
# Fast and Efficient Model for Real-Time Tiger Detection In The Wild

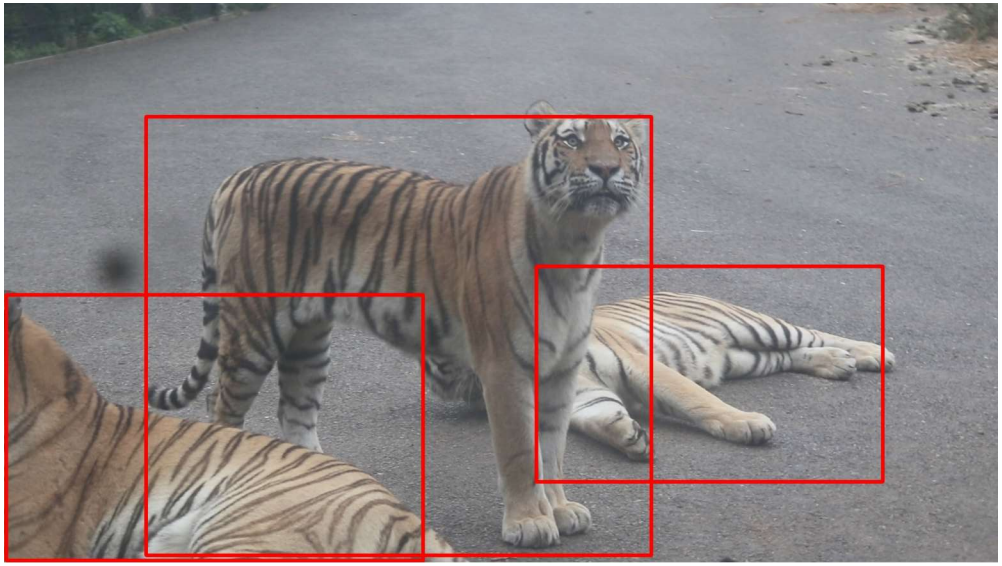
Dmitry Pranchuk,
WANNABY

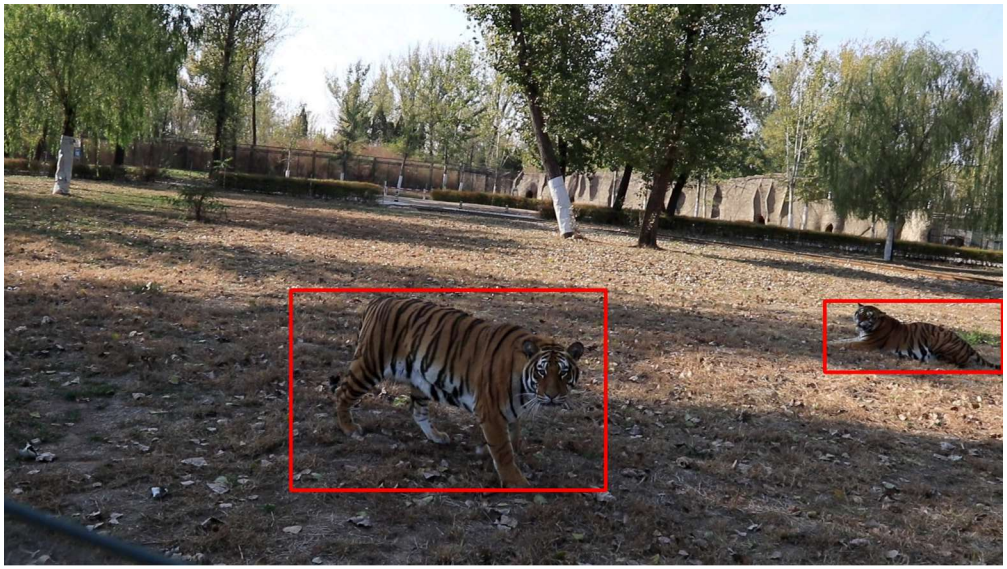
Orest Kupyn,
Faculty of Applied Sciences, UCU

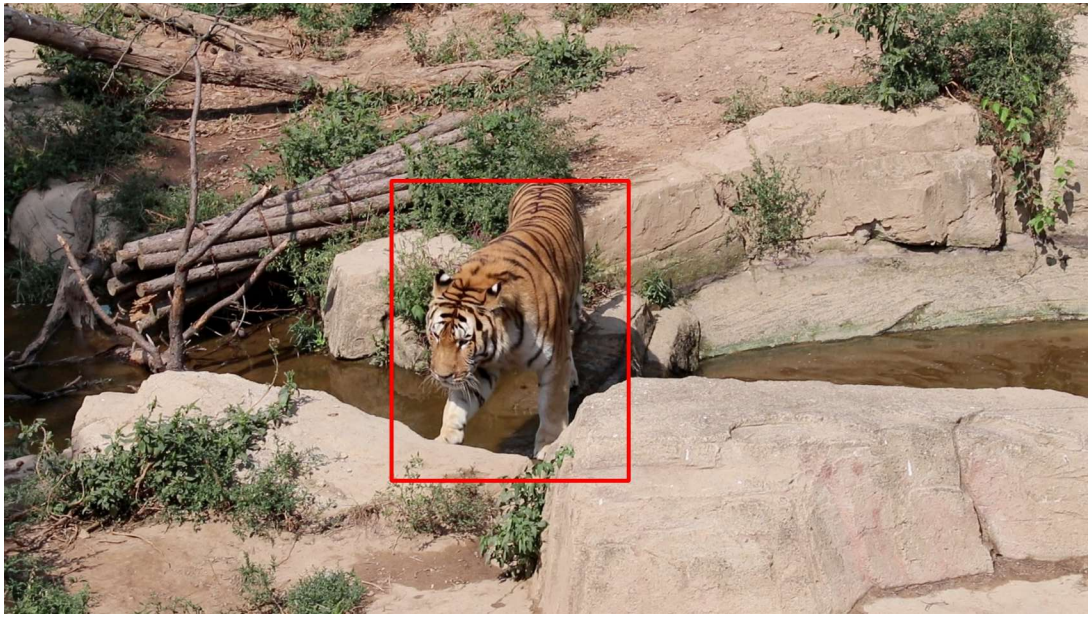

# Agenda

1. Task: detect tigers on images with bounding boxes
2. Data: 4,434 train images of high resolution (1920x1080) with 9,496 bounding boxes
3. Metric: mix of mAP and model FLOPs

# Examples



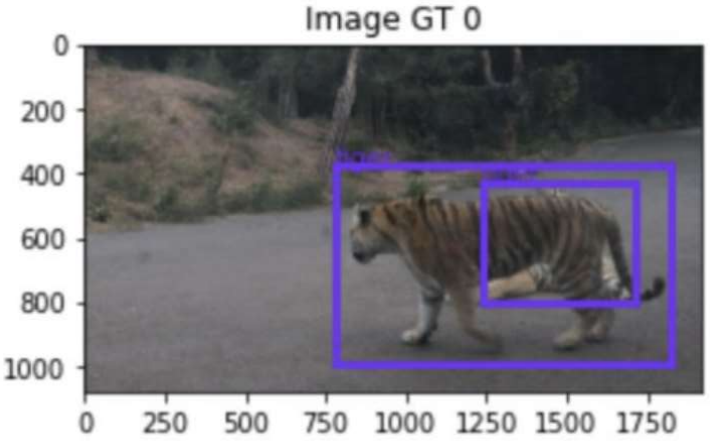


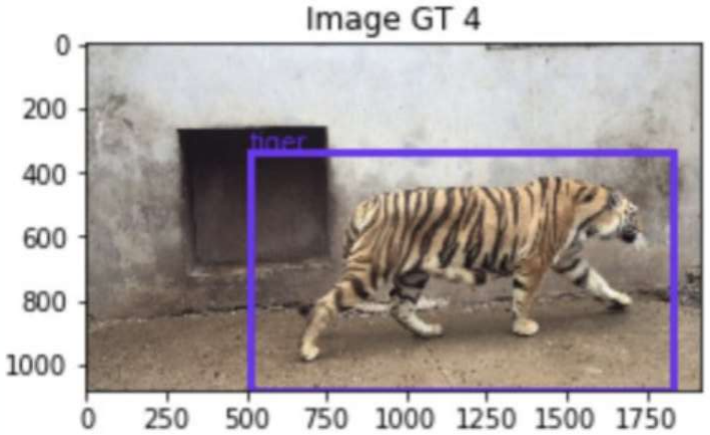


# Problems

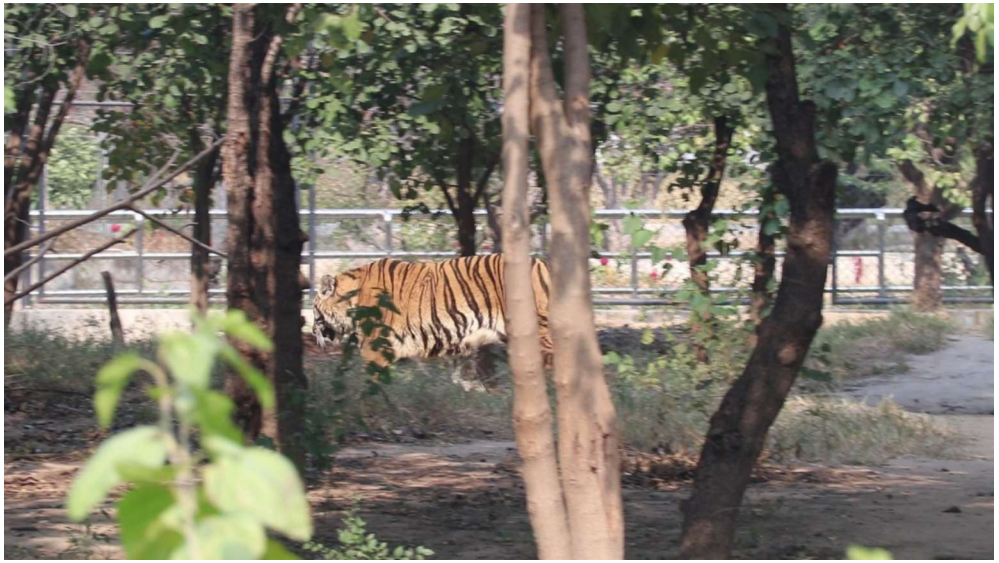
1. Small dataset. 4,434 is not so much and more data can significantly improve result
2. All images were collected from ~10 zoos. It leads to poor variety of data
3. Not always accurate labels
4. Occluded tigers, tigers in bad illumination

# Not accurate labels



# Hard cases

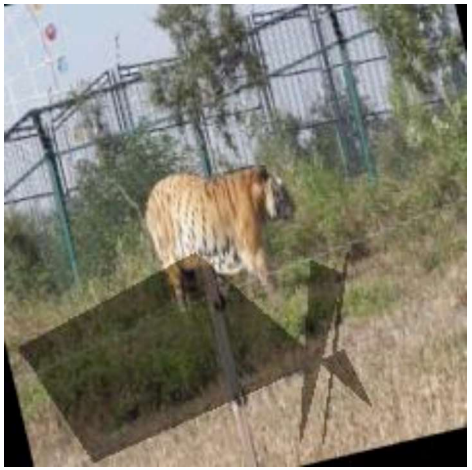
# Baseline

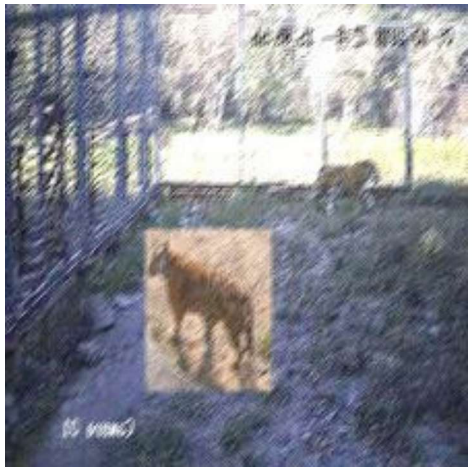
1. Model. RetinaNet +  FPN and SE-ResNeXt-101 pretrained on ImageNet
2. Augmentations. Weak affine (small rotations + flips),
   color (contrast, brightness) and  blur augmentations
3. NMS. Standard non-maximum-suppression method
4. Images resolution. 320x320
5. Metric. 0.50 mAP

# Improvements for Baseline

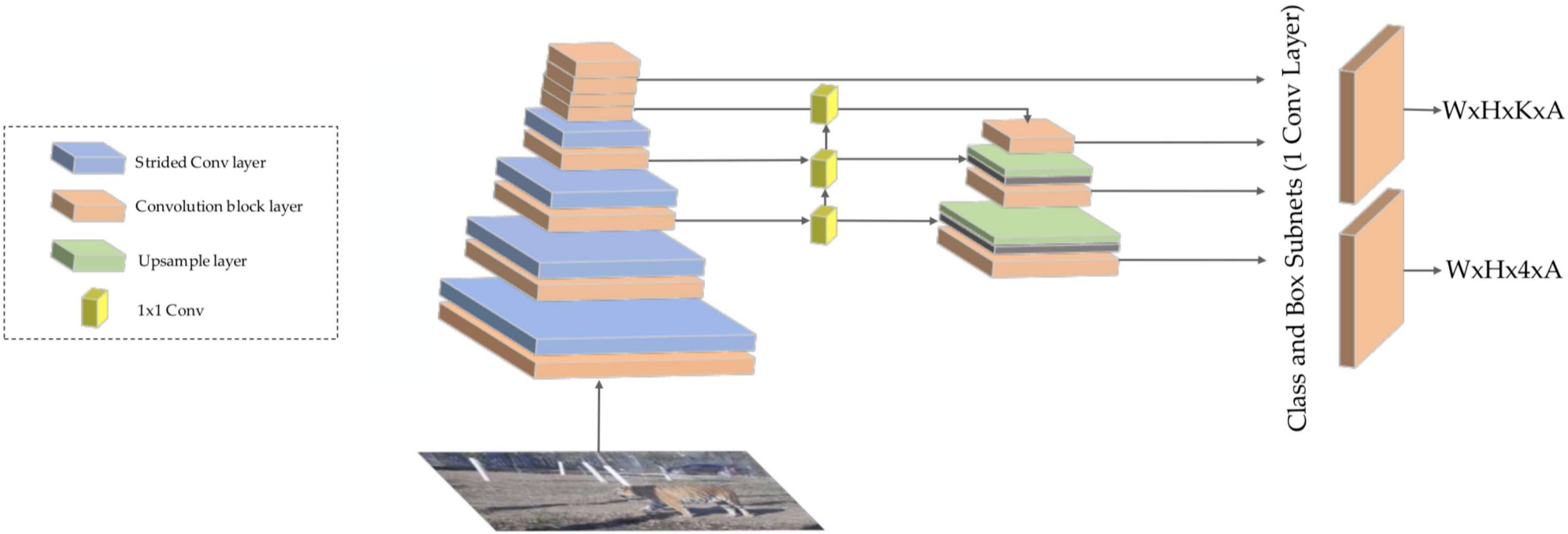
1. Model. Improved RetinaNet (remove poolings -> bigger output heatmaps)
2. Augmentations. Strong augmentations
   - Affine. 360-degrees rotations, shifts, resizes
   - Blur. Aggressive different blur (Gaussian, Motion, Median)
   - Color. Aggressive color transforms: gamma, brightness, contrast
   - Smart augmentations. Rain imitation, sun flare, shadows, cut and insert tigers from one images to others
3. NMS. Mean-NMS (calculate mean value of boxes instead of simple suppression)
4. Metric. 0.61 mAP

# Images After Hard Augmentations





# Architecture Improvements



# FLOPs Optimization

1. Tiny backbone. Different versions of MobileNets. FD-MobileNet finally
2. 1x1 convolutions instead of 3x3 for RetinaNet layers
3. Smaller images. 224x224

# Ablation Study

| | mAP | GFLOPs |
|---|---|---|
| RetinaNet-MobileNetv2 | 0.33 | 0.9 |
| +hard augmentations | 0.42 | 0.9 |
| +mean NMS | 0.43 | 0.9 |
| +slimmer RetinaNet layers and remove pooling | 0.511 | 0.64 |
| + separable convolutions in RetinaNet layers and FD-MobileNet | 0.489 | 0.071 |
| **+ pseudo-labels** | **0.515** | **0.071** |

# Thanks for Attention!

project code is available at:
github.com/KupynOrest/AmurTigerCVWC