组会报告

熊兴旺

2019/07/15

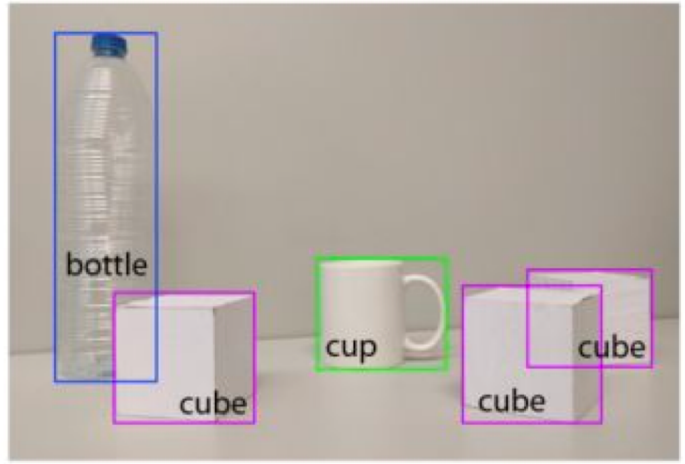
Contents

- Object Detection
- Work in Progress

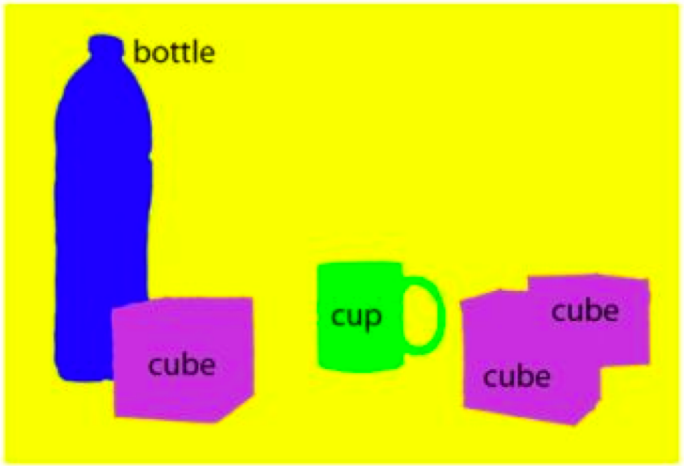
CV Tasks

- Image Classification
- Object Localization
- Sematic Segmentation
- Instance Segmentation

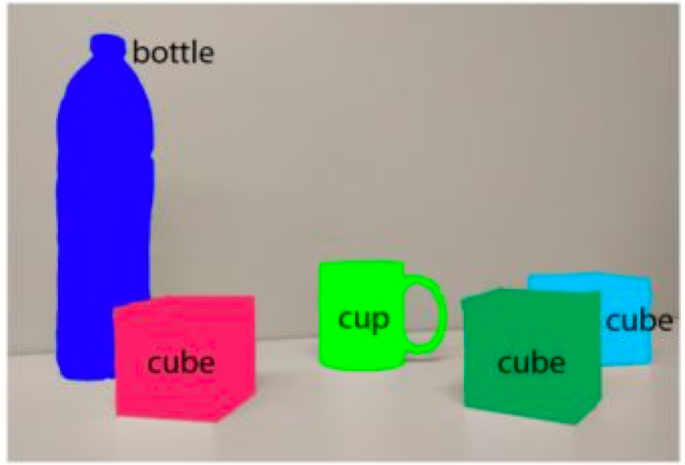
(a) Image classification



(b) Object localization



(c) Semantic segmentation



(d) Instance segmentation

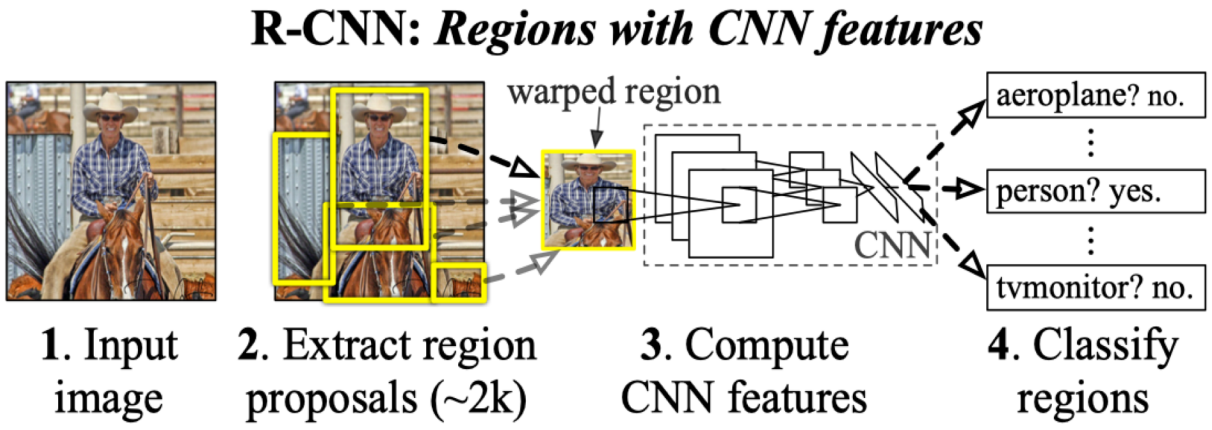
Image Classification → Object Detection

- Image Classification
  - **Backbone Network**, e.g., VGG, ResNet, DenseNet, ZFNet, etc.
  - **Fully Connected & Softmax** layers for cls.
- Object Detection
  - **Region Proposal**
  - **Detection/Classification**
  - **Bounding Box (BBox) Regression**

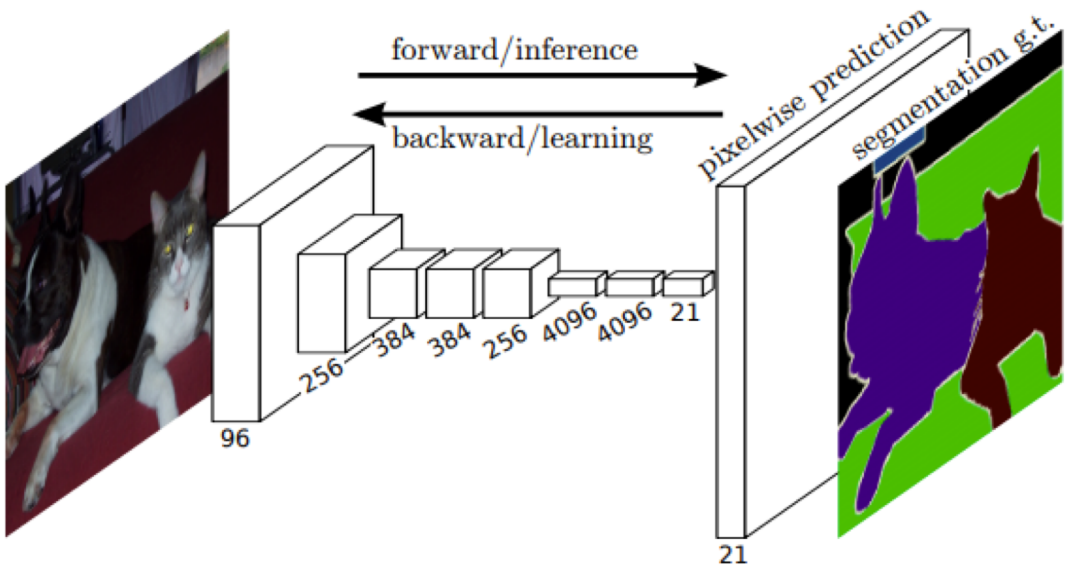
Object Detection = Localization + Classification

Object Detection → Instance Segmentation

- **Object Detection**
  - BBox-level Detection
- **Instance Segmentation**
  - Pixel-level Detection
  - **Fully Convolutional Network (FCN)**

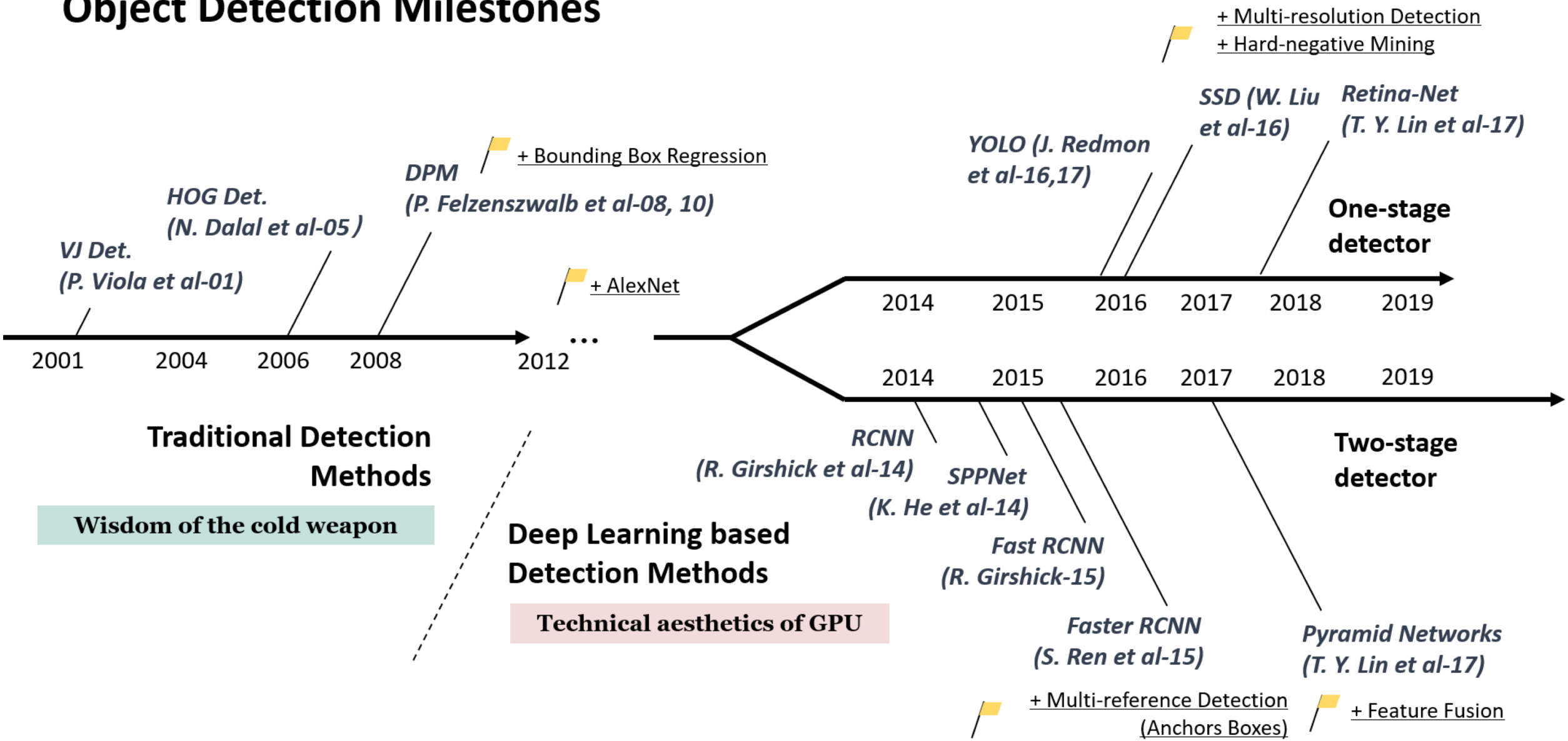


2-stage Object Detection System Overview (R. Girshick, et al, CVPR 2014)



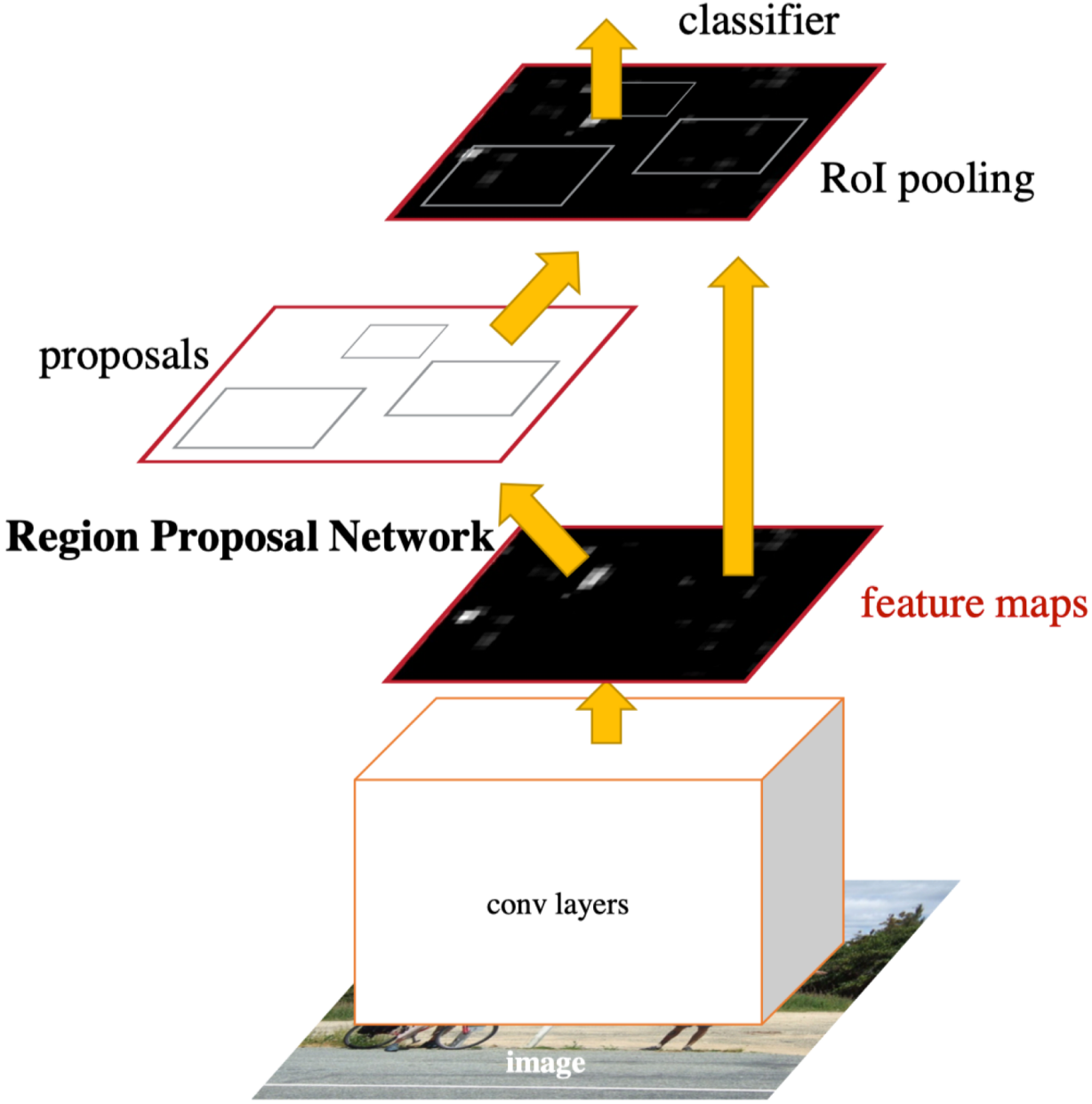
FCN Architecture (Long J, et al, CVPR 2015)

Object Detection Milestones

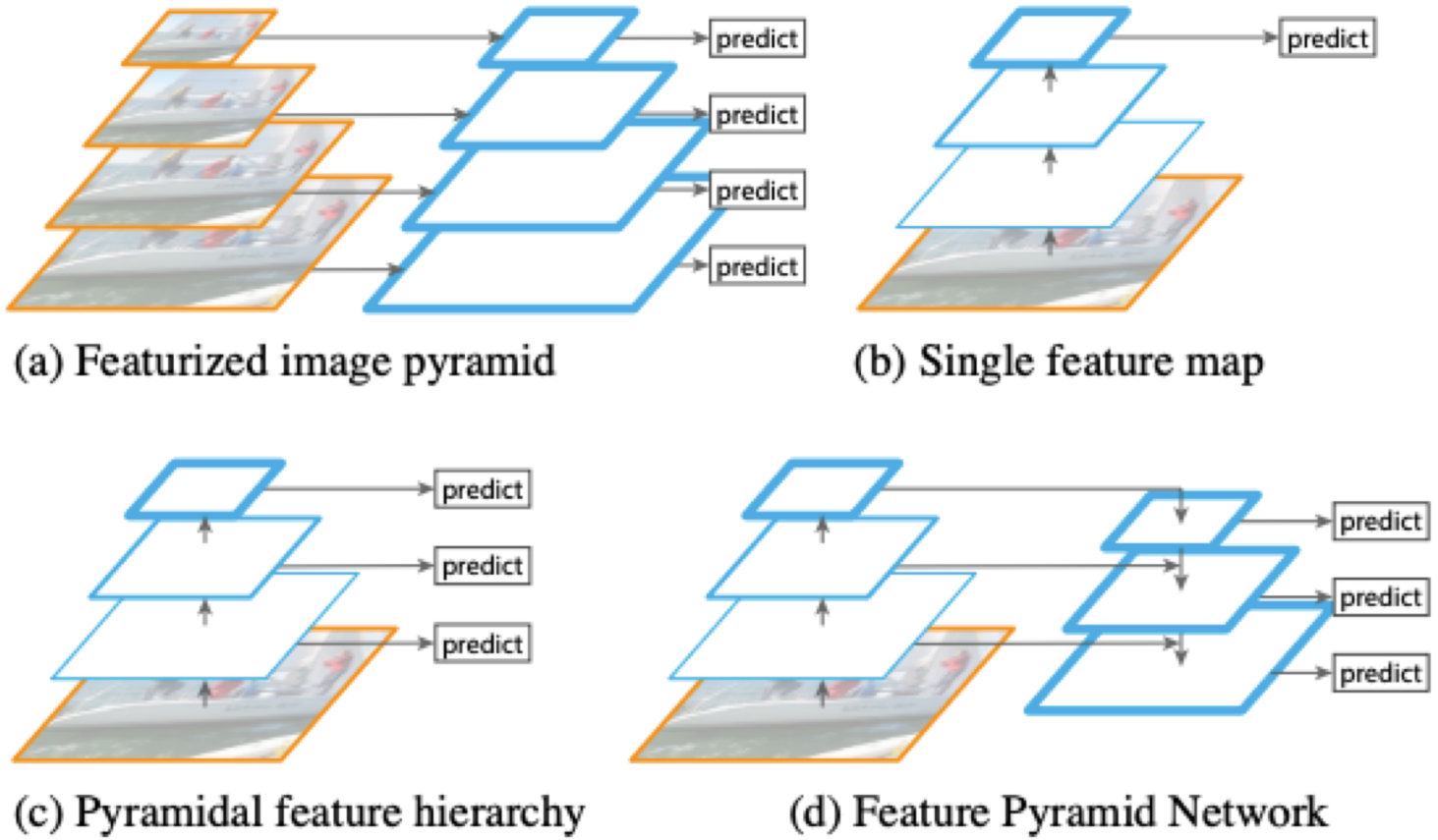


Faster R-CNN

- **用RPN替代Selective Search。**
  RPN 与分类检测网络共享卷积网络。
- **RoI Pooling 进行尺度归一化**
- 检测网络: CNN + FC + Softmax
- 性能：单GPU上，≈5fps



Multi-scale Object Detection



(a) Featurized image pyramid

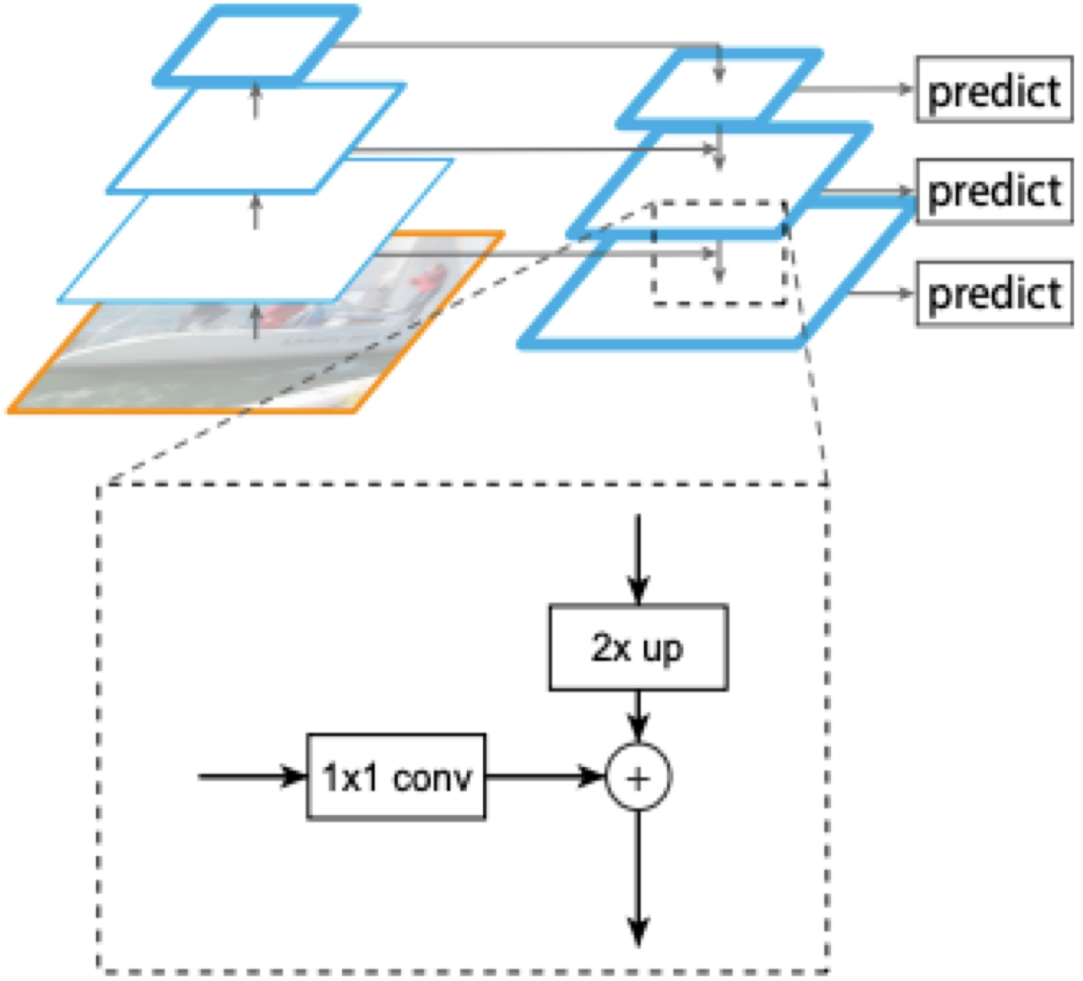
(b) Single feature map

(c) Pyramidal feature hierarchy

(d) Feature Pyramid Network

Feature Pyramid Networks

- **Downsampling**
  - **前向计算**
  - **提取深层特征，加深语义**
- **Upsampling**
  - **保留语义特征**
  - **融合细节特征/位置特征**
  - 上采样的特征与原图的低级特征**相加**



Skip Architecture

Two-stage Object Detection

- 阶段级联 (Stage Cascade)：定位 + 检测/边框回归
- **Regional Proposal** 关注**召回率 (Average Recall, abbr. AR)**，尽可能多选出目标，所以允许我们生成大量的候选框；
- **Detection** 关注**正检率 (Average Precision, abbr. AP)**，尽可能让目标分类正确。

Conclusions

- **Low-level** Vs. **High-level** feature maps.
- **Skip Architectures**: e.g., ResNet, FPN, UNet, DenseNet (concat), etc.
  - add/concat
- **Intuition** are powerful.
- **Simplicity** is the best art.
  - No bells and whistles!
  - No handcraft!

Contents

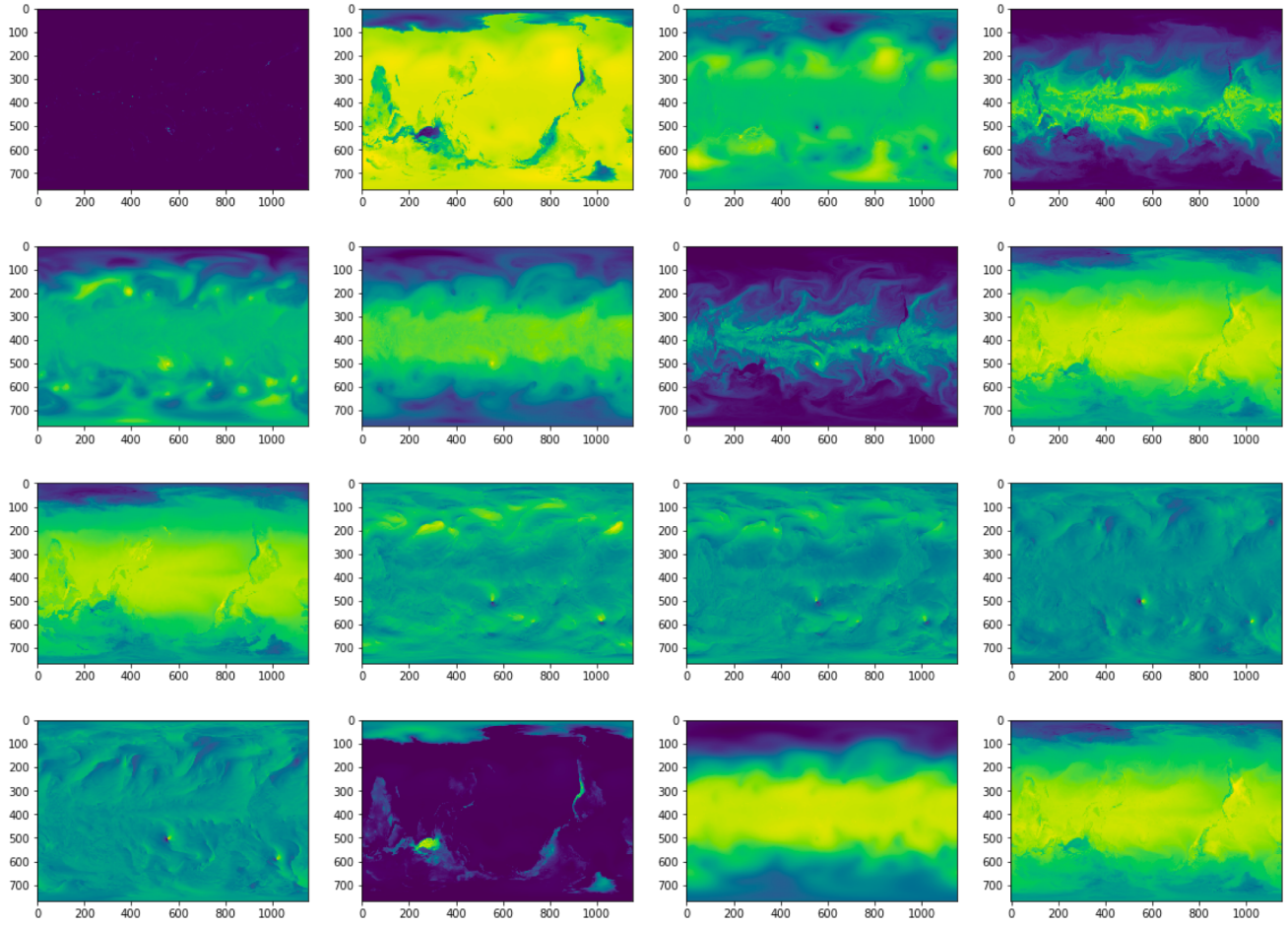
- Object Detection
- Work in Progress

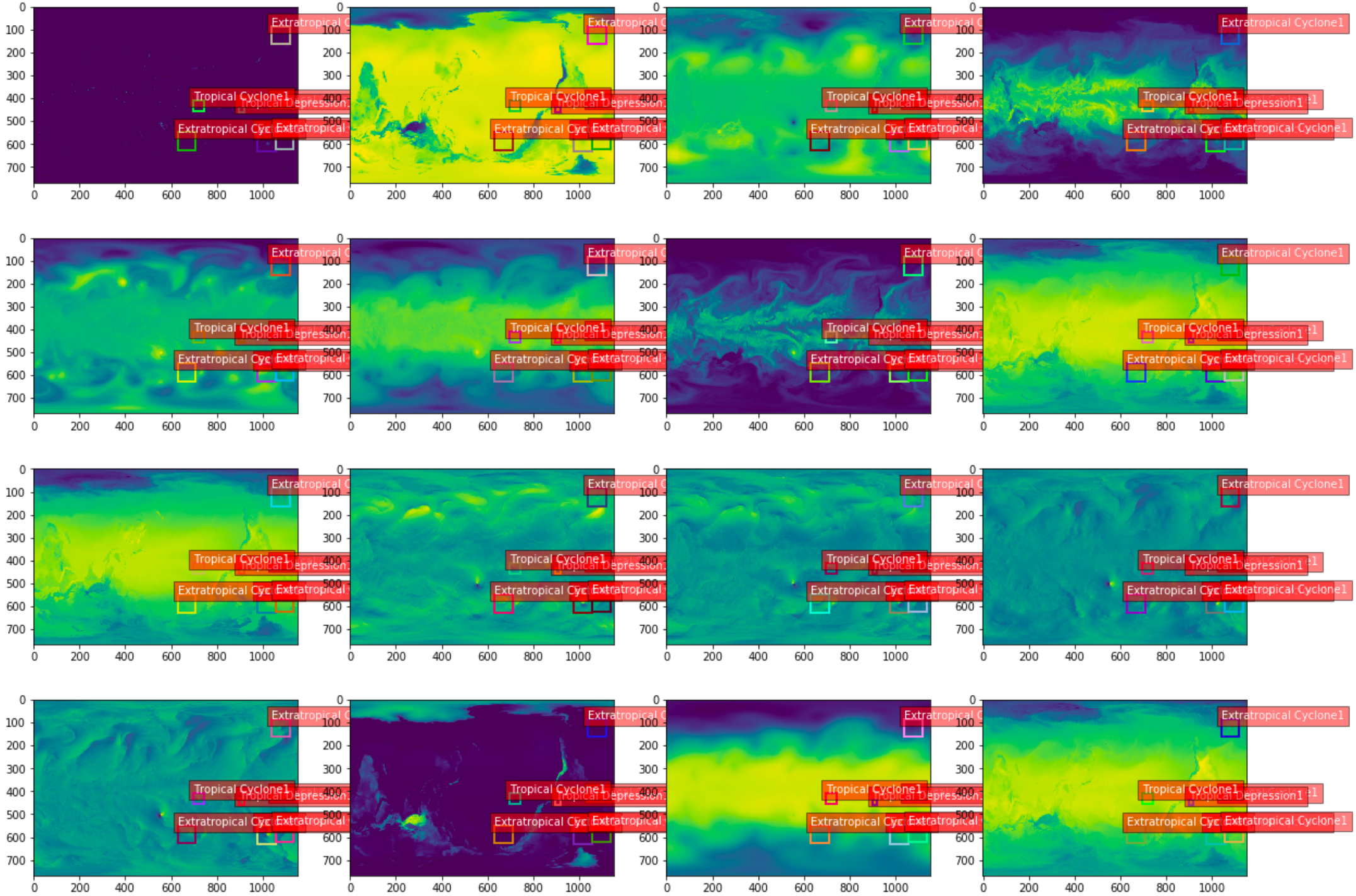
Work in Progress

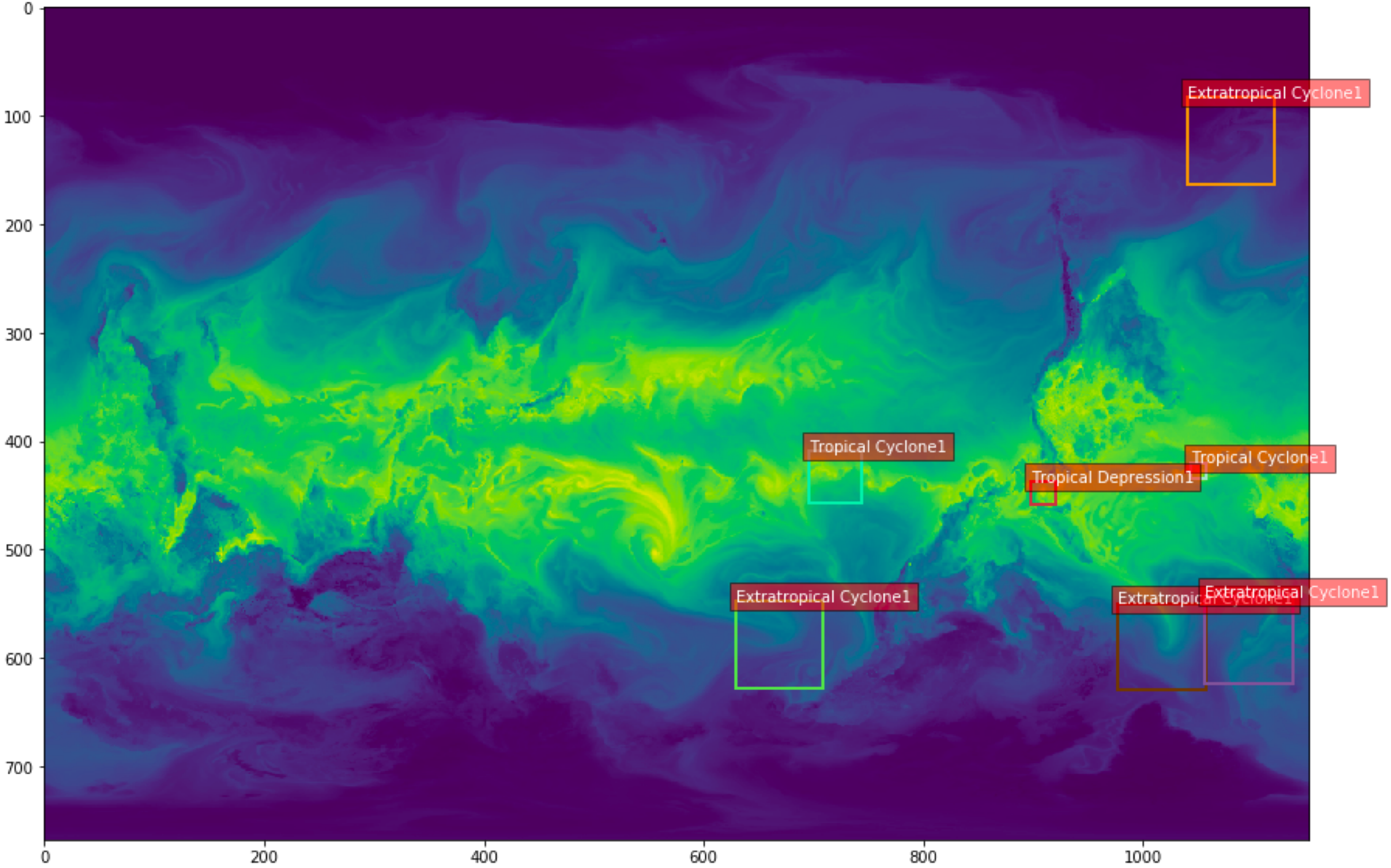
- **IoT Benchmark**: Speech Recognition (Caffe2 Version)
  - Model Transformation (PyTorch → Caffe2)
  - Inference on Mobile
- **HPC Benchmark**: Object Detection on The Extreme Weather Dataset
  - Improvement of Detection Precision, i.e., mean **A**verage **P**recision (mAP)
  - Solution for **O**ut **O**f **M**emory (OOM)

Overview of The Extreme Weather Dataset

- **Large image size**: 55MB
- **Multiple Channels**: 16 channels
- **4-category** object detection:
  - Tropical Depression (热带低气压)
  - Tropical Cyclone (热带气旋)
  - Extratropical Cyclone (温带气旋)
  - Atmospheric River (大气河)




Tropical Cyclone1
Extratropical Cyclone1


Extratropical Cyclone1
Tropical Cyclone1
Tropical Depression1
Tropical Cyclone1
Extratropical Cyclone1
Extratropical Cyclone1
0
100
200
300
400
500
600
700
0
200
400
600
800
1000

Our Solution: Faster R-CNN

- Tensorpack + Horovod
- Backbone: **ResNet-50 + FPN**
- No pretrained model
- Distributed training

Low Detection Precision!

Our Challenges

- Totally different from ImageNet, **does transfer Learning works?**
- Little understanding of the dataset. We don’t know how to evaluate the test results manually.
- Large image size may cause **memory errors**, i.e., OOM errors.
  - We might need to reduce the size of the model.
- There might be some bugs to fix!

Thanks