

# 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. 上一阶段工作总结
- 二. 下一阶段工作计划



## 一、上一阶段工作总结

### 1、目标检测（Object Detection）

基于选框的目标检测网络经典文章：

- 《Rich feature hierarchies for accurate object detection and semantic segmentation》,CVPR2014
- 《Fast R-CNN》,ICCV2015
- 《Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks》,CVPR2016
- 《Mask R-CNN》,ICCV2017

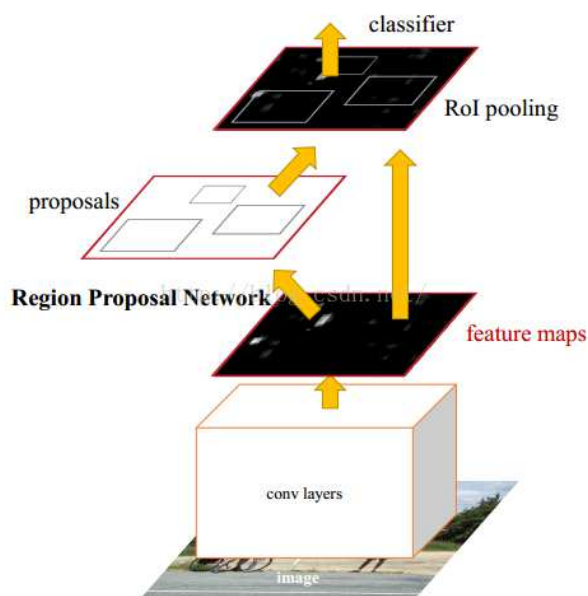


## 基于选框的目标检测网络家族概要

	使用方法	缺点	改进
R-CNN	<ol style="list-style-type: none"> <li>1、SS提取RP;</li> <li>2、CNN提取特;</li> <li>3、SVM分类;</li> <li>4、BB盒回归。</li> </ol>	<ol style="list-style-type: none"> <li>1、训练步骤繁琐（微调网络+训练SVM+训练bbox;</li> <li>2、训练、测试速度慢;</li> <li>3、训练占空间</li> </ol>	<ol style="list-style-type: none"> <li>1、从DPM HSC的34.3%直接提升到了66%（mAP）;</li> <li>2、引入RP+CNN</li> </ol>
Fast R-CNN	<ol style="list-style-type: none"> <li>1、SS提取RP;</li> <li>2、CNN提取特征;</li> <li>3、softmax分类;</li> <li>4、多任务损失函数边框回归。</li> </ol>	<ol style="list-style-type: none"> <li>1、依旧用SS提取RP(耗时2-3s, 特征提取耗时0.32s;</li> <li>2、无法满足实时应用, 没有真正实现端到端训练测试;</li> <li>3、利用了GPU, 但是区域建议方法是在CPU上实现的。</li> </ol>	<ol style="list-style-type: none"> <li>1、由66.9%提升到70%;</li> <li>2、每张图像耗时约为3s。</li> </ol>
Faster R-CNN	<ol style="list-style-type: none"> <li>1、RPN提取RP;</li> <li>2、CNN提取特征;</li> <li>3、softmax分类;</li> <li>4、多任务损失函数边框回归。</li> </ol>	<ol style="list-style-type: none"> <li>1、还是无法达到实时检测目标;</li> <li>2、获取region proposal, 再对每个proposal分类计算量还是比较大。</li> </ol>	<ol style="list-style-type: none"> <li>1、提高了检测精度和速度;</li> <li>2、真正实现端到端的目标检测框架;</li> <li>3、生成建议框仅需约10ms。</li> </ol>
Mask R-CNN	同Faster R-CNN, 添加实例分割	<ol style="list-style-type: none"> <li>1、还是无法达到实时检测目标;</li> <li>2、获取region proposal, 再对每个proposal分类计算量还是比较大。</li> </ol>	<ol style="list-style-type: none"> <li>1、修正ROI Pooling对齐问题;</li> <li>2、添加实例分割功能</li> </ol>



## Faster R-CNN介绍



Faster R-CNN创造性地提出了使用卷积神经网络实现的RPN（Region Proposal Network），解决了Select Search算法选框速度缓慢的瓶颈问题。但是RPN因为anchor box机制的原因，仍然存在以下问题：

1. 大量重复计算
2. 选框个数及大小需预先确定

目前基于选框的目标检测网络几乎都依赖于RPN网络，而无太多的更优替代方案。



中国地质大学  
China University of Geosciences

艰苦朴素 求真务实

1952

## 一、上一阶段工作总结

### 2、实例分割 (Instance Segmentation)

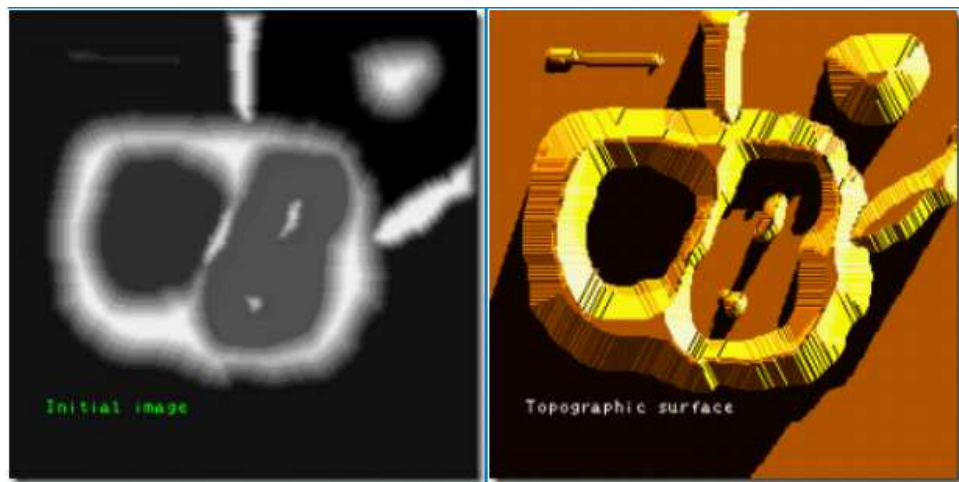
- 《Deep Watershed Transform for Instance Segmentation》, CVPR2017
- 《Pyramid Scene Parsing Network》, CVPR 2017
- 《TheCityscapes Dataset for Semantic Urban Scene Understanding》, CVPR2016





## 实例分割之DWTNet

### 1、分水岭算法（Watershed Transform Algorithm）



分水岭算法是一种图像区域分割法，在分割的过程中，它会把跟临近像素间的相似性作为重要的参考依据，从而将在空间位置上相近并且灰度值相近（求梯度）的像素点互连接起来构成一个封闭的轮廓。

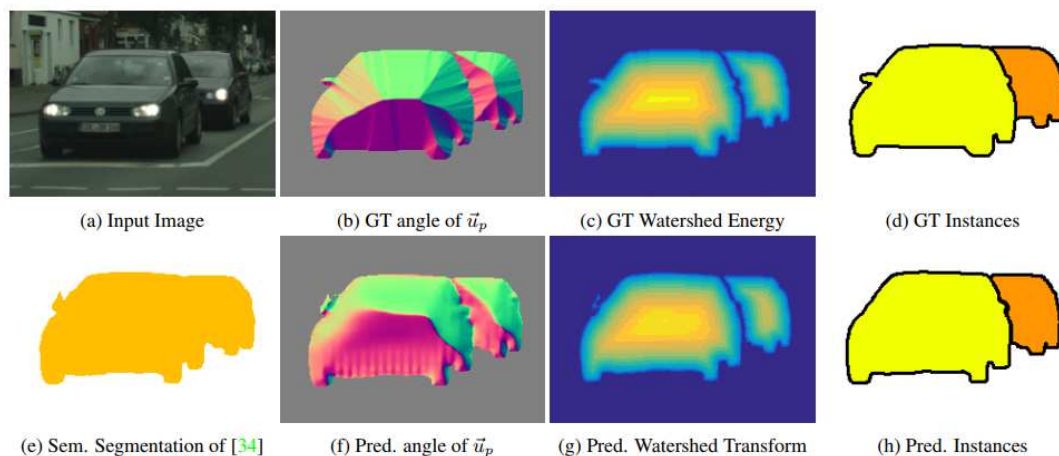
但是，由于图像的像素复杂，直接使用该算法并不会得到很好的效果，所以，应在使用该算法之前，先进行图像处理。

在实例分割中，处理的原则是：保留实例边界信息的前提下，简化像素矩阵，使其被分水岭算法分割时得到高精度度。



## 实例分割之DWTNet

### 2、DWTNet基本思路



以图像和语义分割的结果作为输入，计算每一个像素指向最近距离边界的向量，基于此，生成一张分水岭能量图，最后使用WT算法进行分割。

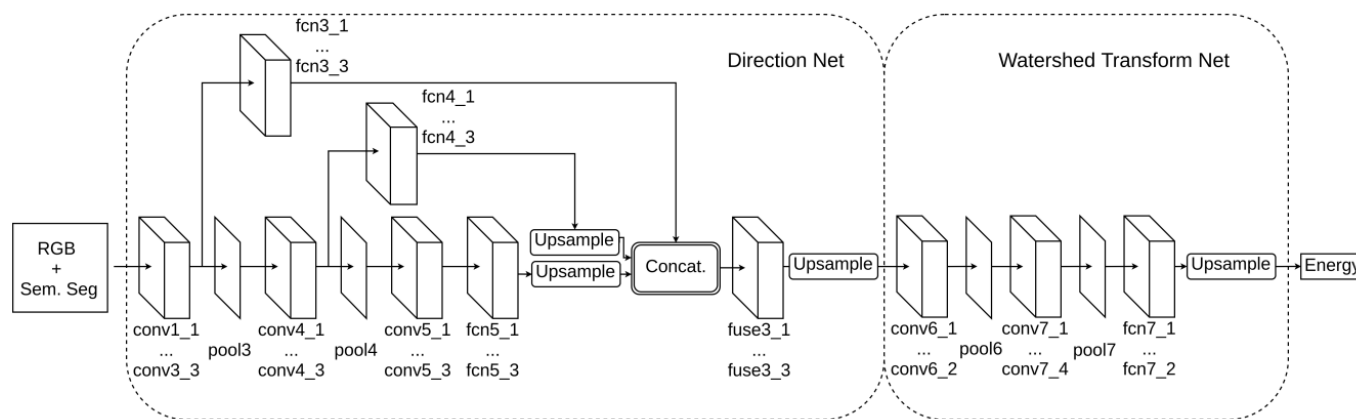
Figure 4: Our network takes the RGB image (a) and the semantic segmentation (e) as input, and predicts a unit vector at each foreground pixel pointing directly away from the nearest boundary (f). Based on this, we then predict a modified watershed transform energy (g), upon which we perform cut at a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





## 实例分割之DWTNet

### 3、DWTNet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。



## 实例分割之DWTNet

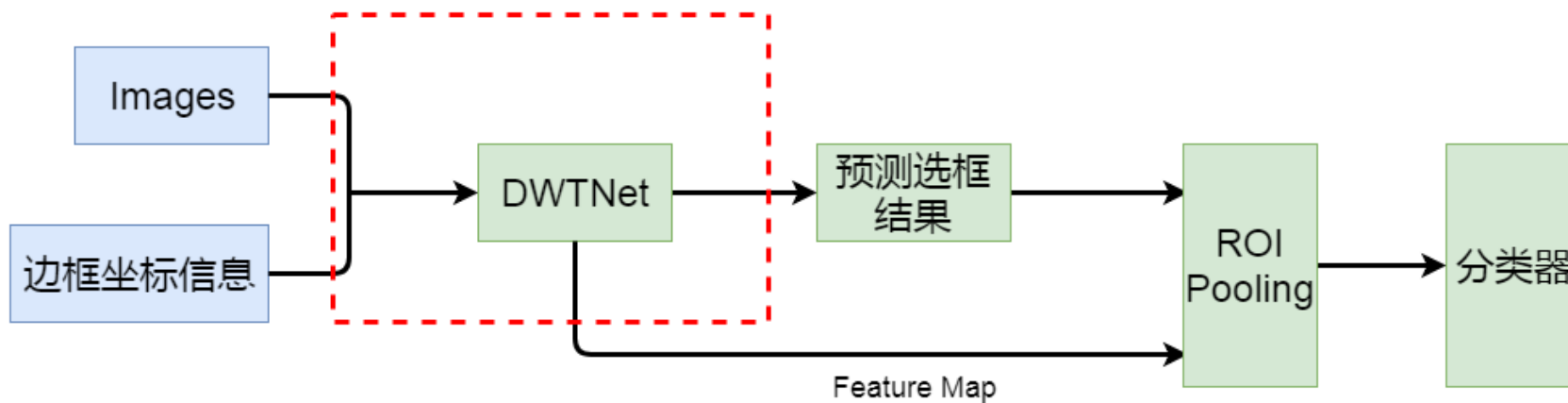
### 4、DWTNet网络结构分析

分水岭算法本身具有优秀的图像目标分割能力，DWTNet 通过方向向量在不丢失边界信息的情况下，简化了图像像素，以使其适用于实例分割。该模型在challenging Cityscapes Instance Level Segmentation task中获得了优于state-of-the-art两次的结果。

但是，分水岭算法并未被使用到目标检测的选框问题中，如果将目标检测问题中的图像选框看做是物体边界，将整张图像看做是DWTNet中的语义分割的结果，该网络就可以用于生成目标检测问题中的选框。



## 结合 Faster R-CNN 和 DWTNet





## 二、后续工作计划

- 1、复现DWTNet（已复现，实验结果还未达到论文所述）；
- 2、
  - ① 确定修改后的DWTNet的损失函数；
  - ② 将DWTNet语义分割部分去掉，添加选框坐标信息，并修改网络结构；
  - ③ 确定网络训练的方式；
- 3、将修改训练后的DWTNet和Faster R-CNN进行组合，并进行训练；