在计算之前是否先可以将所有物体划分为几个独立的关联区域，然后再进行计算。
- 3、是否可以借鉴另外的注意力模型来完成检测？



《Optimizing the Trade-off between Single-Stage and Two-Stage Object Detectors using Image Difficulty Prediction》

方向: object detection

出处: CVPR2018

摘要:

目前，主要有两类最好的目标检测模型：

(1) 两步检测模型：先使用RPN网络得到ROI；然后再进行检测。这种模型准确率高，但是速度慢；

(2) 一步检测模型：这类模型将检测问题看做是一个回归问题，输出为物体类别的概率，以及Bounding Boxes。这种模型准确率低，但是速度快。鉴于以上模型的特性，作者依据图像的复杂程度，做了一个图像检测选择器：将复杂度低的图像送入一步检测模型进行检测，反之，送入两步检测模型，以求在速度和检测准确度两个方面获得平衡。



《Optimizing the Trade-off between Single-Stage and Two-Stage Object Detectors using Image Difficulty Prediction》

作者在进行图像复杂程度预测的时候依据的是CNN features 和 v 支持向量回归（线性回归）：

Algorithm 1: Easy-versus-Hard Object Detection

```
1 Input:  
2  $I$  – an input test image;  
3  $D_{fast}$  – a fast (single-stage) object detector;  
4  $D_{slow}$  – a slow (two-stage) object detector;  
5  $P$  – an image difficulty predictor;  
6  $t$  – a threshold for dividing images into easy or hard;  
  
7 Computation:  
8 if  $P(I) \leq t$  then  
9    $B \leftarrow D_{fast}(I)$ ;  
10 else  
11    $B \leftarrow D_{slow}(I)$ ;  
  
12 Output:  
13  $B$  – the set of predicted bounding boxes.
```



中国地质大学
China University of Geosciences

艰苦朴素 求真务实

地大精神

《Optimizing the Trade-off between Single-Stage and Two-Stage Object Detectors using Image Difficulty Prediction》

总结与反思

虽然作者在两类模型的前面加了一个选择器，但是，两类模型的训练并没有所谓的图像复杂度要求，后续，是否可以在训练的时候，就把图像复杂度考虑进去，即训练两步检测模型的时候，只使用复杂图像，而训练一步检测模型的时候，只使用简单图像；这样训练样本和测试样本的对应，应该也会提高检测的效果。



中国地质大学
China University of Geosciences

艰苦朴素 求真务实

《Pyramid Scene Parsing Network》

方向：（场景分析）Scene Parsing

出处：ICCV2017

摘要：

基于大量无限制的物体名词和形形色色的场景，场景解析是一件极其困难的任务。本文提出了名为金字塔场景分析网络（Pyramid scene parsing network, PSPNet），该网络利用金字塔池化模型（Pyramid pooling module）将基于不同区域的上下文信息组合，以得到全局上下文信息（global context information）。

PSPNet提供了一种像素级别的预测框架，它在ImageNet scene parsing challenge 2016, PASCAL VOC 2012 以及 Cityscapes 都获得了第一。



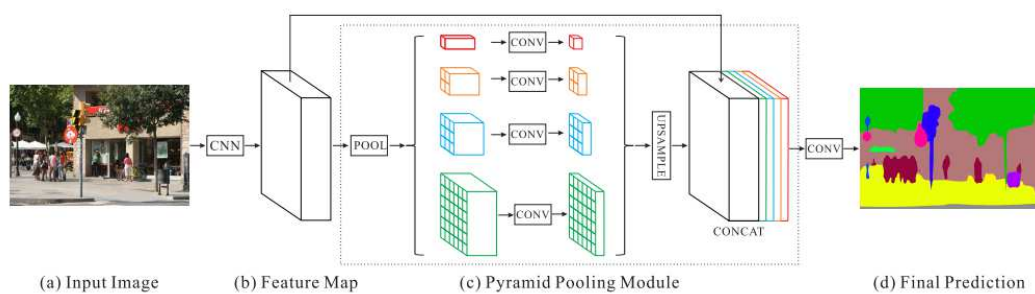
中国地质大学
China University of Geosciences

艰苦朴素 求真务实

地大精神

《Pyramid Scene Parsing Network》

网络结构





中国地质大学
China University of Geosciences

艰苦朴素 求真务实

地大精神

《Pyramid Scene Parsing Network》

总结与反思

- 1、PSPNet是通过多尺度整合的方法来进行场景分析的，该网络输入一张图片，先使用基础网络（如ResNet）得到特征图，然后使用金字塔池化模型分别对局部特征进行学习；将池化后的层进行降采样，以得到相同尺寸的层，最后对不同池化得到的层进行整合，以得到最后的预测结果。
- 2、在《Deep Transform for Instance Segmentation》中，作者就是使用PSPNet来得到输入图像的semantic Segmentation的。



中国地质大学
China University of Geosciences

艰苦朴素 求真务实

地大精神

二 . 实验情况

本学期在实验方面，主要学习了深度学习框架 Pytorch，在此基础上分别使用数据集电力检测图像、coco、VOC2007+2014进行实验。

框架: pytorch

数据集: 电力检测图像、COCO数据集、VOC数据集

算法: Faster R-CNN / Mask R-CNN



二 . 实验情况

框架: pytorch; 数据集: 电力设备检测图像; 网络: Faster-RCNN

```
AP for class1 = 0.7270
AP for class2 = 0.7383
AP for class3 = 0.1250
AP for class4 = 0.0000
AP for class5 = 0.4153
AP for class6 = 0.3215
AP for class7 = 0.7318
AP for class8 = 0.6212
AP for class9 = 1.0000
AP for class10 = 0.0545
AP for class11 = 0.0000
Mean AP = 0.4304
```

GROUND_TRUTH

step 6,734 Thu May 31 2018 10:36:32 GMT+0800 (中国标准时间)



default

GROUND_TRUTH

step 6,734 Thu May 31 2018 10:36:36 GMT+0800 (中国标准时间)



default_val

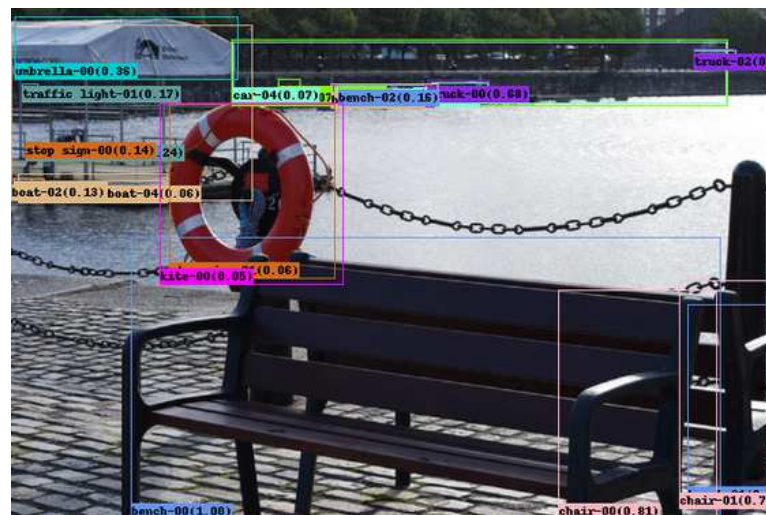


中国地质大学
China University of Geosciences

艰苦朴素 求真务实

二 . 实验情况

框架: pytorch;
数据集: coco;
网络: faster-RCNN





中國地質大學
China University of Geosciences

艰苦朴素 求真务实

二 . 实验情况

框架: pytorch;
数据集: VOC;
网络: faster-RCNN





三 . 研究现状分析总结

Faster R-CNN系列网络对于图像中大目标的检测有很高的准确率，但是对于小目标检测有很低的准确率。原因在于**RPN**网络进行特征提取的时候会进行池化操作，小目标物体的特征会在池化的过程中消失，所以，即使设置很低的**IOU**也无法准确检测到小目标。那么对于小目标检测的核心问题即是如何在特征提取的时候，保留其特征。



四 . 创新思路

4.1 目前思路整理

现行目标检测网络（如Faster R-CNN）由于池化操作是降采样，故而池化会使得小目标物体的特征值消失，从而常规目标检测算法在小目标检测场景中并不适应。针对小目标检测的修改可以从以下几个角度入手：

- **设置合适的Anchor:**

小目标物体对应的特征图也是很小的，故而可以通过减小anchor box的大小，来提高小目标物体的检测率。

- **对原图进行切割送入CNN:**

小目标物体在特征图中消失的根本原因在于图片大小和物体大小的比例过大，故而可以通过对原图切割，增大小目标物体的相对大小，以让其特征在特征图中显示出来，被RPN检测到，这里面临的核心问题是如何进行切割。

- **使用GAN来扩大检测目标:**

检测小目标物体的场景中，在把图片送入CNN之前，如果可以用GAN进行处理扩大小目标物体的尺寸，以使得RPN网络可以成功检测到小目标物体。这里面临的核心问题在于：如何确定小目标物体的位置。



四 . 创新思路

4.2 下一步计划

- 拓宽文献阅读范围：不只局限于CVPR，且保证数量和质量
- 实验深入
- 学术论文撰写：争取在下学期结束前完成学术论文的撰写
- 在实验过程中适当加入一些最新的成果