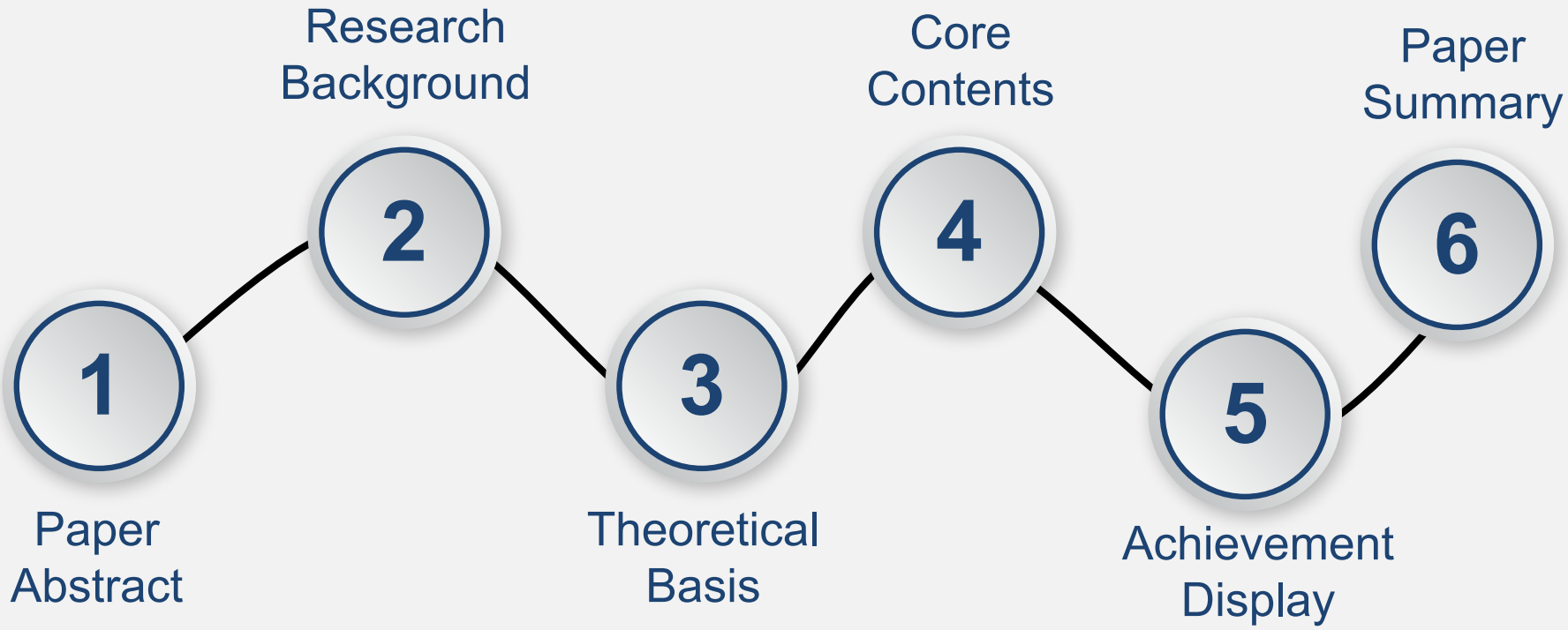
A Comparison of R-CNN and Faster R-CNN at Vehicle Detection Using the UA-DETRAC Video Dataset

Students : Xiaofo Geng and Lulu Zha

Supervisor : Suzanne Little

Catalog


1
Paper Abstract
2
Research Background
3
Theoretical Basis
4
Core Contents
5
Achievement Display
6
Paper Summary

01

Paper Abstract

Theme

Using R-CNN and Faster R-CNN algorithms to perform image detection and comparing their performance in the UA-DETRAC dataset.

Paper Abstract

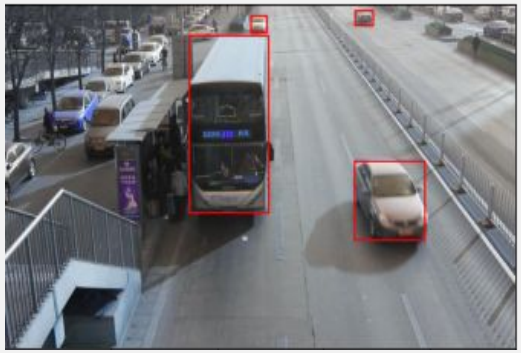


Figure 1

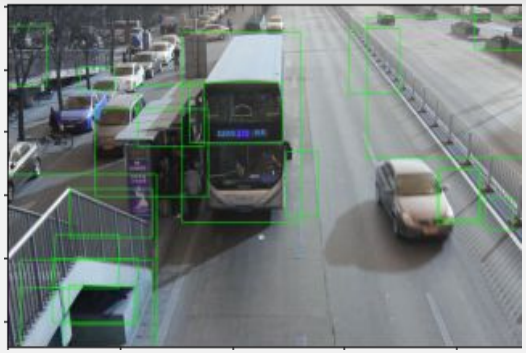


Figure 2

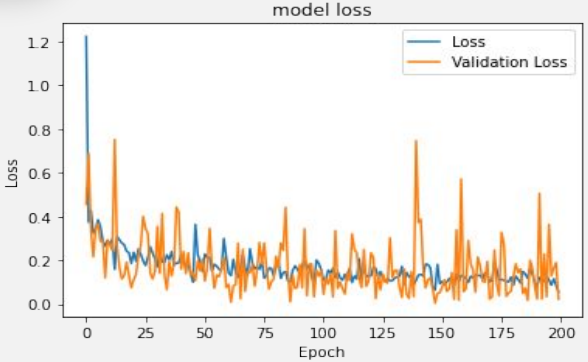


Figure 3

Figure 4

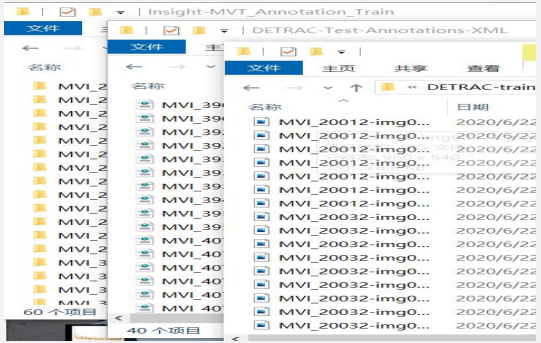


Figure 5

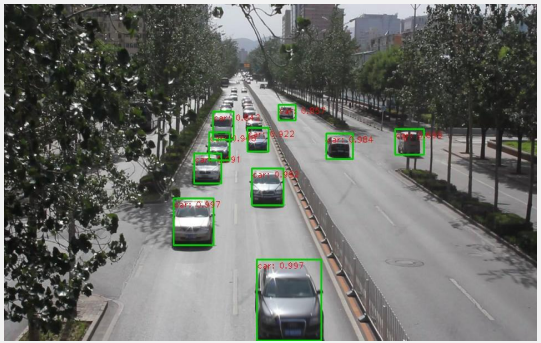


Figure 6

02

Research Background

Research Background

Geoffrey Hinton → **Deep Learning(DL)**

DL's development period: (2006 ~ 2012)
In 2006, Hinton and Salakhutdinov published "Reducing the Dimensionality of Data with Neural Networks" [1].

DL's explosive period: (after 2012)
In 2012, Hinton's research team won the ImageNet competition by constructing the AlexNet [2].

Research Background

Deep Learning → **Computer Vision**

In 2009, the ImageNet dataset was released to test whether computer vision could recognize all things in nature and overcome the problem of overfitting[3].

In 2012, Alexnet won the championship in the ImageNet competition [2].

In 2014, VGG won the ImageNet competition [4].

Research Background

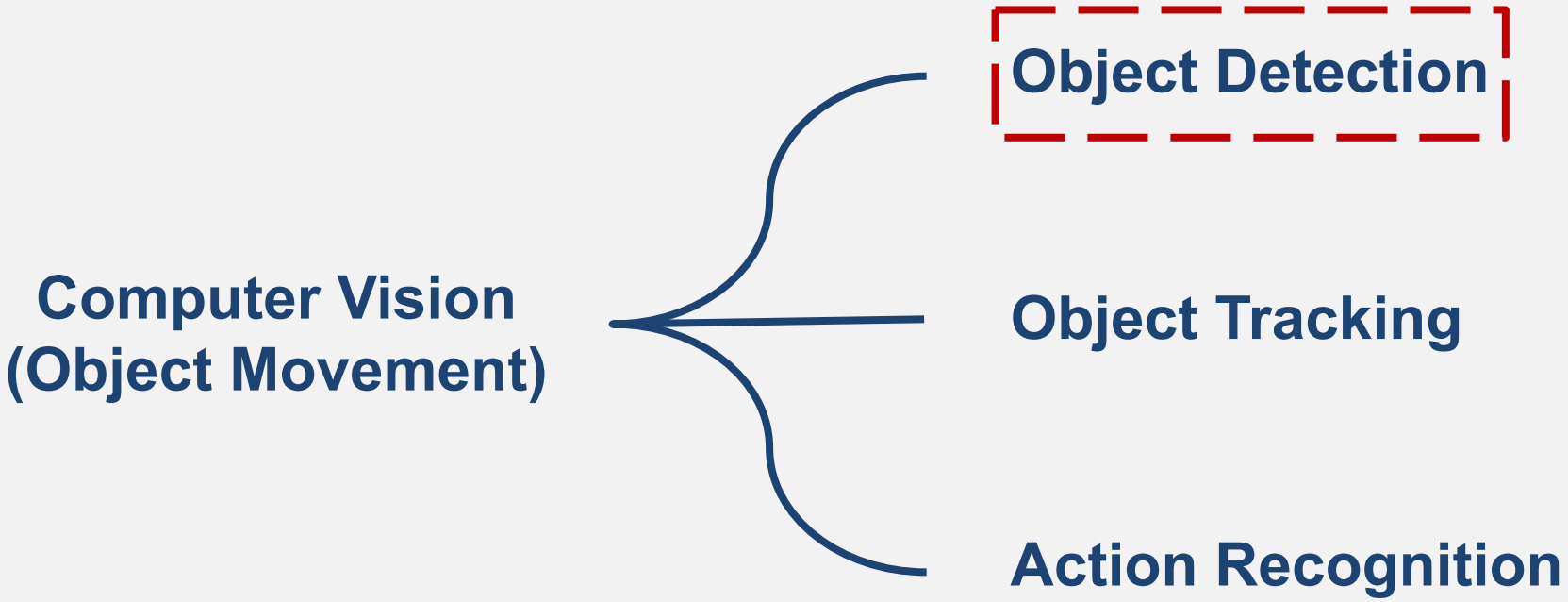


Figure 7

Figure 8

The Challenge of Object Detection

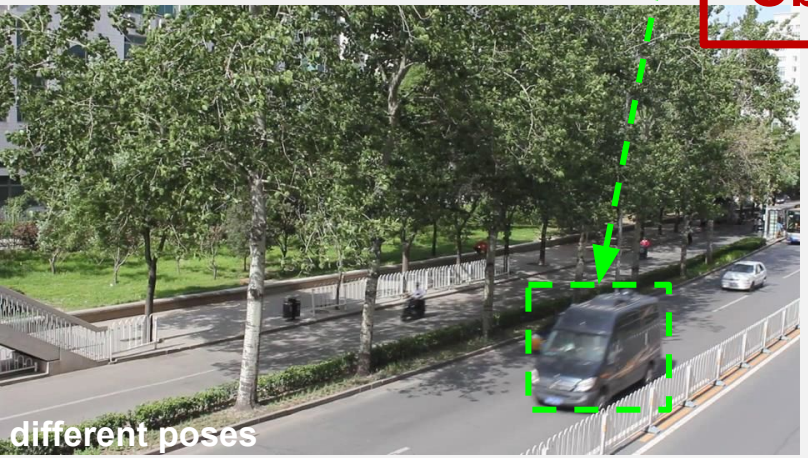


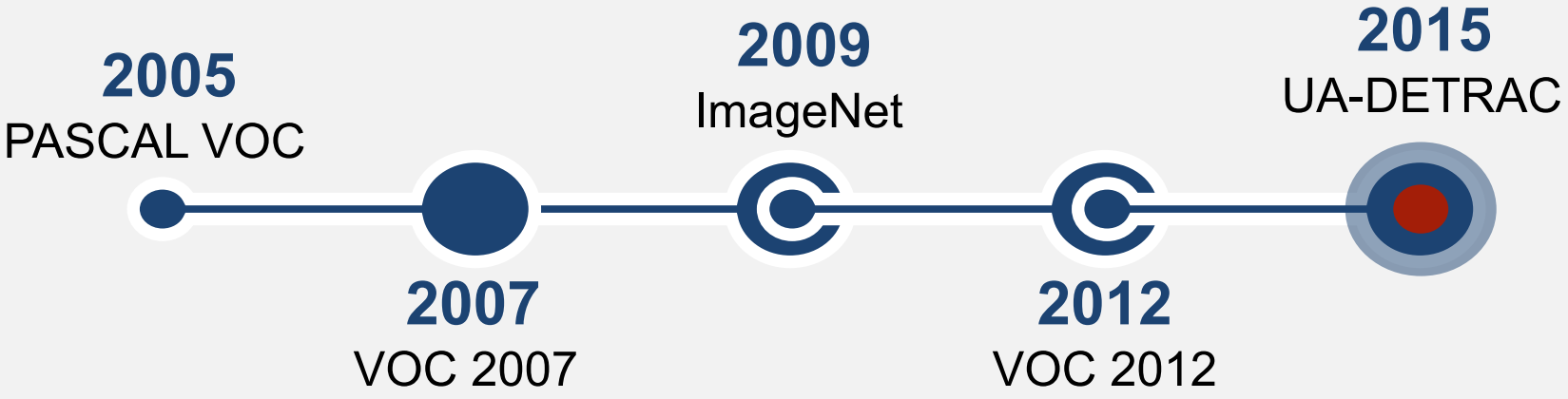
Figure 9

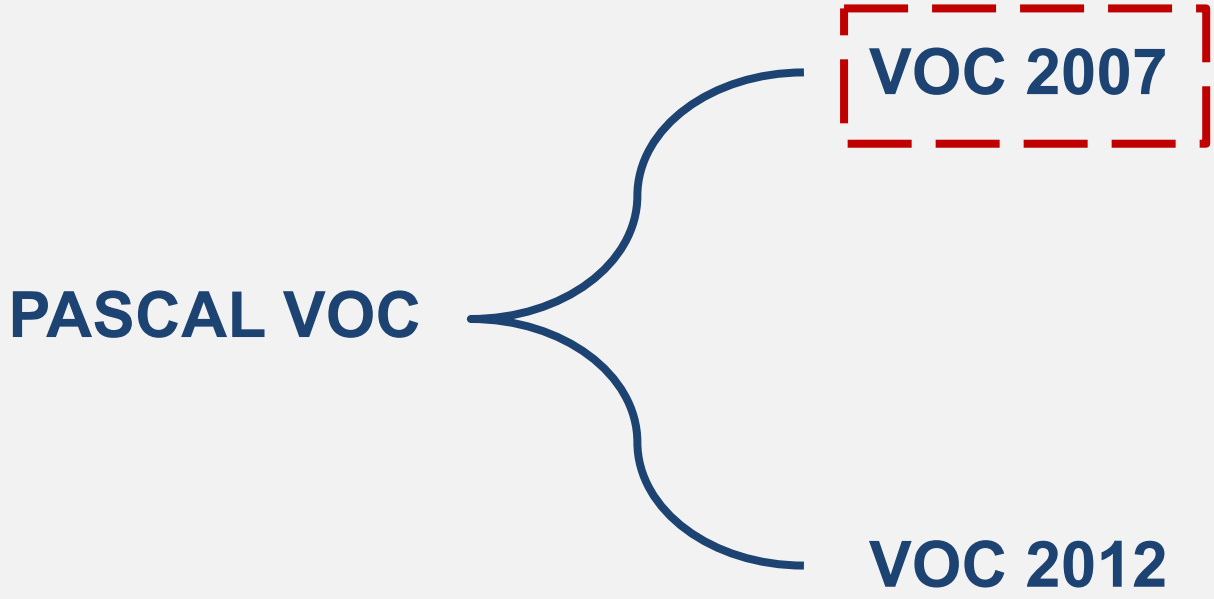
Figure 10

03

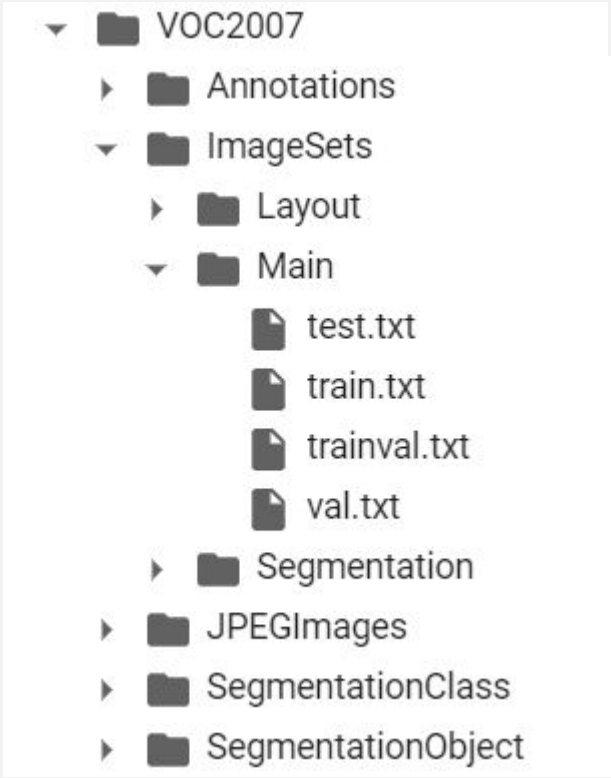
Theoretical Basis

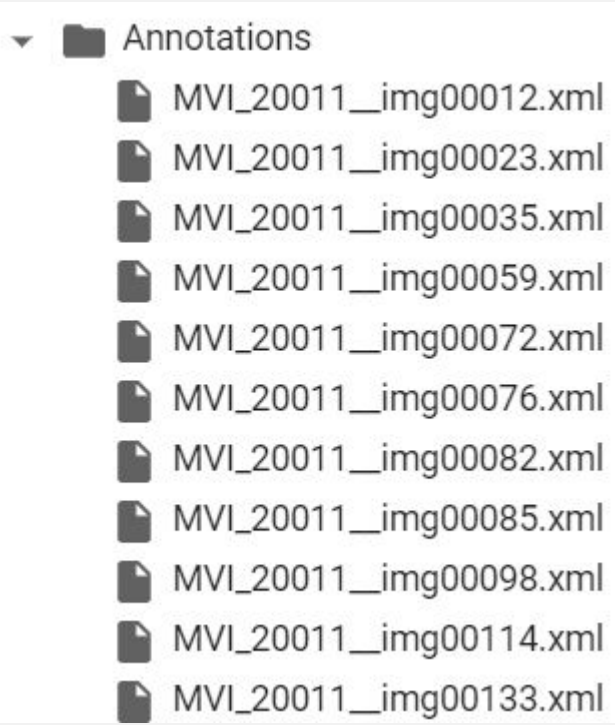
Timeline

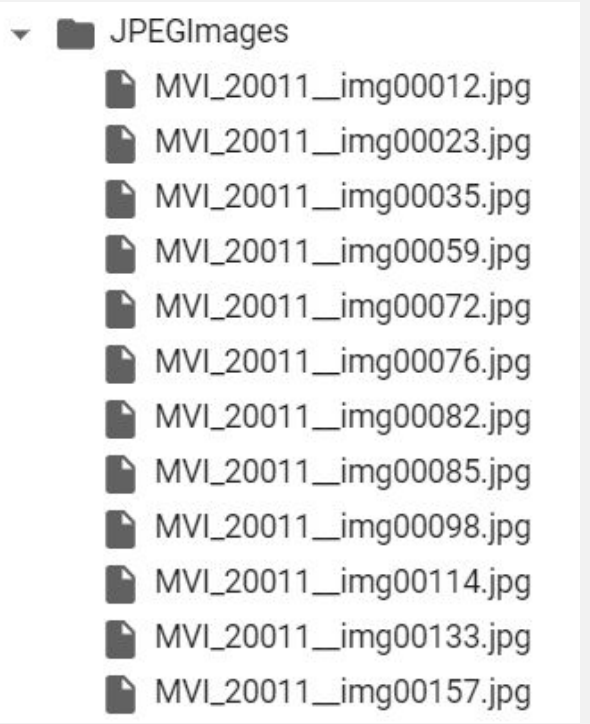

2005
PASCAL VOC
2007
VOC 2007
2009
ImageNet
2012
VOC 2012
2015
UA-DETRAC


PASCAL VOC
VOC 2007
VOC 2012

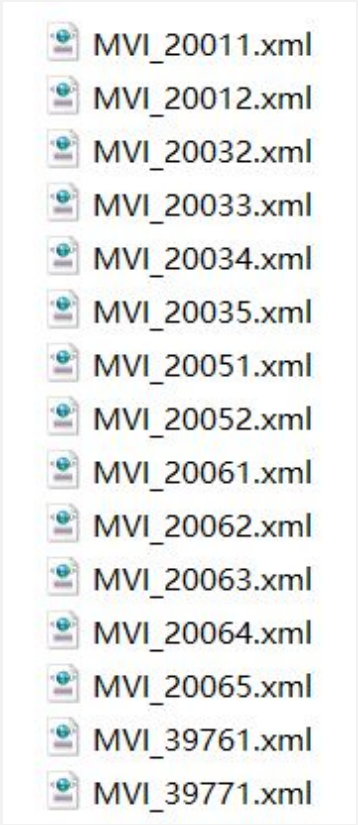
VOC 2007


VOC2007
Annotations
ImageSets
Layout
Main
test.txt
train.txt
trainval.txt
val.txt
Segmentation
JPEGImages
SegmentationClass
SegmentationObject


Annotations
MVI_20011__img00012.xml
MVI_20011__img00023.xml
MVI_20011__img00035.xml
MVI_20011__img00059.xml
MVI_20011__img00072.xml
MVI_20011__img00076.xml
MVI_20011__img00082.xml
MVI_20011__img00085.xml
MVI_20011__img00098.xml
MVI_20011__img00114.xml
MVI_20011__img00133.xml


JPEGImages
MVI_20011__img00012.jpg
MVI_20011__img00023.jpg
MVI_20011__img00035.jpg
MVI_20011__img00059.jpg
MVI_20011__img00072.jpg
MVI_20011__img00076.jpg
MVI_20011__img00082.jpg
MVI_20011__img00085.jpg
MVI_20011__img00098.jpg
MVI_20011__img00114.jpg
MVI_20011__img00133.jpg
MVI_20011__img00157.jpg

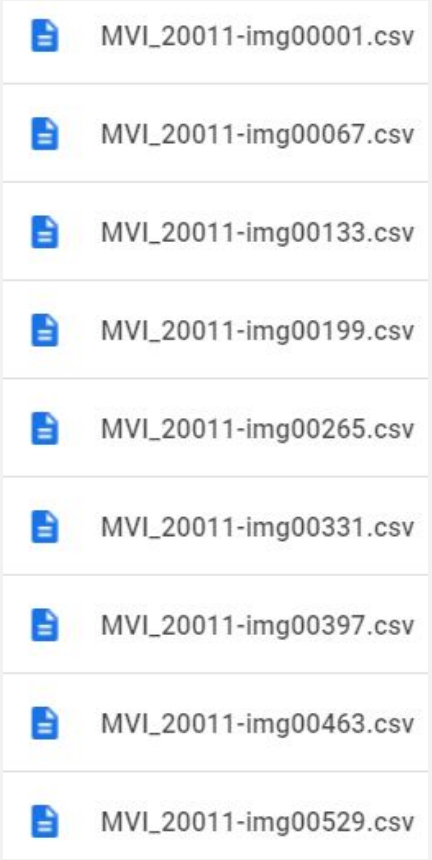
UA-DETRAC

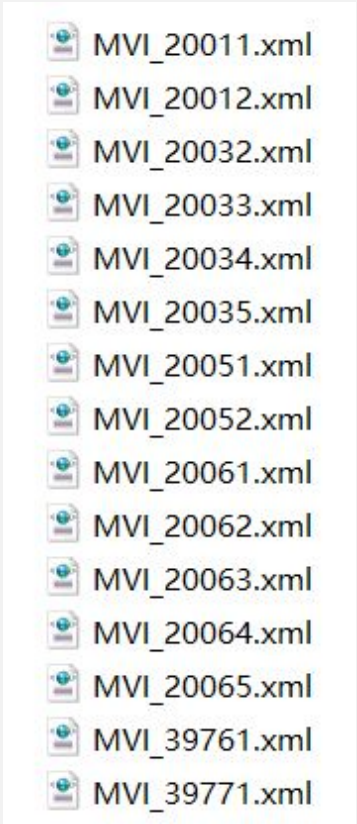


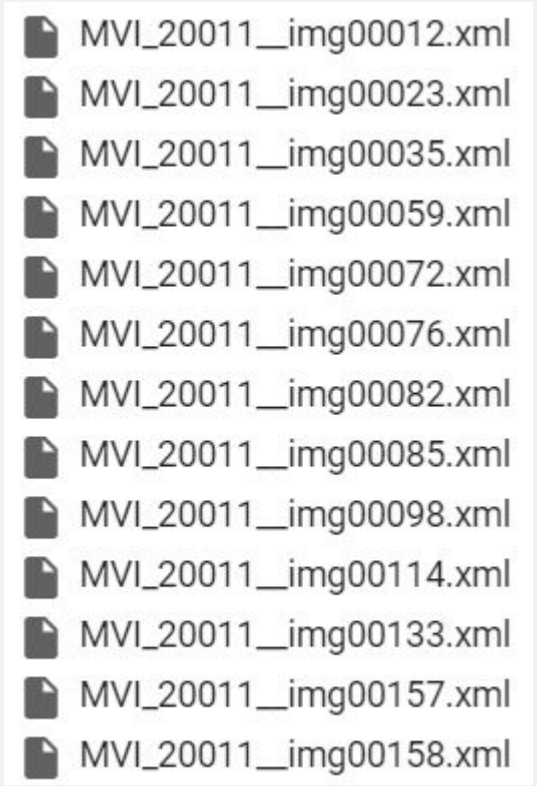

```
<?xml version="1.0" encoding="UTF-8"?>
- <sequence name="MVI_20011">
    <sequence_attribute sence_weather="sunny" camera_state="unstable"/>
  - <ignored_region>
      <box height="63.5" width="181.75" top="24.75" left="778.75"/>
      <box height="33.5" width="29.75" top="94.75" left="930.75"/>
      <box height="37.5" width="178.5" top="13.75" left="523.75"/>
      <box height="94.5" width="270.5" top="34.75" left="207.75"/>
      <box height="117.5" width="145.5" top="131.75" left="182.75"/>
      <box height="121.5" width="88.75" top="231.75" left="0.5"/>
      <box height="97.5" width="57.5" top="151.75" left="123.75"/>
    </ignored_region>
  - <frame num="1" density="7">
    - <target_list>
      - <target id="1">
          <box height="162.2" width="160.05" top="378.8" left="592.75"/>
          <attribute vehicle_type="car" truncation_ratio="0.1" trajectory_length="5" speed="6.859" orientation="18.488"/>
        </target>
      - <target id="2">
          <box height="43.06" width="47.2" top="120.98" left="557.65"/>
          <attribute vehicle_type="car" truncation_ratio="0" trajectory_length="72" speed="1.5055" orientation="19.398"/>
        </target>
      - <target id="3">
          <box height="30.08" width="35.25" top="88.27" left="545.2"/>
          <attribute vehicle_type="car" truncation_ratio="0" trajectory_length="105" speed="0.5206" orientation="2.7525"/>
        - <occlusion>
            <region_overlap height="1.52" width="27.45" top="88.27" left="553" occlusion_status="1" occlusion_id="5"/>
          </occlusion>
        </target>
      - <target id="4">
          <box height="25.925" width="28.0" top="67.5" left="508.35"/>
          <attribute vehicle_type="car" truncation_ratio="0" trajectory_length="132" speed="0.52707" orientation="349.06"/>
        </target>
      - <target id="5">
          <box height="19.695" width="29.55" top="70.095" left="553"/>
          <attribute vehicle_type="car" truncation_ratio="0" trajectory_length="151" speed="0.49822" orientation="58.543"/>
        </target>
```

04 Core Contents

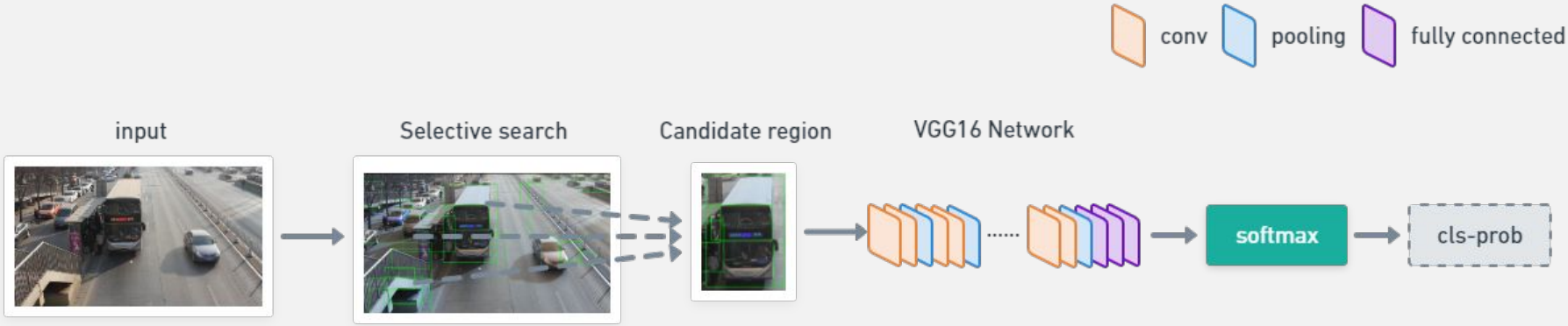
csvs ← UA-DETRAC xmls → VOC xmls





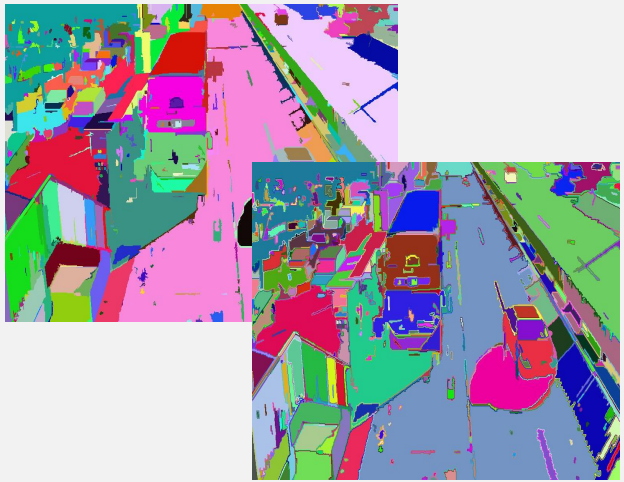


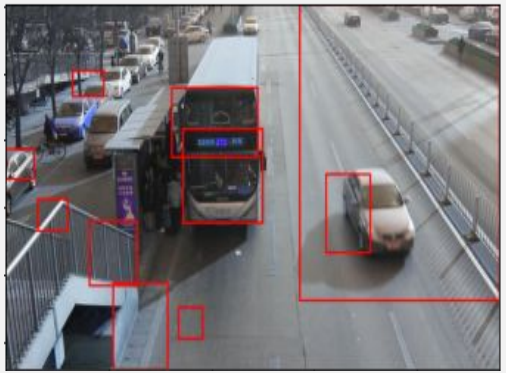
R-CNN

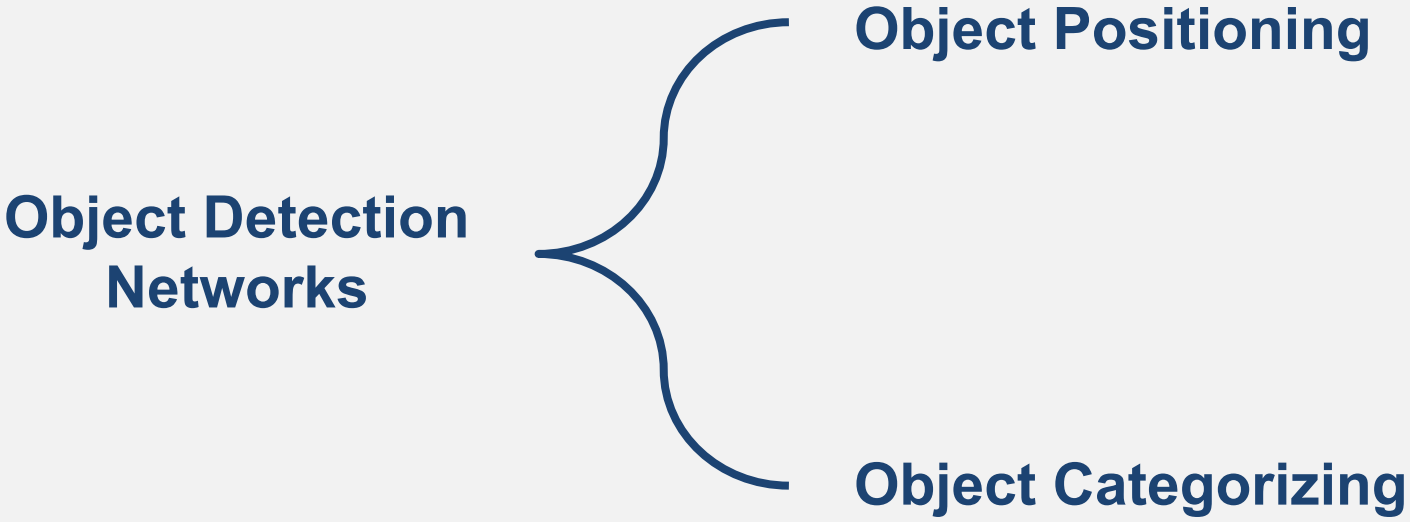

conv
pooling
fully connected
input
Selective search
Candidate region
VGG16 Network
softmax
cls-prob

Selective Search

Selective Search [5] is a hierarchical grouping-based algorithm. It splits images based on four attributes: color, texture, size, and shape.

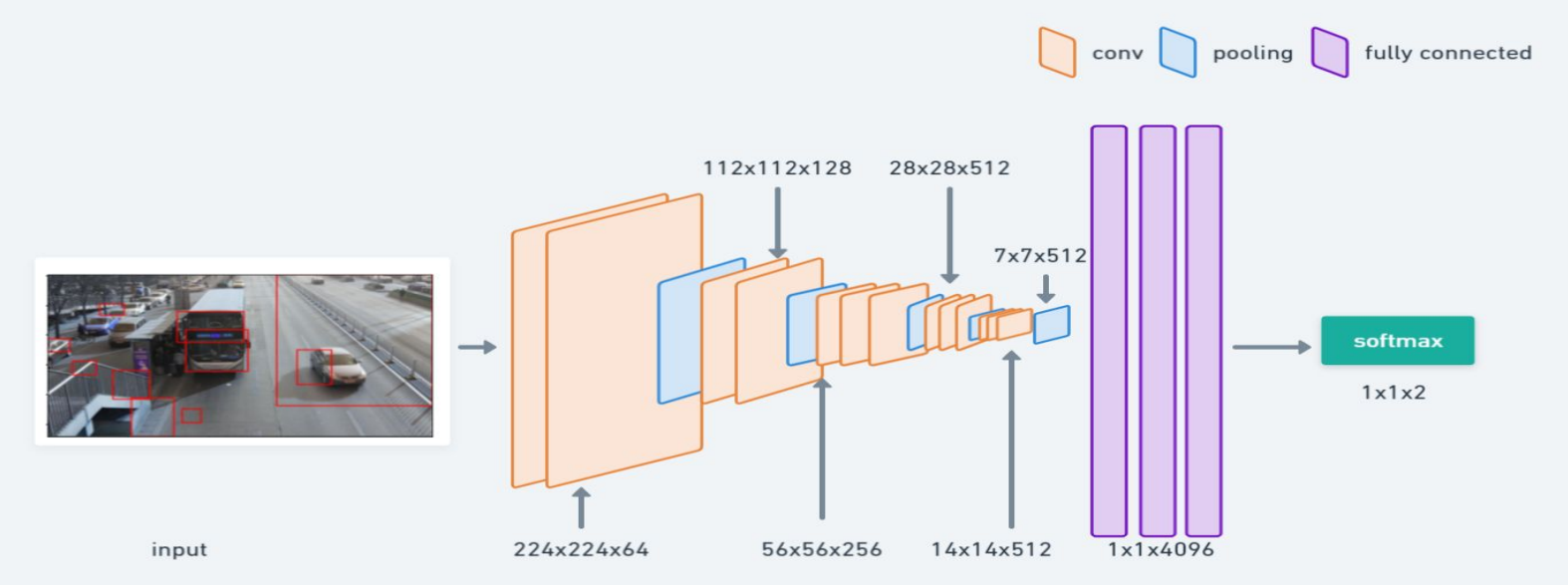




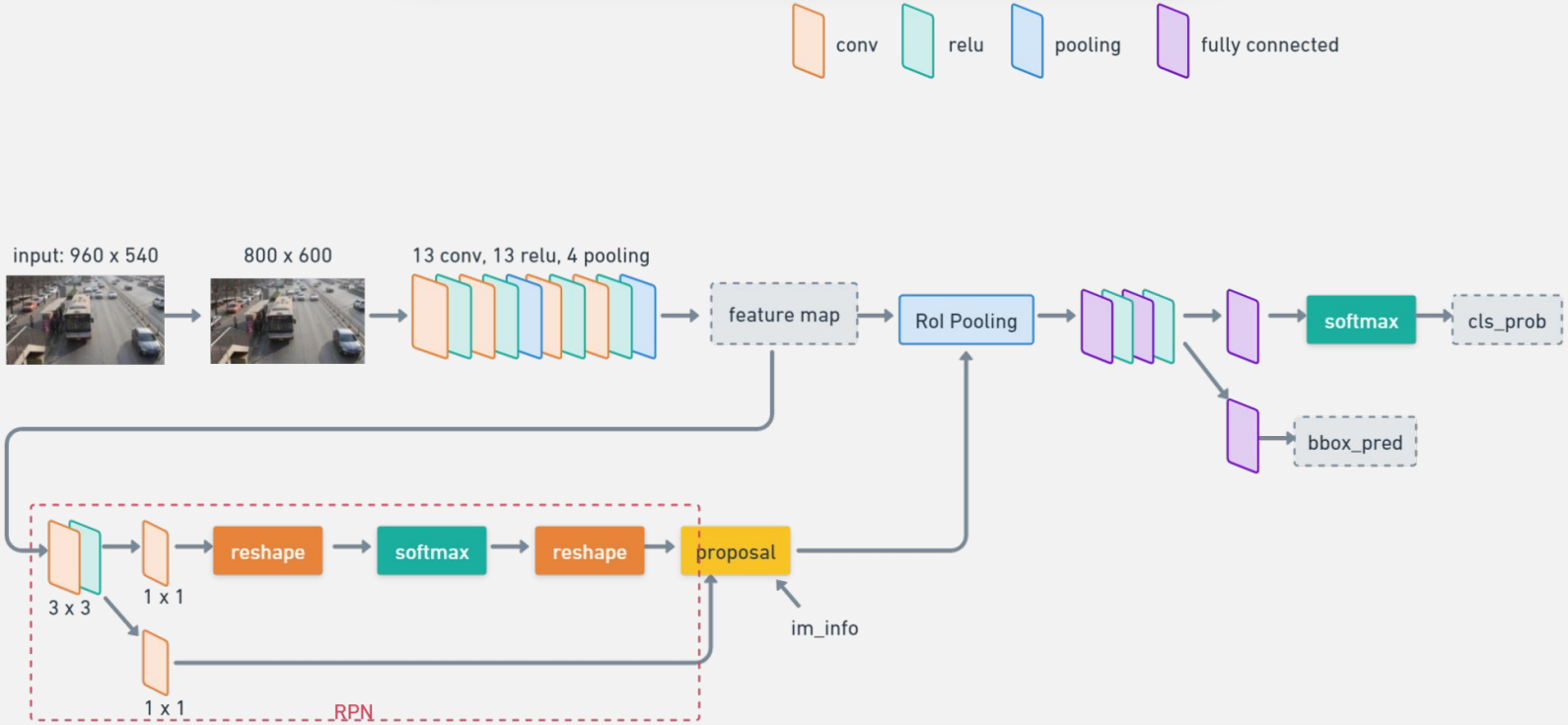

Object Detection Networks
Object Positioning
Object Categorizing

CNN(VGG16)

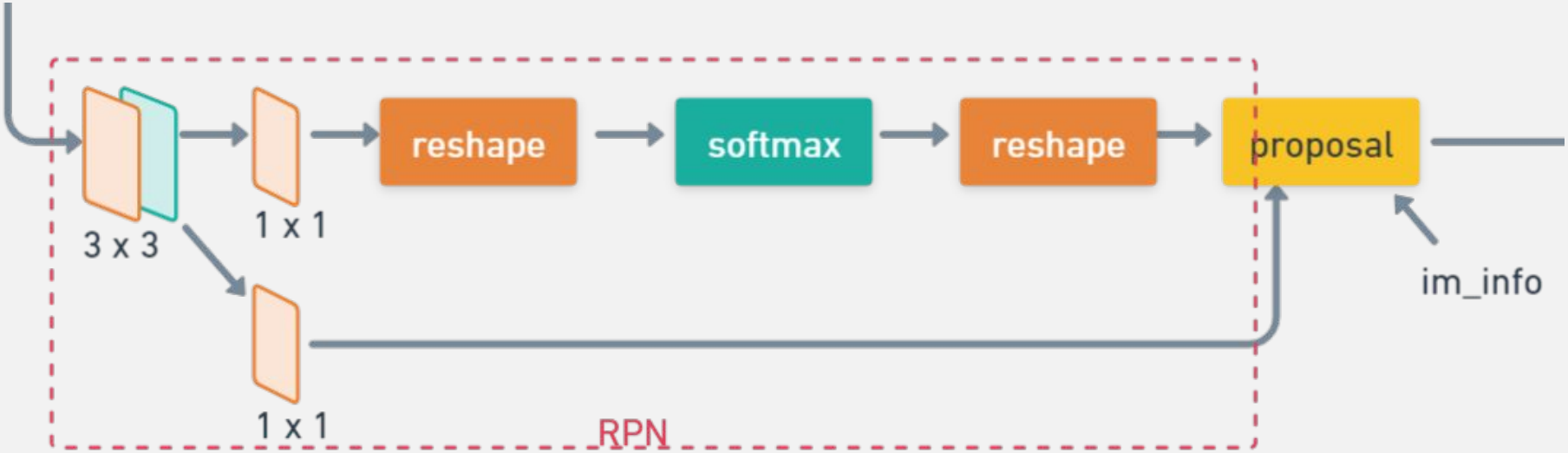
The structure of VGG16 [4] neural network is very simple, but at the same time deep enough to the number of network layers, in turn, ensures the correct classification rate.



Faster R-CNN

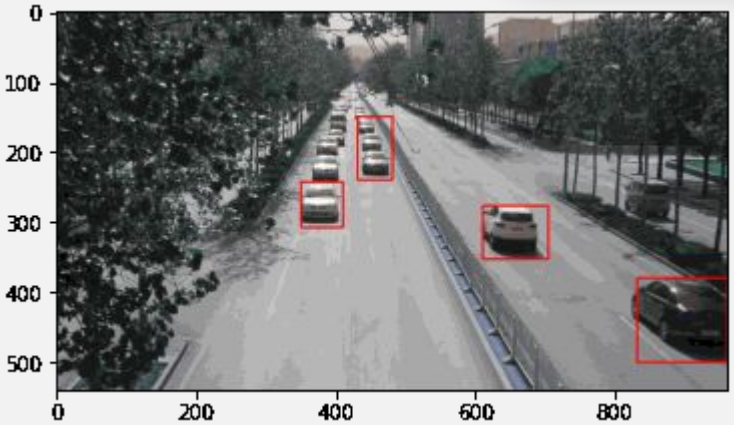

conv
relu
pooling
fully connected
input: 960 x 540
800 x 600
13 conv, 13 relu, 4 pooling
feature map
RoI Pooling
softmax
cls_prob
bbox_pred
3 x 3
1 x 1
reshape
softmax
reshape
proposal
im_info
1 x 1
RPN

RPN



05 Achievement Display

Achievement Display

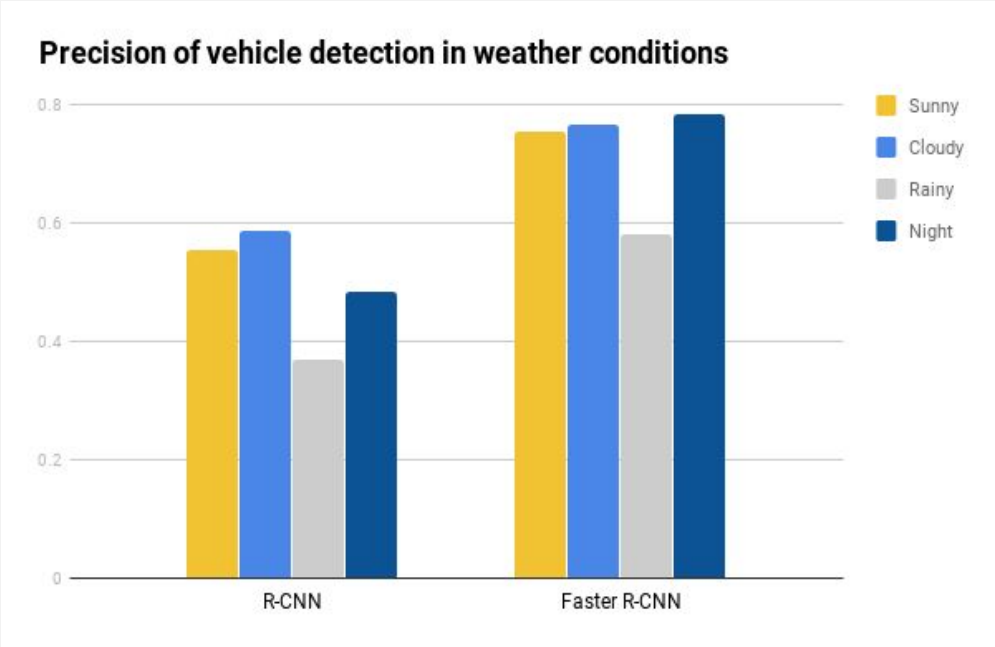


The result of R-CNN

The result of Faster R-CNN

Algorithms	Training time	Prediction time	mAP
R-CNN	7h27min	27s/Image	0.492
Faster R-CNN	32min	0.1s/Image	0.713

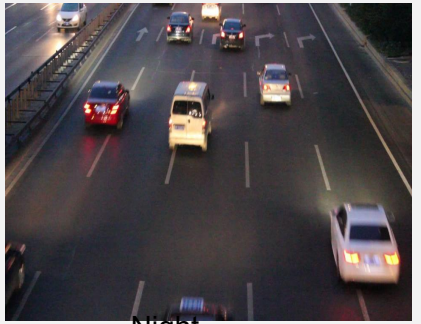
Achievement Display



Sunny

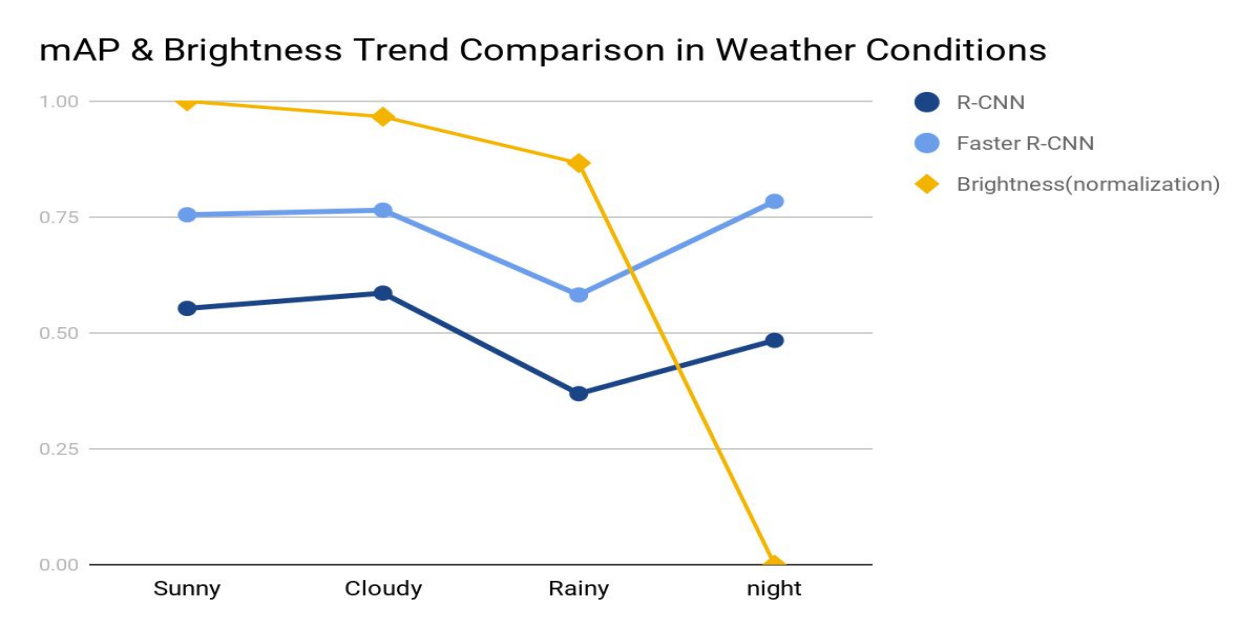
Rainy

Cloudy



Night

Achievement Display



There is no correlation between the precision of object detection and the brightness of the image.

06

Paper Summary

Paper Summary

Process	Results	Conclusion	Future work
Compared the performance of R-CNN and Faster R-CNN for the vehicle detection problem in the UA-DETRAC dataset.	1.The Faster R-CNN is faster and more precise than the R-CNN. 2. Weather conditions could affect the precision of algorithms.	1.Using the RPN instead of the Selective Search to build region proposals is the main reason 2. Brightness is not the direct reason for weather affecting precision	1.Other different object detection algorithms can be compared(e.g. YOLO) 2.Explore the reason why weather conditions have an effect on the precision

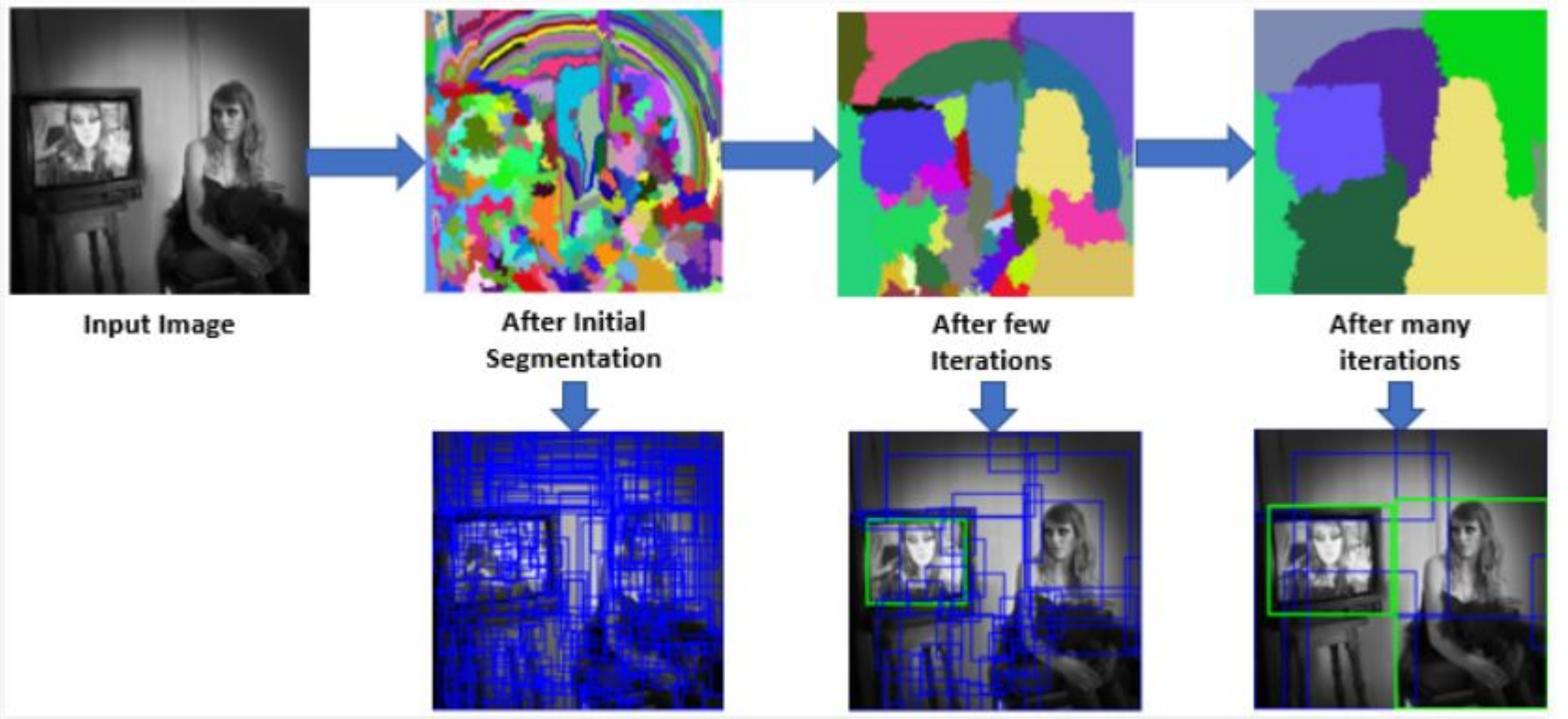
Reference

1. Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.
2. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
3. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. *In 2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
4. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv: 1409. 1556*, 2014.
5. Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. International journal of computer vision, 104(2):154–171, 2013.

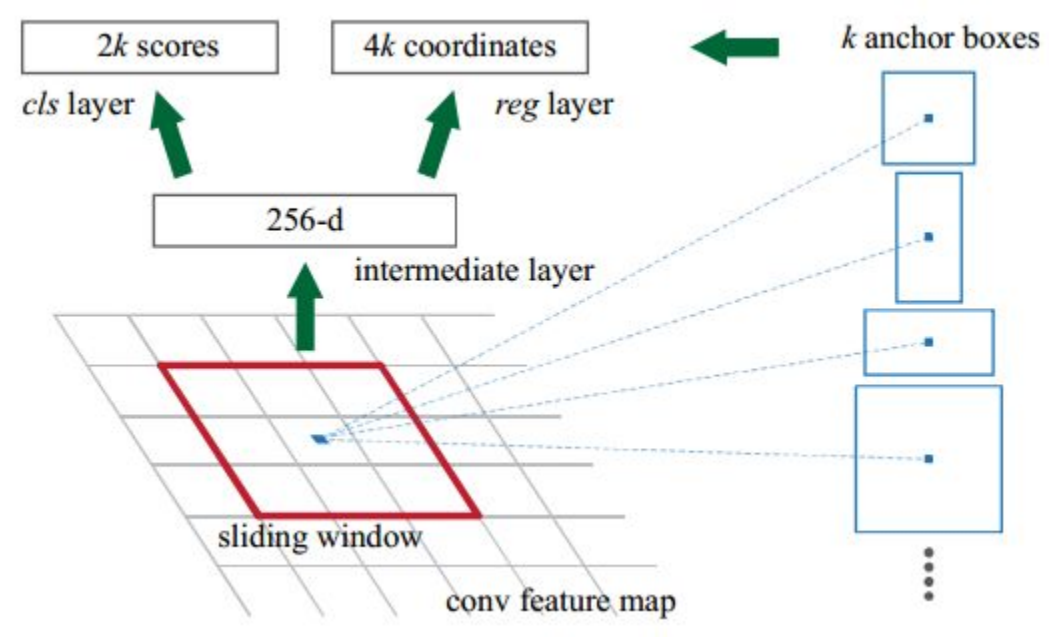
Questions?

THANKS!

Appendix


Input Image
After Initial Segmentation
After few Iterations
After many iterations

Appendix


2k scores
4k coordinates
k anchor boxes
cls layer
reg layer
256-d
intermediate layer
sliding window
conv feature map