March Madness Series: Neptune - Facial Detection and Re-Identification Marathon Match

Miras Amir
X5 Retail Group

Problem

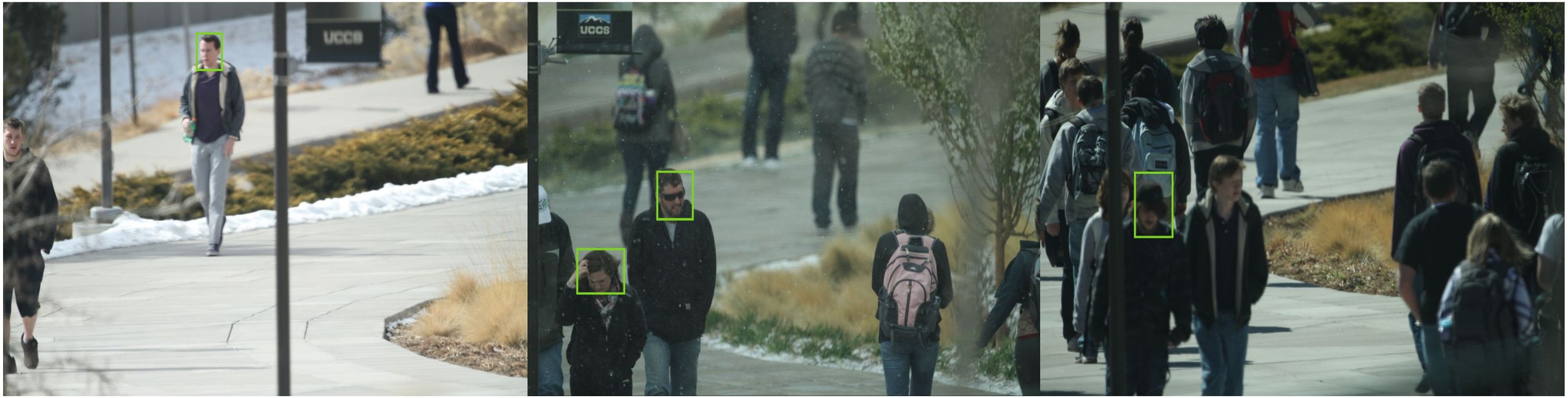
Task #1: face detection

Task #2: face detection and re-identification of found faces

- Metric: mean Average Precision (mAP)

https://topcoder.com/challenges/30086997
https://topcoder.com/challenges/30086998

Data



- UnConstrained College Students (UCCS) Dataset
- **10k** images in a train set, **2.6k** images in a test set
- **70k** faces, **1732** identifiers
- Unknown faces are marked up with class **-1**
- Average image size: **5184 x 3456**

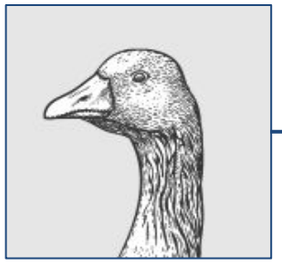
Submission format

Public Leaderboard

CSV solution.csv

/code

- Dockerfile
- train.sh
- test.sh
- model.pth
- <your code>



deadline

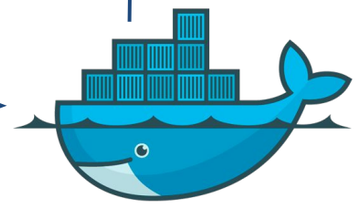
Private Leaderboard

CSV solution.csv

1 days: test.sh <data>

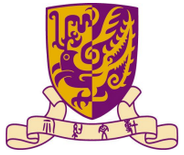
3 days: train.sh <data>

topcoder™




0.9
Face Detection

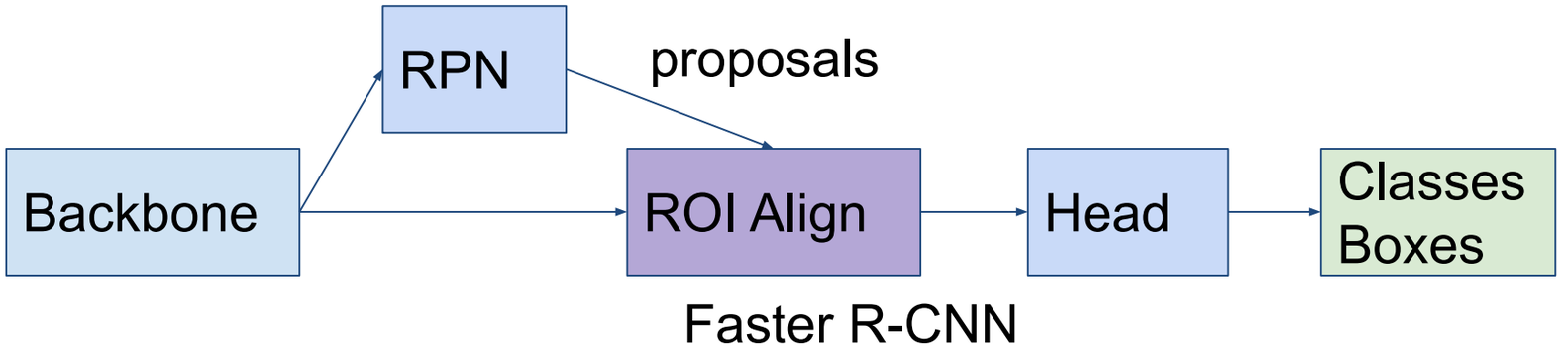
Framework

	owner	code	fp16	SOTA model	box AP
maskrcnn-benchmark		+	+	MaskRCNN X-101-32x8d-FPN	42.2
mmdetection		+	-	Cascade MaskRCNN with DCN R-101-FPN	45.8

https://github.com/facebookresearch/maskrcnn-benchmark
https://github.com/open-mmlab/mmdetection

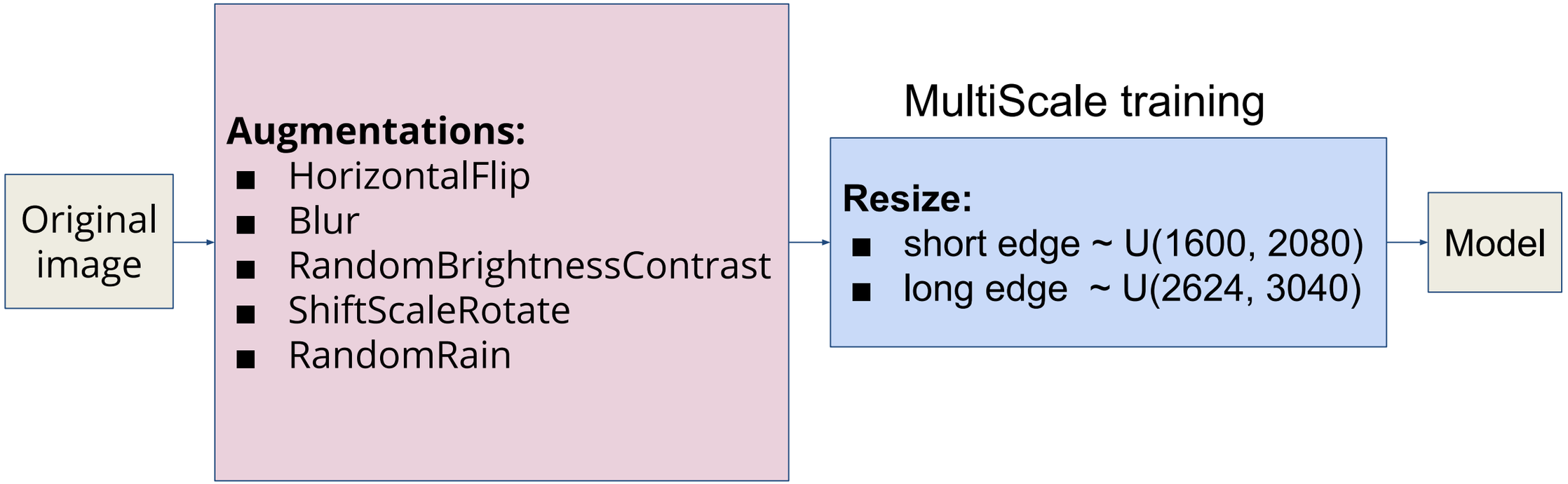
Detection baseline

- Faster R-CNN with ResNet50 backbone
- Pre-train from MS COCO
- Image size: 2048 x 1024
- Optimizer: SGD
- Post-processing: Non Maximum Suppression (NMS)
- Public LeaderBoard Score: **49.126**



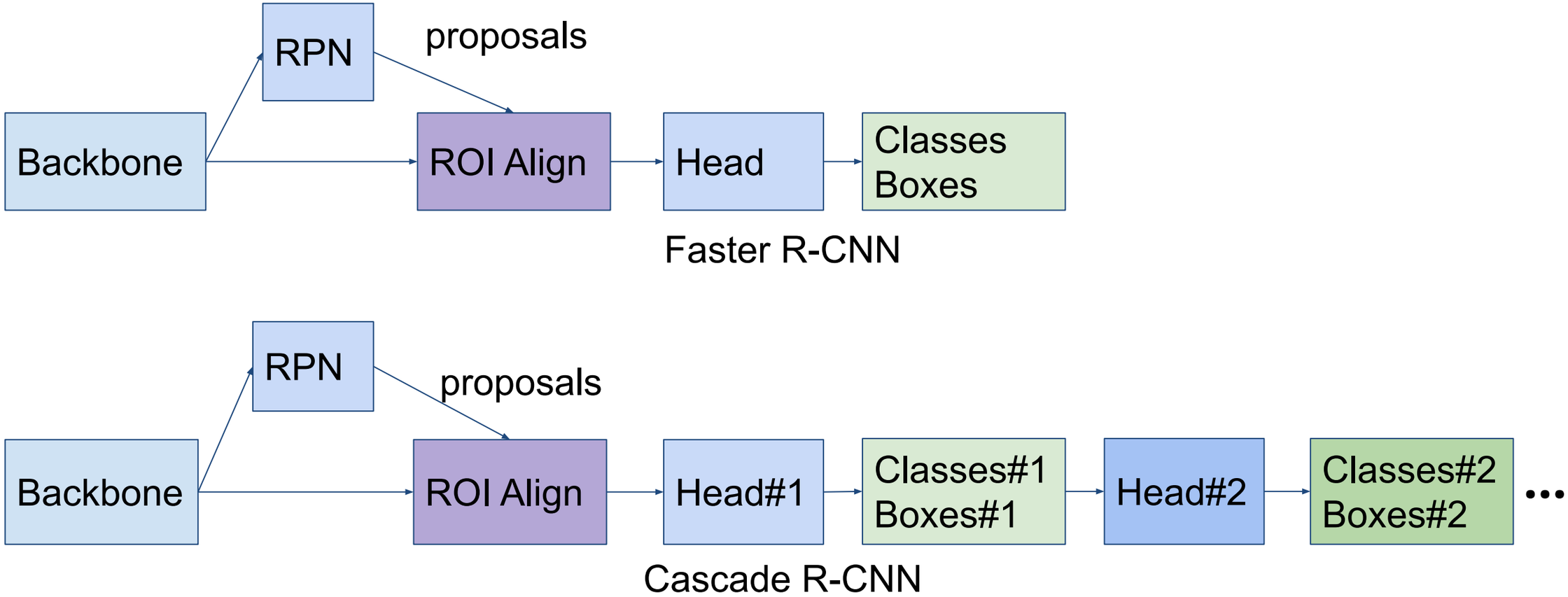
https://arxiv.org/abs/1506.01497

Preprocessing



https://github.com/albu/albumentations

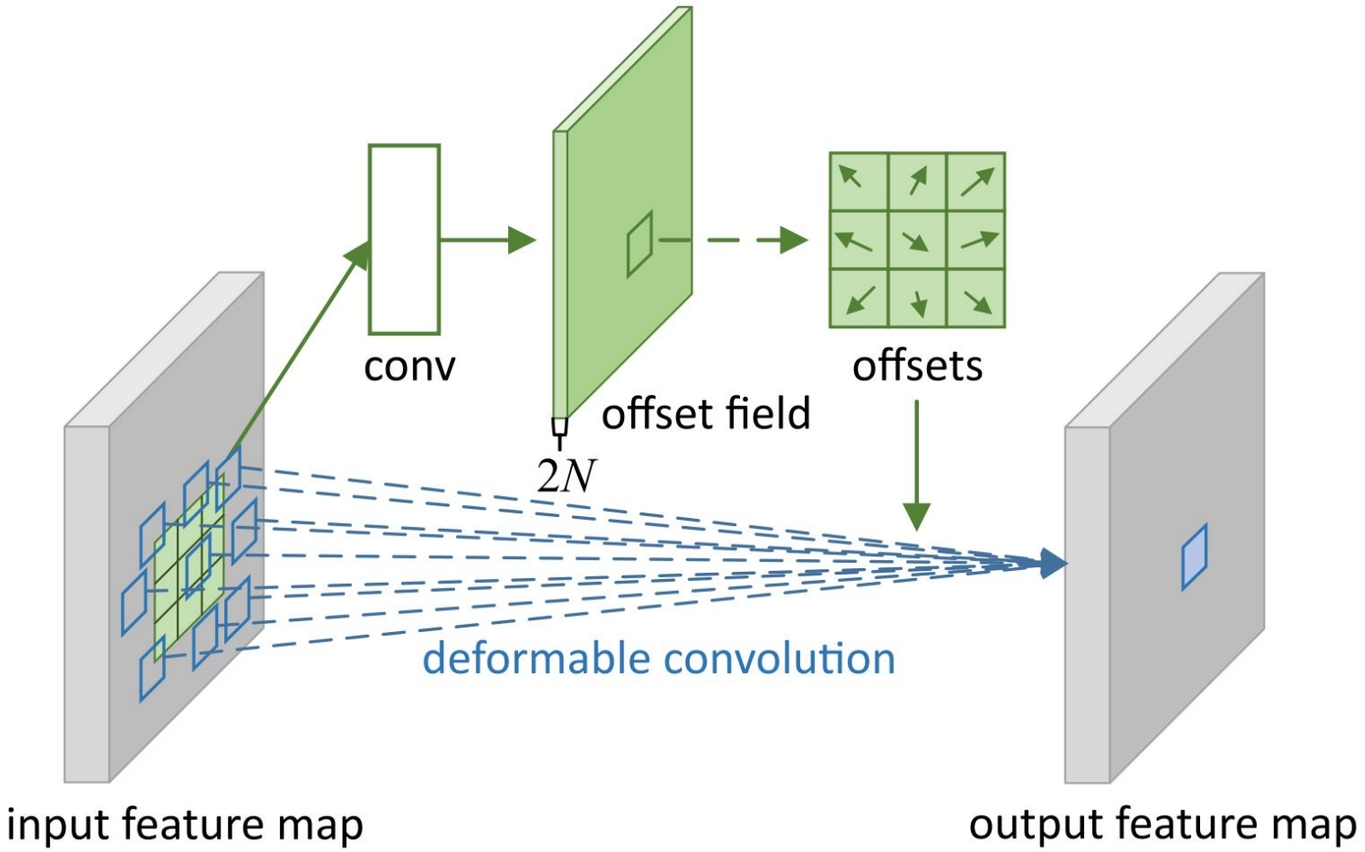
Cascade R-CNN



https://arxiv.org/abs/1712.00726

Deformable Convolutions

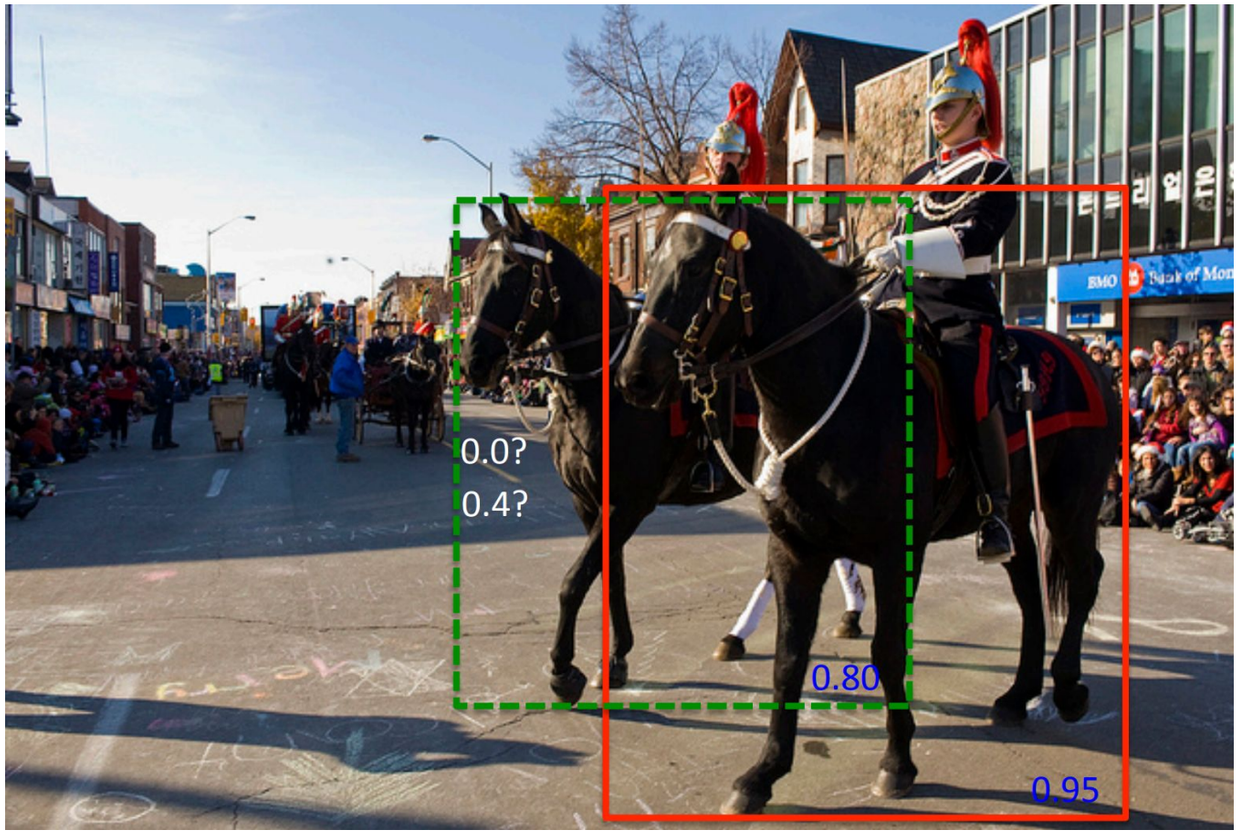
Dynamic and learnable receptive field



https://arxiv.org/abs/1703.06211

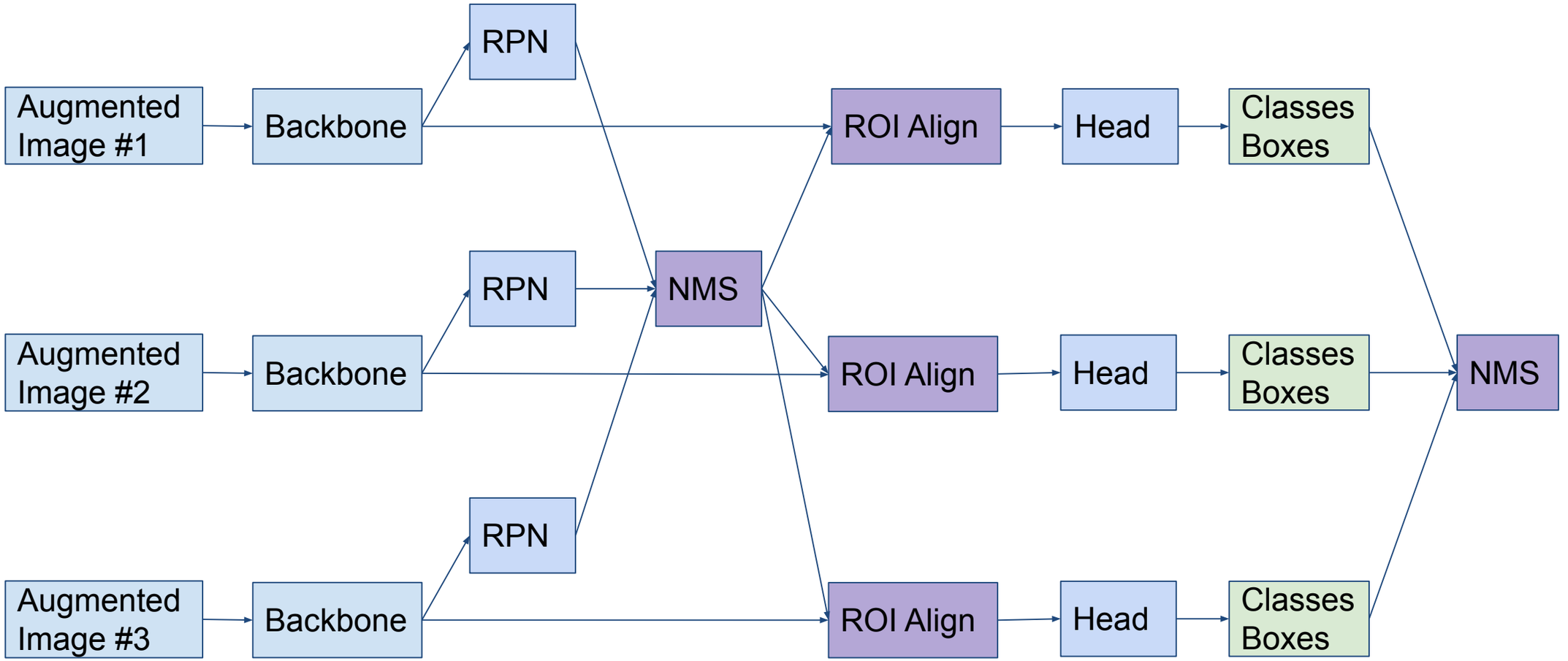
Soft-NMS

Decay detection scores of contiguous objects instead of setting them to 0



https://arxiv.org/abs/1704.04503

Test Time Augmentation



Final model

- Cascade R-CNN with ResNet50 backbone
- Pre-train from MS COCO
- Deformable convolutions in first layer
- MultiScale training
- Post-processing: Soft-NMS
- Ensemble: TTA (original + horizontal flip)

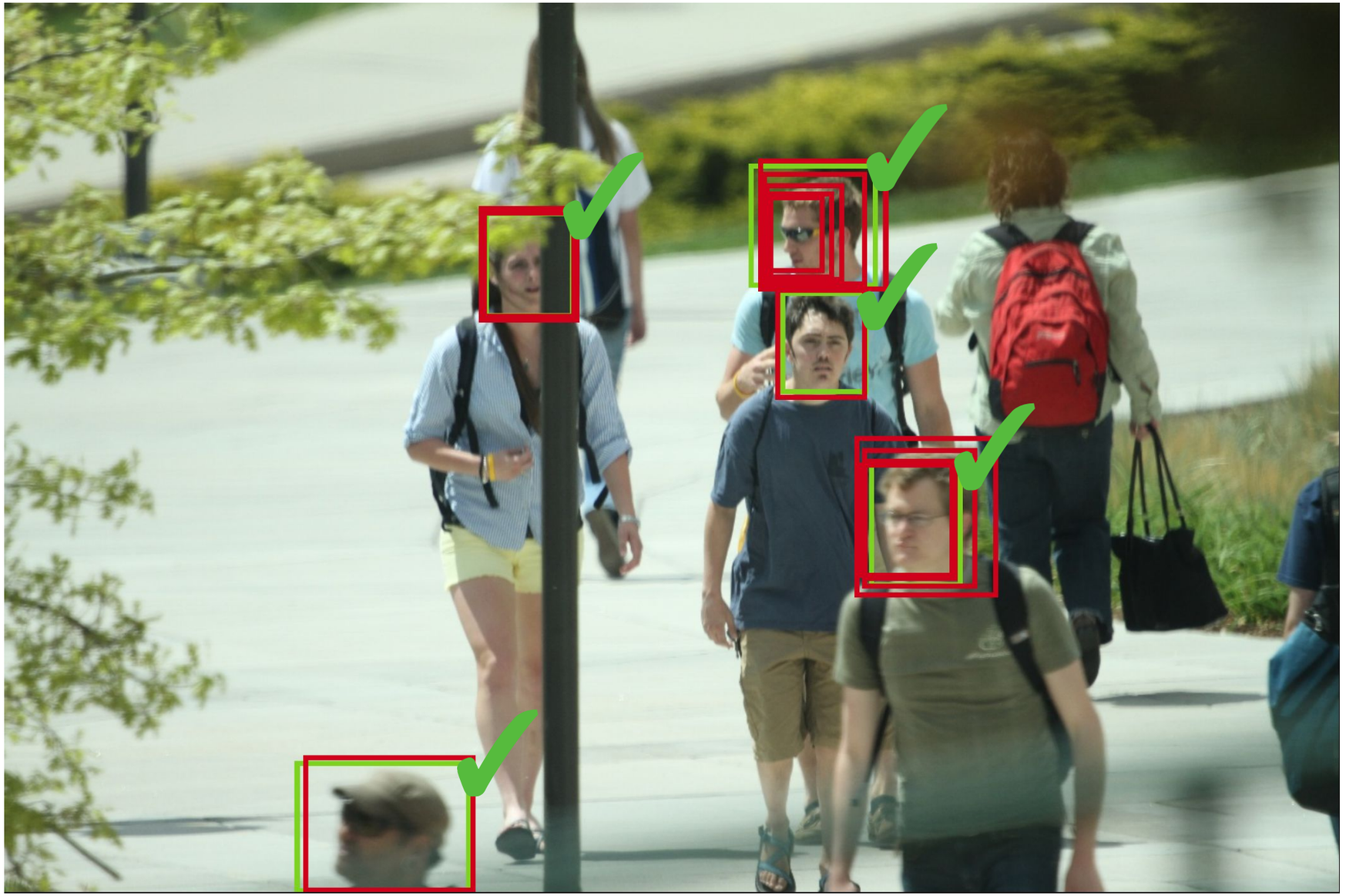
Training details

- SGD
  - lr=0.02
  - momentum=0.9
  - weight_decay=0.0001
- batch_size: 8 = 2 images per gpu x 4 gpus Tesla V100
- 15 epochs
- Learning Rate Scheduler: MultiStep, Milestones= [7, 11]
- Training time: 12 hours
- Public LeaderBoard Score: **49.126 ⇒ 53.505**

Results

: prediction

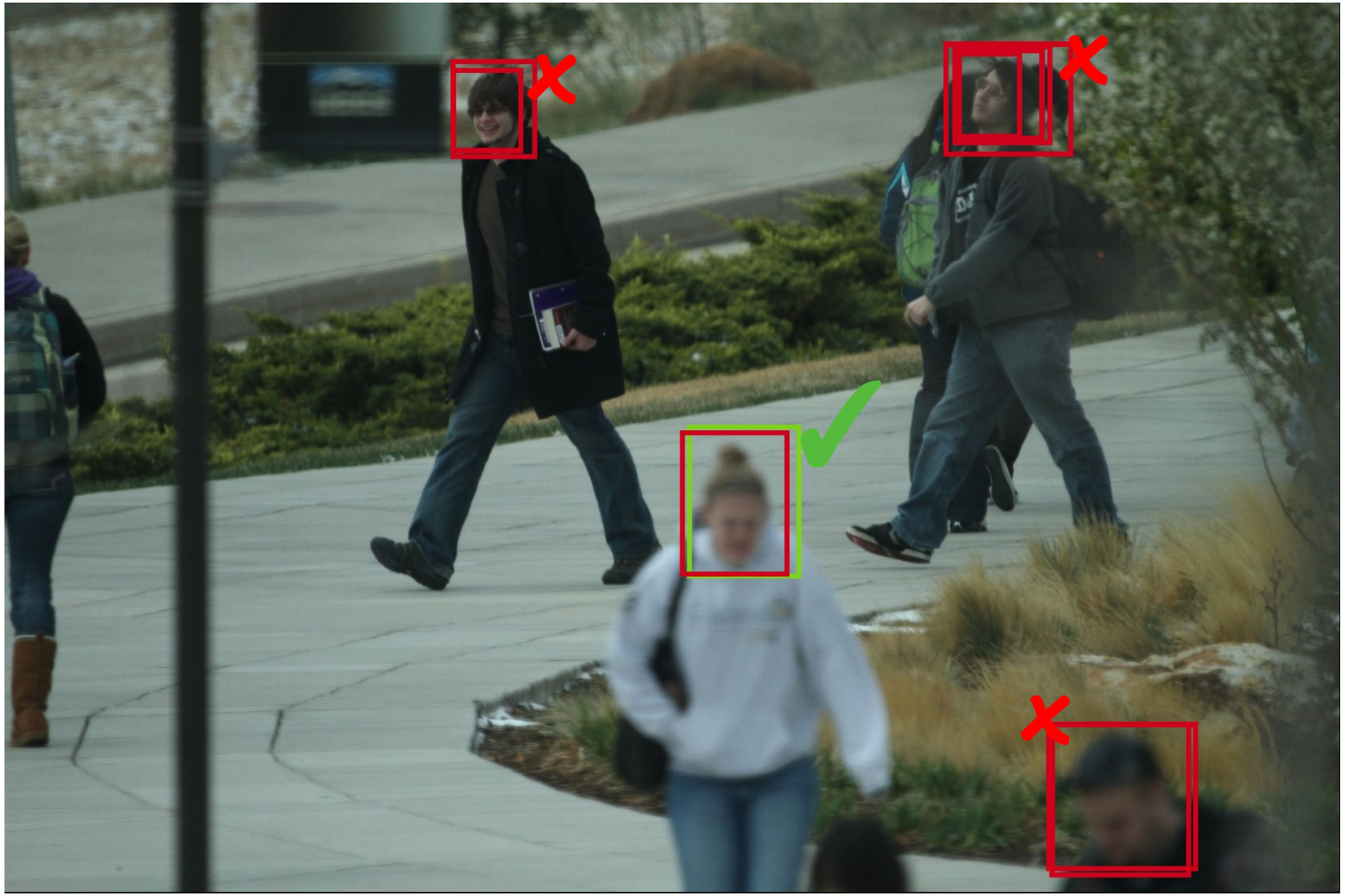
: ground truth



Results

: prediction

: ground truth

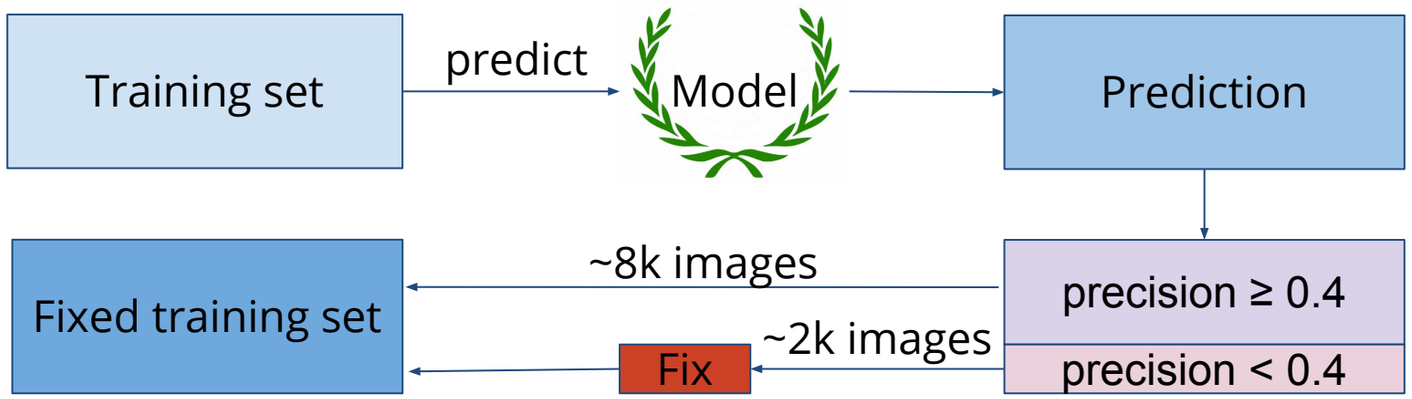


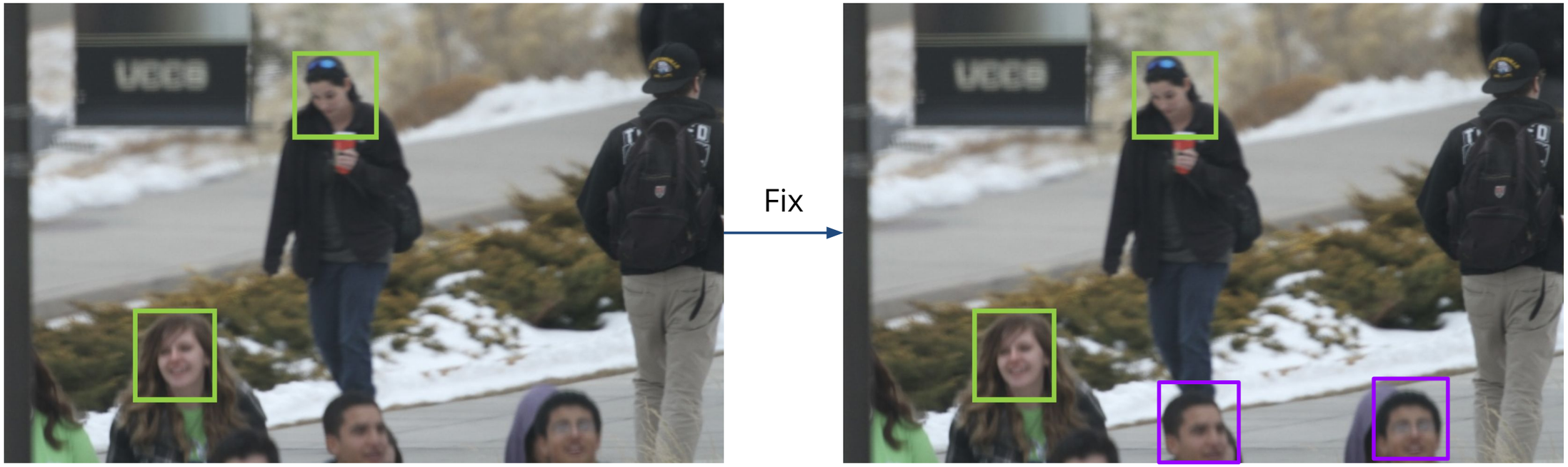
Results

: prediction

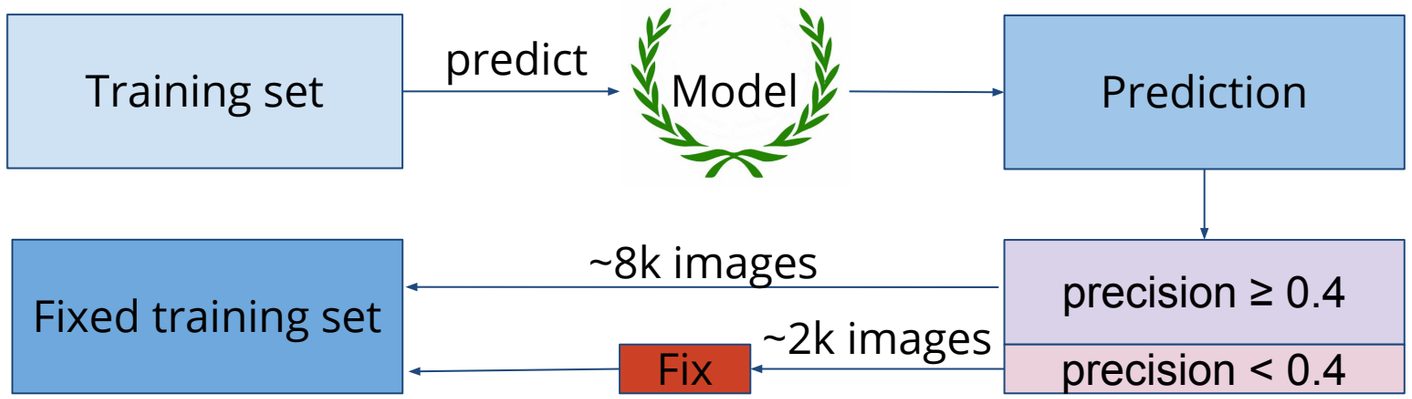
: ground truth

Annotation fix

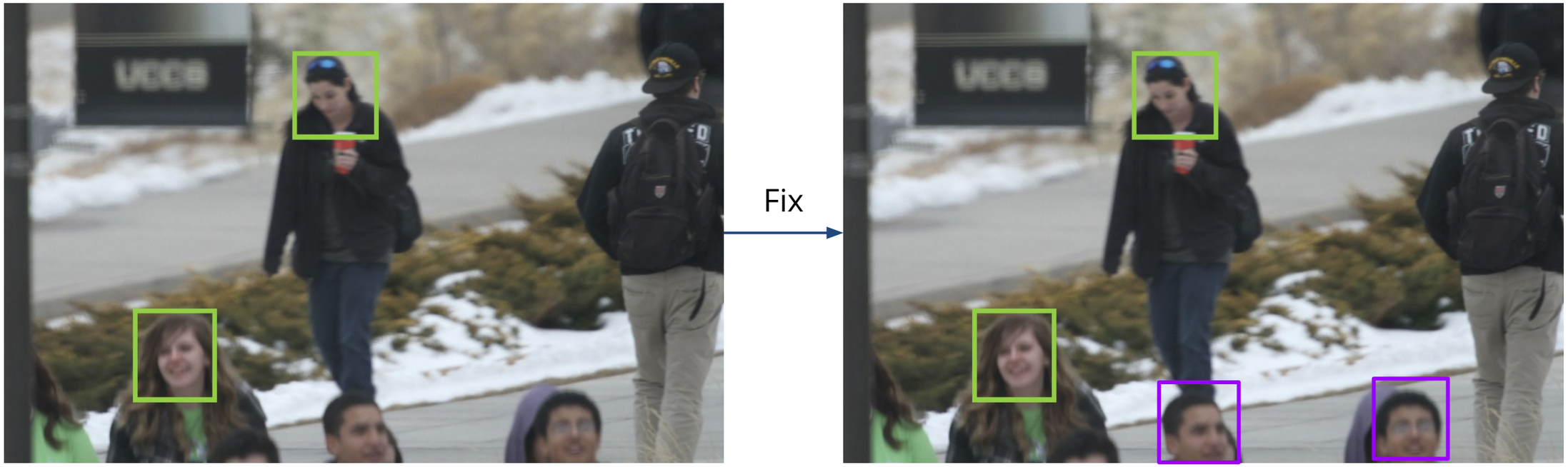




Annotation fix

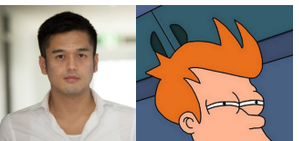


Public Leaderboard Score:
53.505 ⇒ 54.649



Leaderboard: Facial Detection Marathon

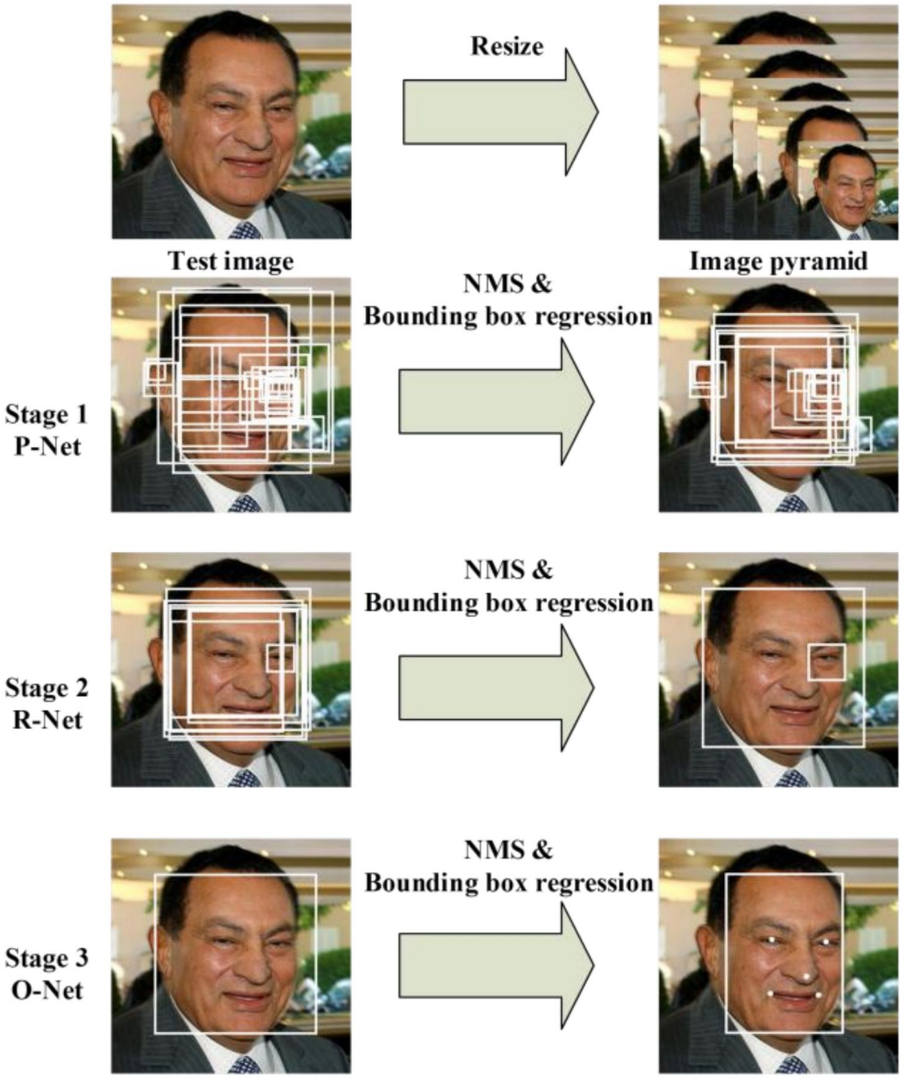
Rank		Handle	Score	
Final	Provisional		Final	Provisional
1	1	AlexZarichkovyi	54.67737	54.91
2	2	amirassov	54.43033	54.64966
3	4	ZFTurbo	53.76424	53.4076
4	5	n01z3	53.69196	53.33211
5	3	MaksimovKA	53.30687	53.72221



Face Re-identification

MTCNN

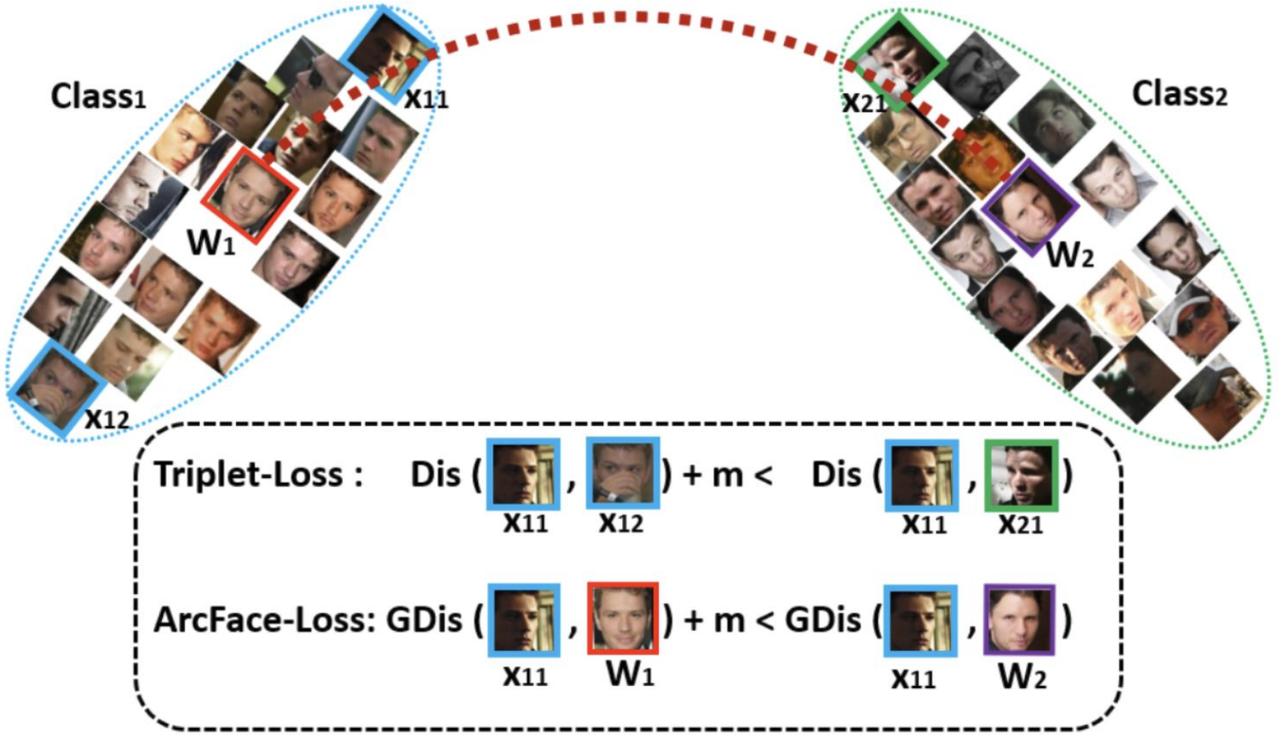
Cascaded structure with three stages of carefully designed deep convolutional networks that predict face and landmark location.



https://arxiv.org/abs/1604.02878

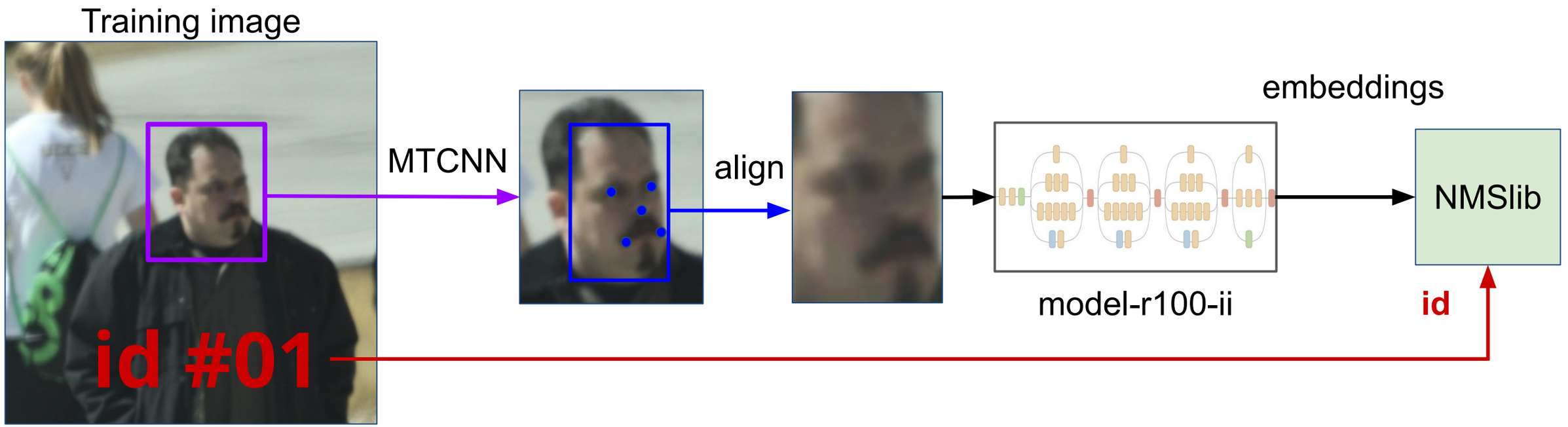
ArcFace

Directly maximise decision boundary in angular (arc) space based on the L2 normalised weights and features



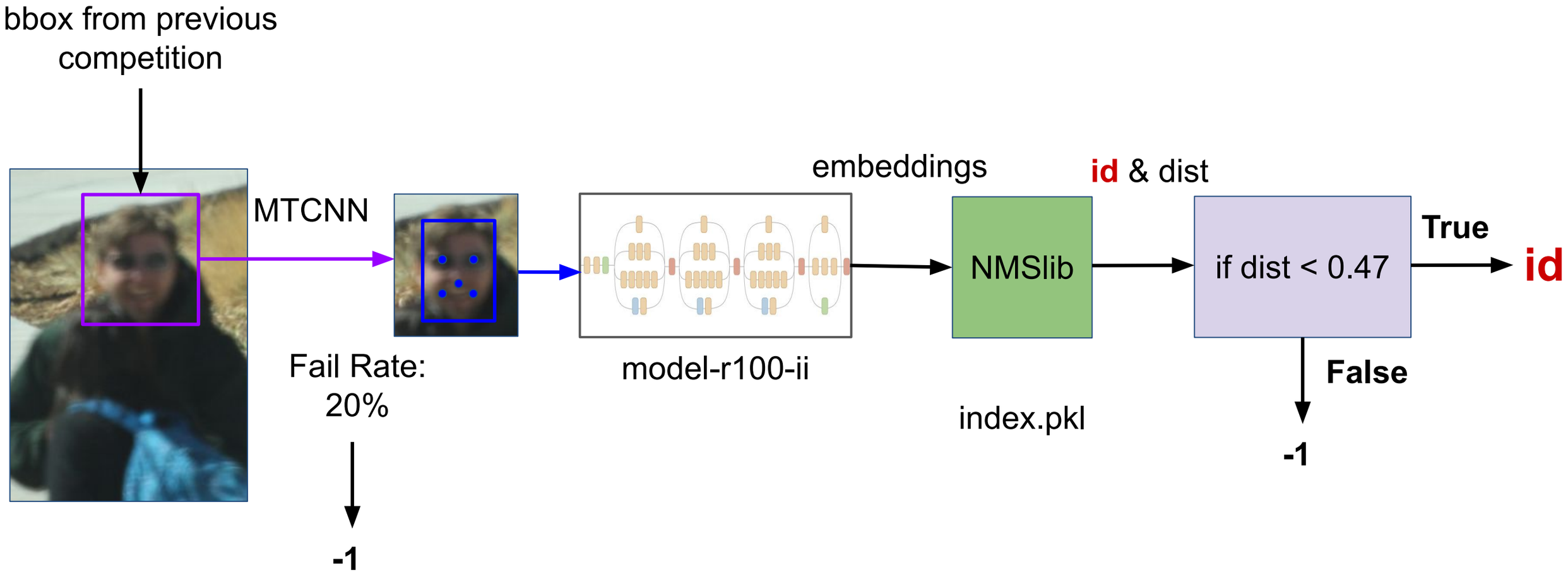
https://arxiv.org/abs/1801.07698
https://github.com/deepinsight/insightface

Re-identification: training pipeline



https://github.com/nmslib/nmslib

Re-identification: test pipeline



Leaderboard: Facial Re-Identification Marathon

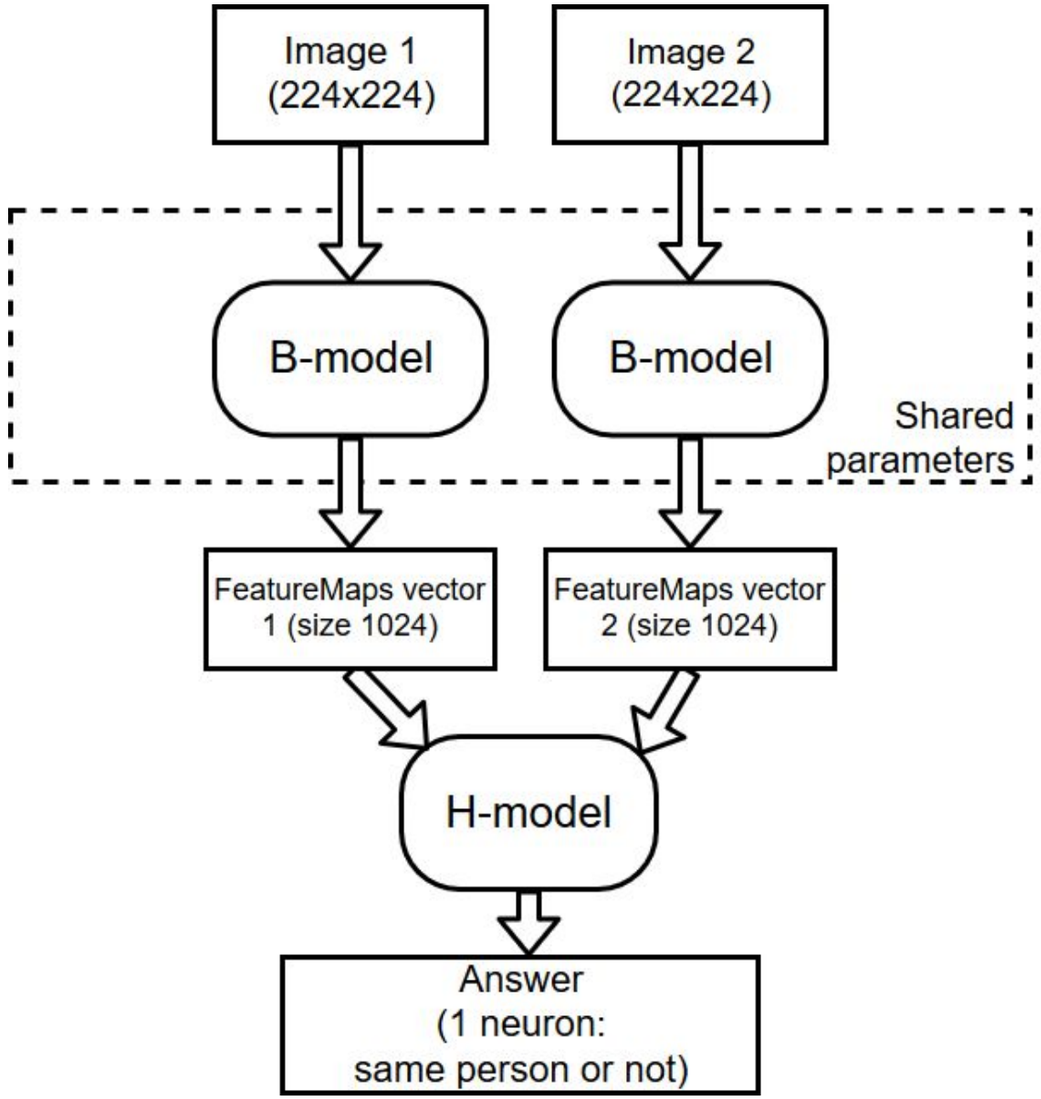
Rank		Handle	Score	
Final	Provisional		Final	Provisional
1	1	ZFTurbo	50.33887	50.97573
2	2	amirassov	48.58725	50.40111
3	3	n01z3	45.18998	46.76524
4	16	vadik	44.28962	0
5	4	MaksimovKA	43.04459	45.42458

Detection: 1st place solution (AlexZarichkovyi)

- Cascade R-CNN + ResNet-50 (with deform convs) + FPN
- Pre-train from MS COCO instance segmentation task
- Multi-scale training: from 1200 x 800 to 1800 x 1200
- SSD photometric augmentations: random contrast, saturation, hue
- Weight decay tuning
- TTA (multi-scale testing [1800x1200 and 1500x1000] and horizontal flipping)
- Software: mmdetection, PyTorch 1.0

Re-identification: 1st place solution (ZFTurbo)

- Detection: Retinanet ensemble with different backbones
- Image size: 800 x 1200
- NMS, TTA (horizontal flipping)
- Re-identification: Siamese Network with DenseNet121
- Calculate several metrics (e.g. sum, product, absolute distance, squared distance) between the two feature vectors in H-model.



Thank you for your attention!

https://github.com/amirassov/topcoder-facial-marathon