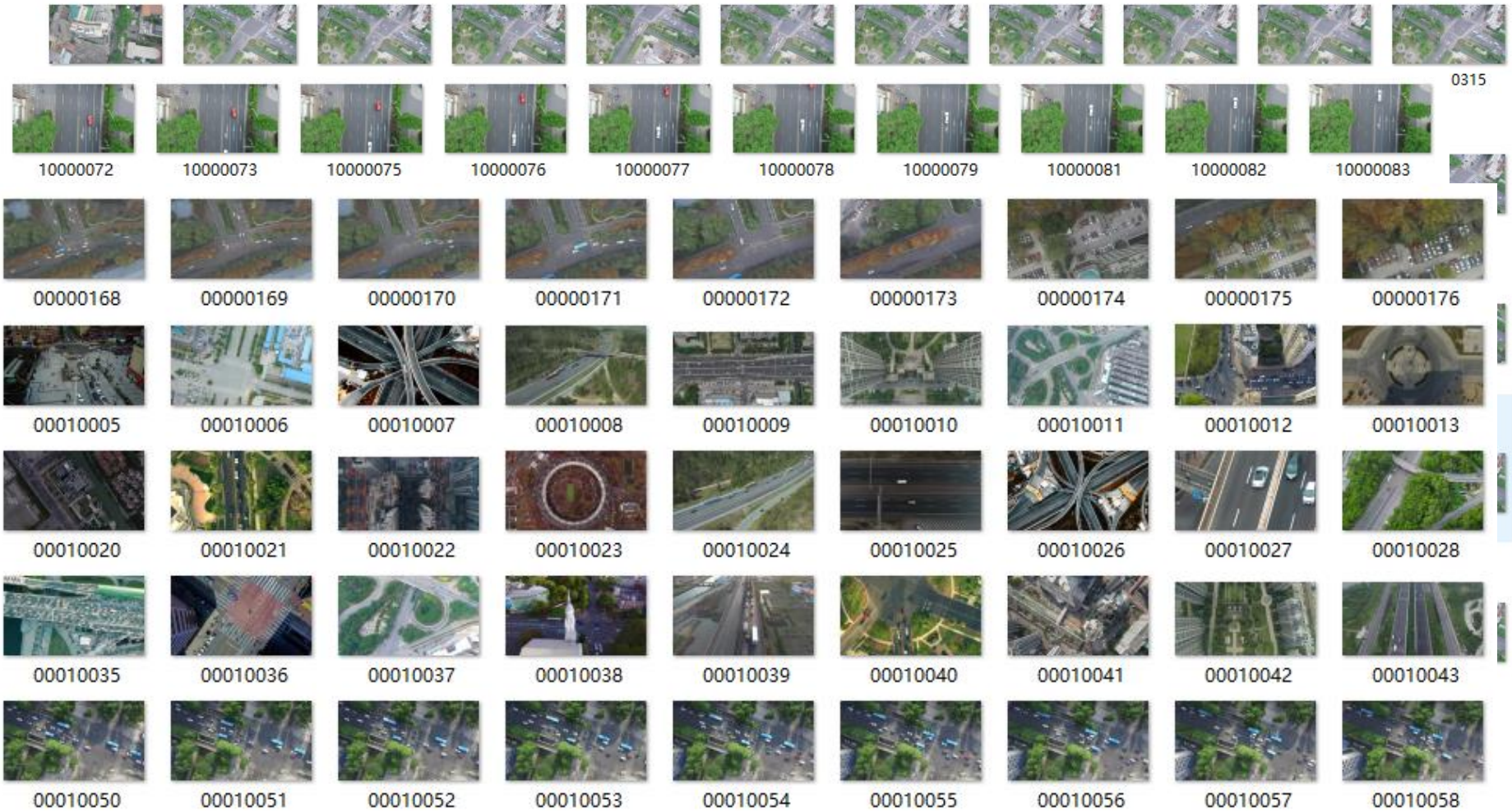
Vehicle Detection in Aerial Images

Hao Xu

05/09/2019

Dataset Review



1. imbalanced, miss the detection of hard samples
2. input resolution ranges from 400 ~ 6000: different input resolution for inference and training.

Retinanet

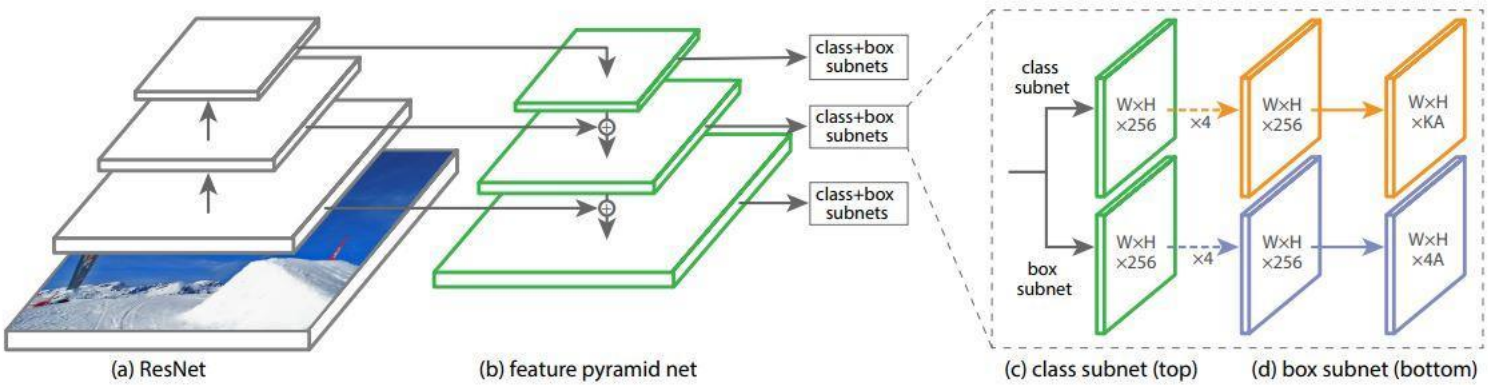


Figure 3. The one-stage **RetinaNet** network architecture uses a Feature Pyramid Network (FPN) [20] backbone on top of a feedforward ResNet architecture [16] (a) to generate a rich, multi-scale convolutional feature pyramid (b). To this backbone RetinaNet attaches two subnetworks, one for classifying anchor boxes (c) and one for regressing from anchor boxes to ground-truth object boxes (d). The network design is intentionally simple, which enables this work to focus on a novel focal loss function that eliminates the accuracy gap between our one-stage detector and state-of-the-art two-stage detectors like Faster R-CNN with FPN [20] while running at faster speeds.

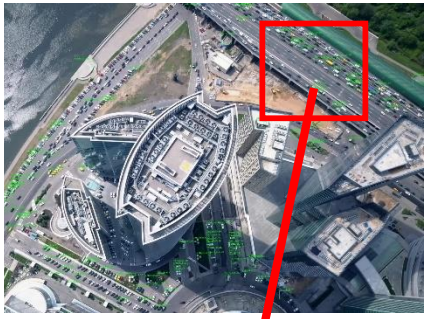
Highly effective and efficient one-stage object detector:

- ResNext101 + FPN + 2 FCN detection heads
- Focal loss, balance the background and foreground samples

Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.

Overlapped Slide Windows

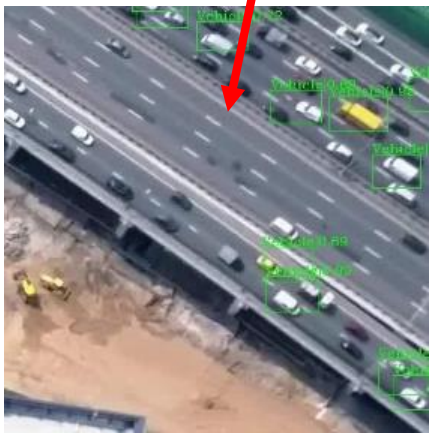
- Keep consistent input resolution for training and inference



original input

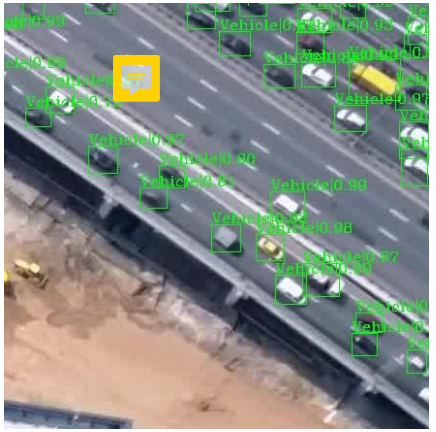
zero pad outside the window

slide window



<

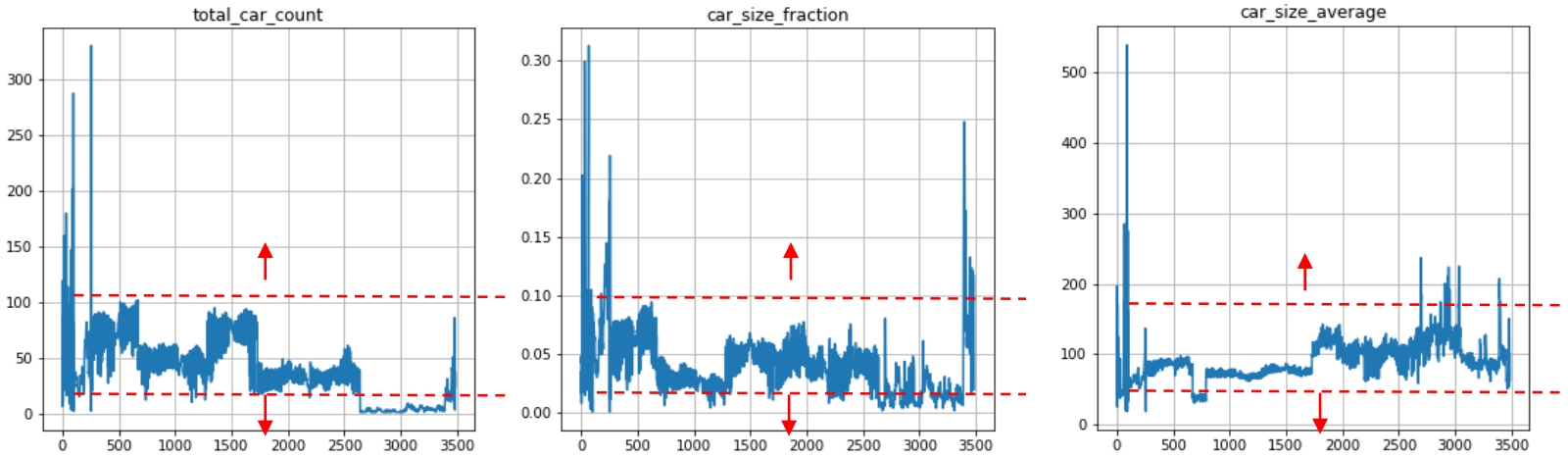
≈



The large receptive field of the network can compound some noise when resolution increases too much than that in the training phase.

Hard Sample Resampling

- Using statistical method to determine hard samples

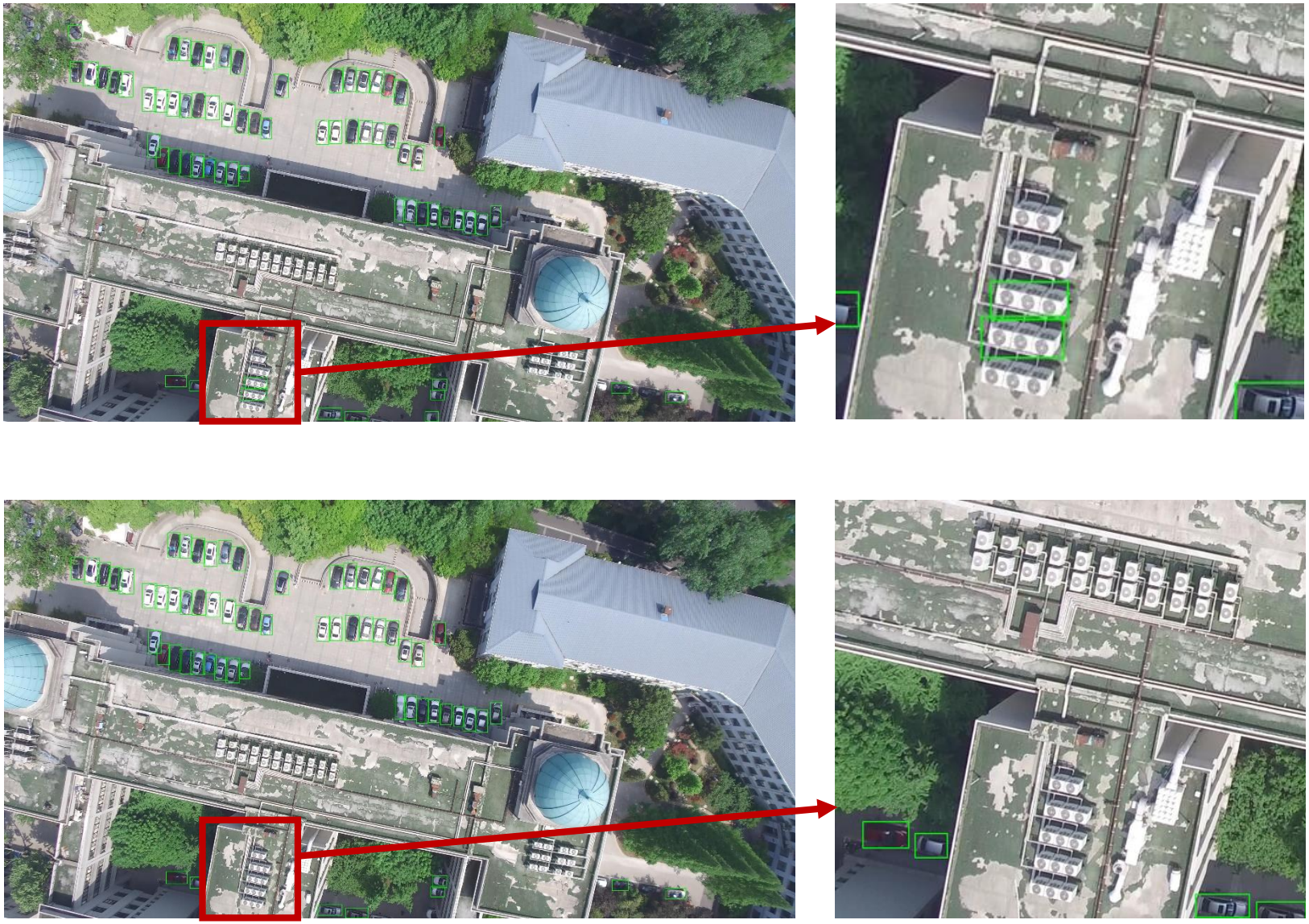


Criteria for difficulty sample:

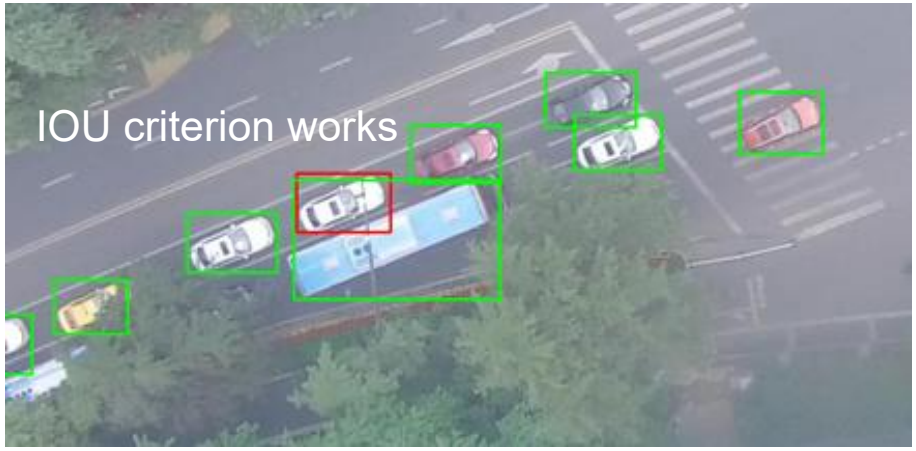
- number of cars: too large or too small
- foreground area ratio: too small or too large
- car size: too large or too small

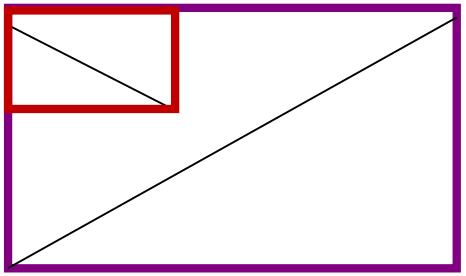
Hard Sample resampling

reduce false response



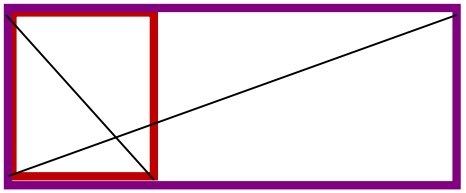
Modified NMS- topdown perspective





True

Overlapping boxes that are close to three sides are impossible in this view. The the inside box must be false detection.



False

Other Techniques

FPN	Fuse the low level features and high-level features from different stages. Improve the accuracy on small objects.
Backbone	Higher accuracy on ImageNet (85.4% top-1)
Focal Loss	Suppress loss of the easily classified samples and enlarge those hard samples.
Data Augmentation	Random rotation, random crop, random resize.
Guided Anchor	Using low level feature map to guide the anchor generation.

Result

- Train set: 4484；Test set: 3871
- 10 Fps inference speed for 1080 input resolution on NVIDIA-1080ti.
- Achieve 93.6 F1-score and rank 2nd place over 132 teams.