Advanced Applications of Computer Vision

张海鹏

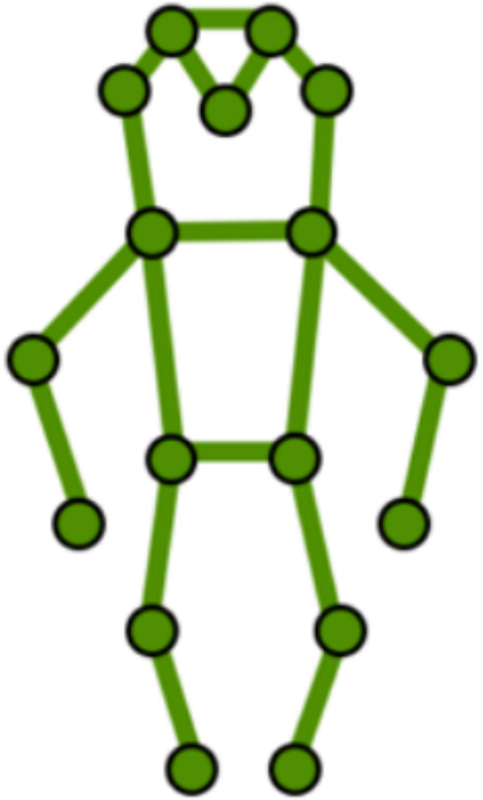
2018-05-28

Outline

- Human Keypoint Detection
- Re identification(Person, Vehicles, etc.)
- Face

Keypoint Detection

人体骨骼关键点检测, 2017, AI Challenger



Keypoints, 2017, COCO

Object Keypoint Similarity

$$OKS = \frac{\sum_i e^{-\frac{d_i^2}{2s^2k_i^2}} \delta(v_i > 0)}{\sum_i \delta(v_i > 0)}$$

- d_i - the Euclidean distances between each ground truth and detected key point
- v_i - the visibility flags of the ground truth
- $s * k_i$ - the standard deviation of the gaussian
- perfect predictions will have $OKS = 1$
- predictions with key points wrong by more than a few standard deviations will have $OKS \sim 0$

http://cocodataset.org/#keypoints-eval

Problem Modeling

- Single person pose estimation
- Multi person pose estimation
  - Top-down
  - Bottom-up

	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR	AR^{50}	AR^{75}	AR^{M}	AR^{L}	date
Megvii (Face++)	0.721	0.905	0.789	0.679	0.781	0.787	0.947	0.848	0.743	0.847	2017-10-29
oks	0.714	0.894	0.781	0.659	0.791	0.772	0.936	0.834	0.718	0.845	2017-10-29
bangbangren	0.706	0.880	0.765	0.656	0.792	0.774	0.936	0.830	0.718	0.850	2017-10-29
G-RMI	0.691	0.859	0.752	0.660	0.745	0.751	0.907	0.807	0.697	0.824	2017-10-29
FAIR Mask R-CNN	0.689	0.892	0.752	0.637	0.768	0.754	0.932	0.812	0.702	0.826	2017-10-29
SJTU	0.680	0.867	0.747	0.633	0.750	0.735	0.908	0.795	0.686	0.804	2017-10-29

Solution-SenseTime

- Top-down method
    - Person detection(Re-implement FPN+Mask-RCNN, ResNet50)
    - Pose estimation(Stacked Hourglass(v1) 8 stacks)
- Inference
    - Box proposal rescoring(sort box candidates by the product of box score and key point score)
    - OKS-NMS(+IoU NMS)

BlockQNN(Q learning)
MPII(Transfer learning)
5 days, 32 GPUs

method	AP (validation)
Box score	70.3
Keypoint score	56.1
Rescoring	71.5

Solution-SenseTime

- Data selection(KS between box center and keypoint center and keypoint number of an instance)

$$KS(C_{box}, C_{kps}) = e^{\frac{-\|C_{box} - C_{kps}\|^2}{2 * area_{box}}}$$

- External data

1. Train a hourglass 8 stacks with COCO only data

2. Use the model above to select hard examples in AICKD(AI Challenger Keypoint Dataset)

3. Joint train with COCO data and hard examples of AICKD

4. only prop the loss of common annotations with COCO for AICKD

Solution-SenseTime

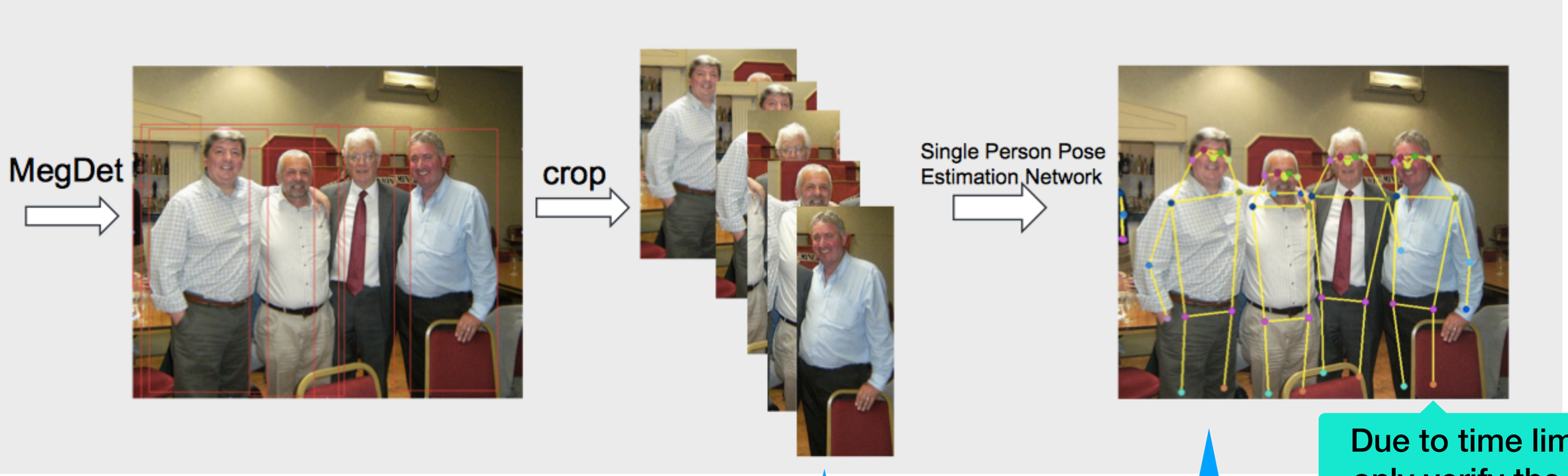
- For top-down methods, single person pose estimation module is **much more important** than detection module

- A direct simple CNN regression model can solve complicated pose estimation problems in COCO dataset, including heavily occlusion, large variance and crowding cases

- **Hourglass** shows great performance for single pose estimation task, but it is not the only choice

Method	AP (validation set)
Hourglass 8 stacks naïve	70.3
++ data selection	70.8
++ proposal rescoring	71.5
++ OKS-NMS	71.7
++ external data	73.0
++ ground truth box	75.5

Solution-Face++

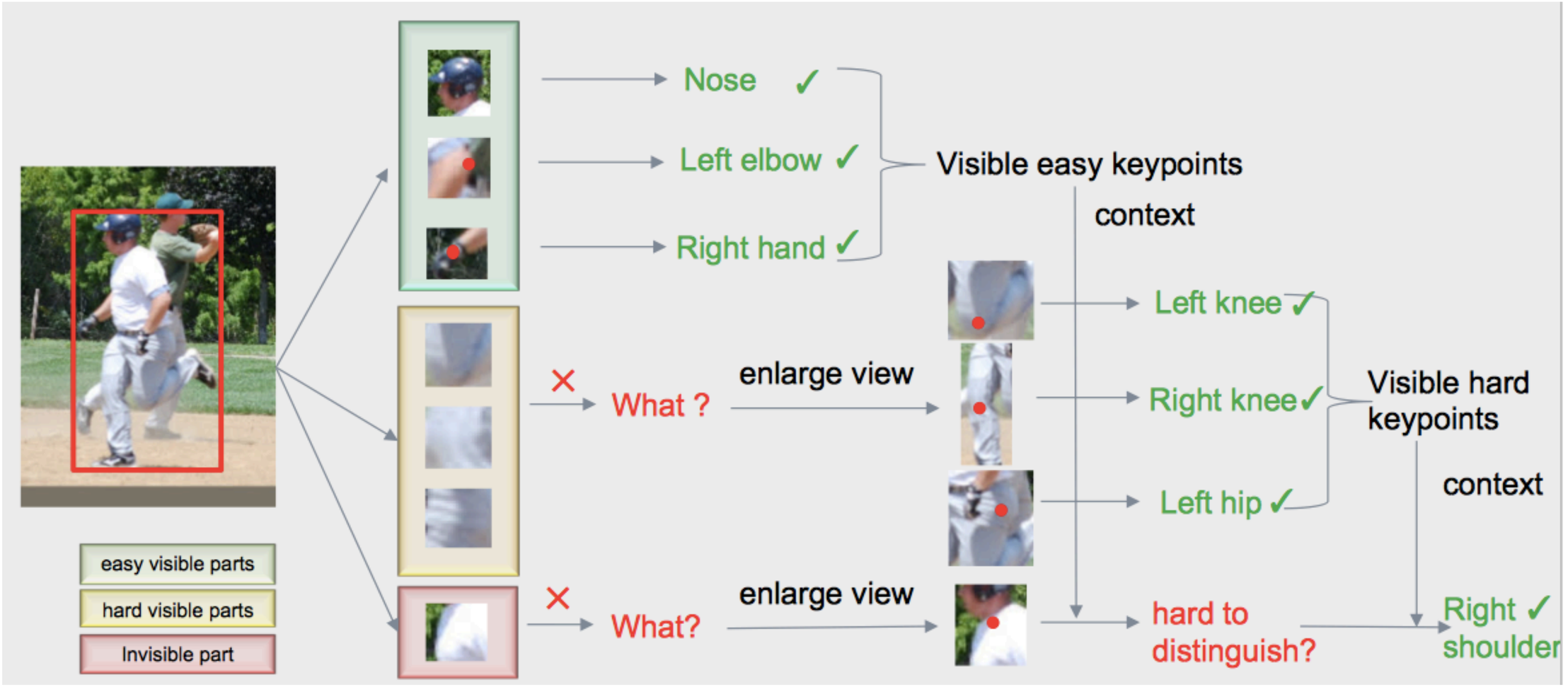


MegDet

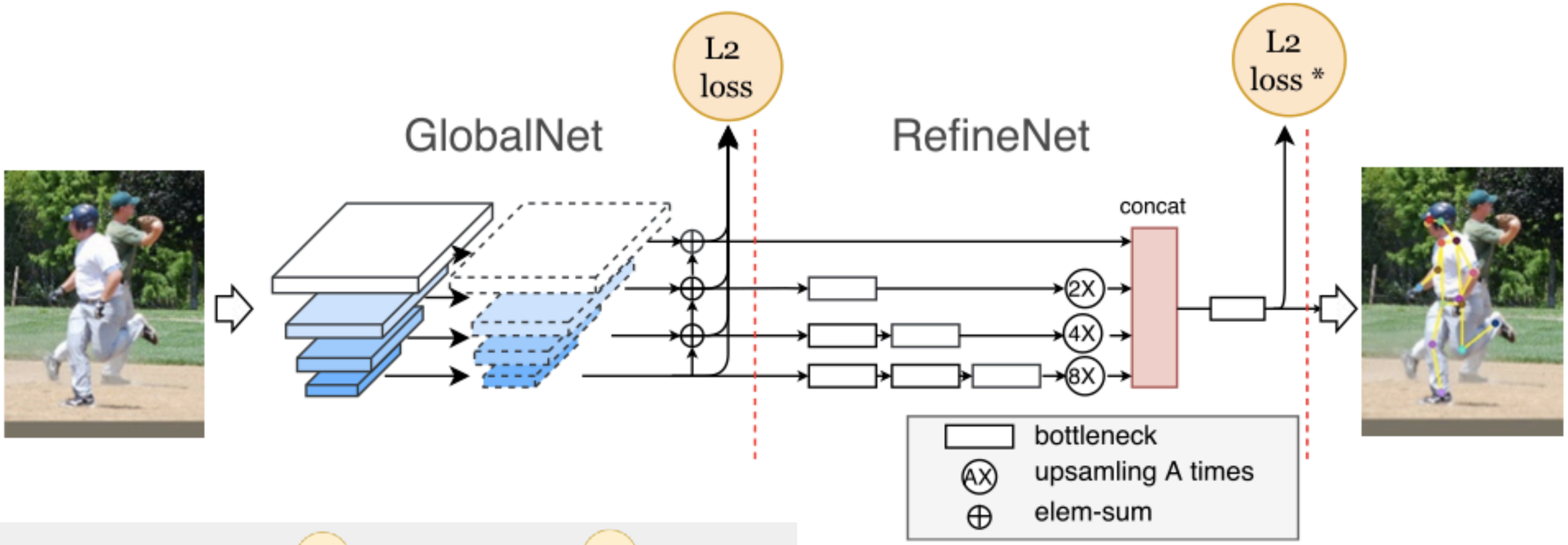
Due to time limited, we only verify the result of hourglass 2 stacks.

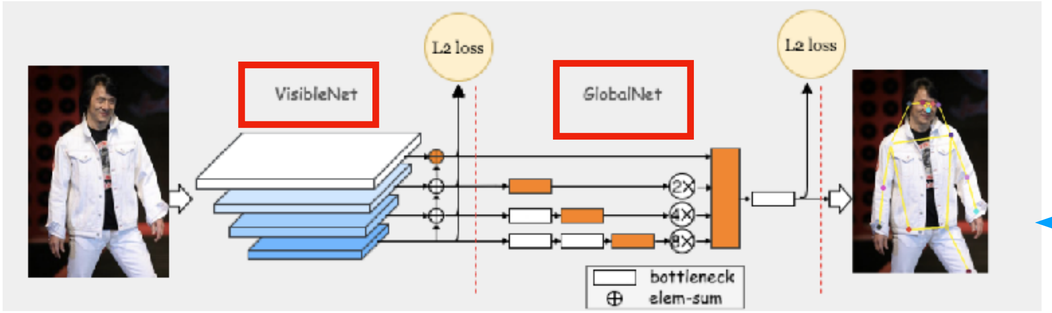
Two stages are enough for key point localization for better trade off(Guess: Hourglass stages harm the spatial resolution).

Solution-Face++



Solution-Face++





COCO 2017, Keynote, Face++

《Cascaded_Pyramid_Network_for_Multi_Person_Pose_Estimation》, CVPR2018

Solution-Face++

- Data augmentation(+0.4AP)
  - Crop augmentation
  - Random scales
  - Rotation
- Large Batch(+0.4~0.7AP)
- Segmentation supervision(+0.2~0.6AP)
  - Enhance the network's ability to distinguish the detected person from crowded scene
- Ensemble(+1.1~1.5AP)
  - Heatmap merge

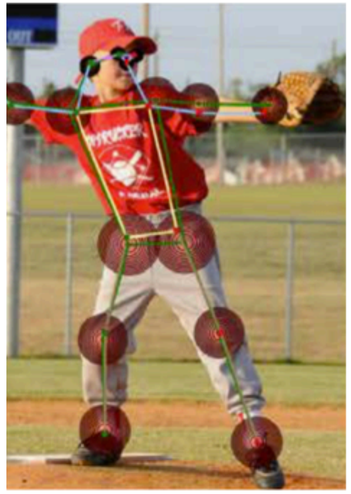
Data augmentation is the key to enhance robustness of network, especially in CNN

Large batch technique is not only applicable in object detection, but also in key point

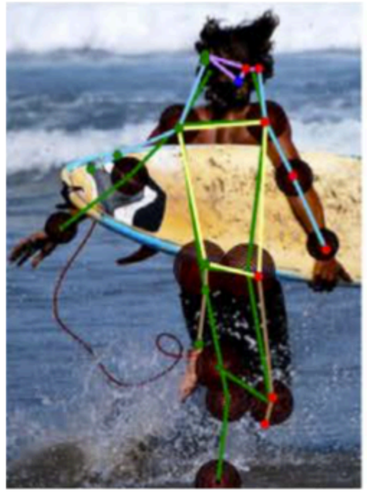
Segmentation supervision is also an universe skill in training CNN

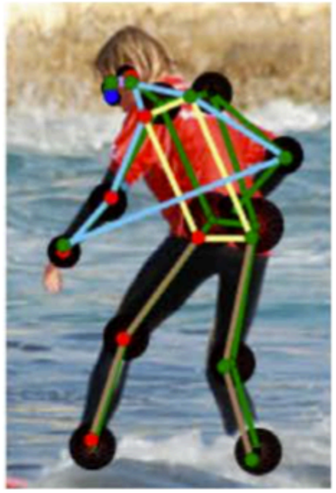
Taxonomy of Errors

JITTER

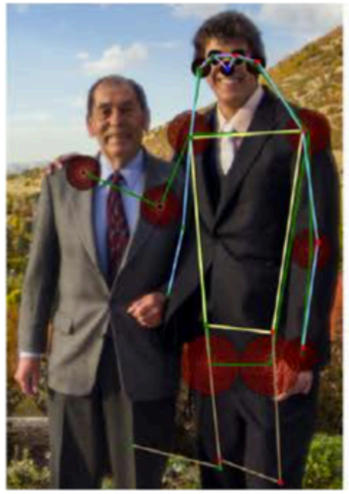


INVERSION



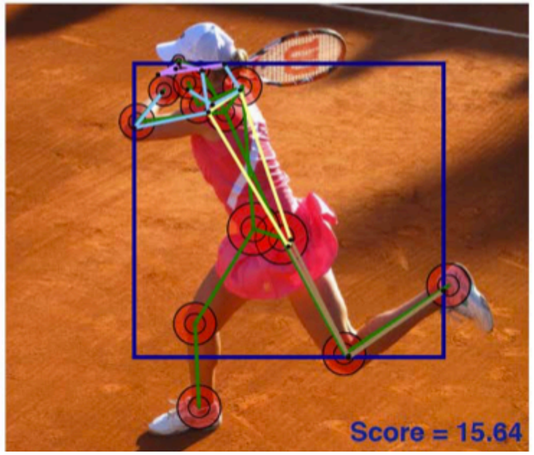


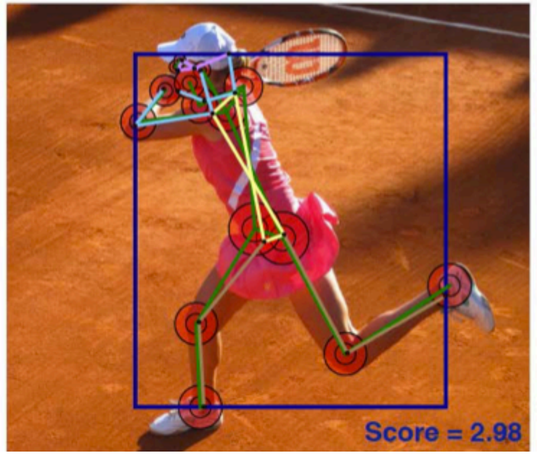
SWAP



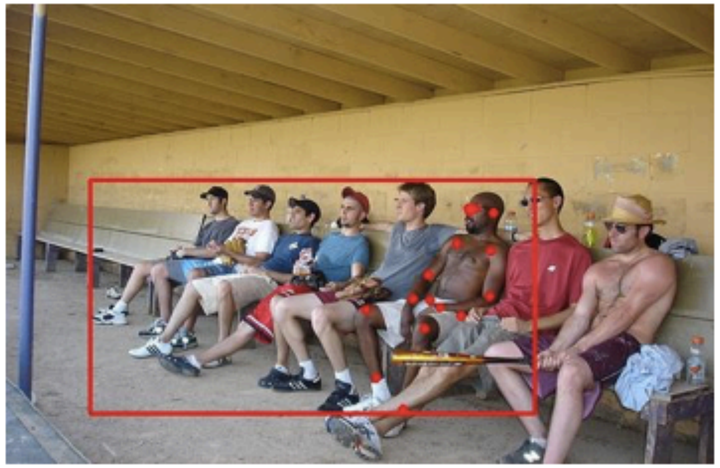
MISS

SCORING





False annotations in COCO

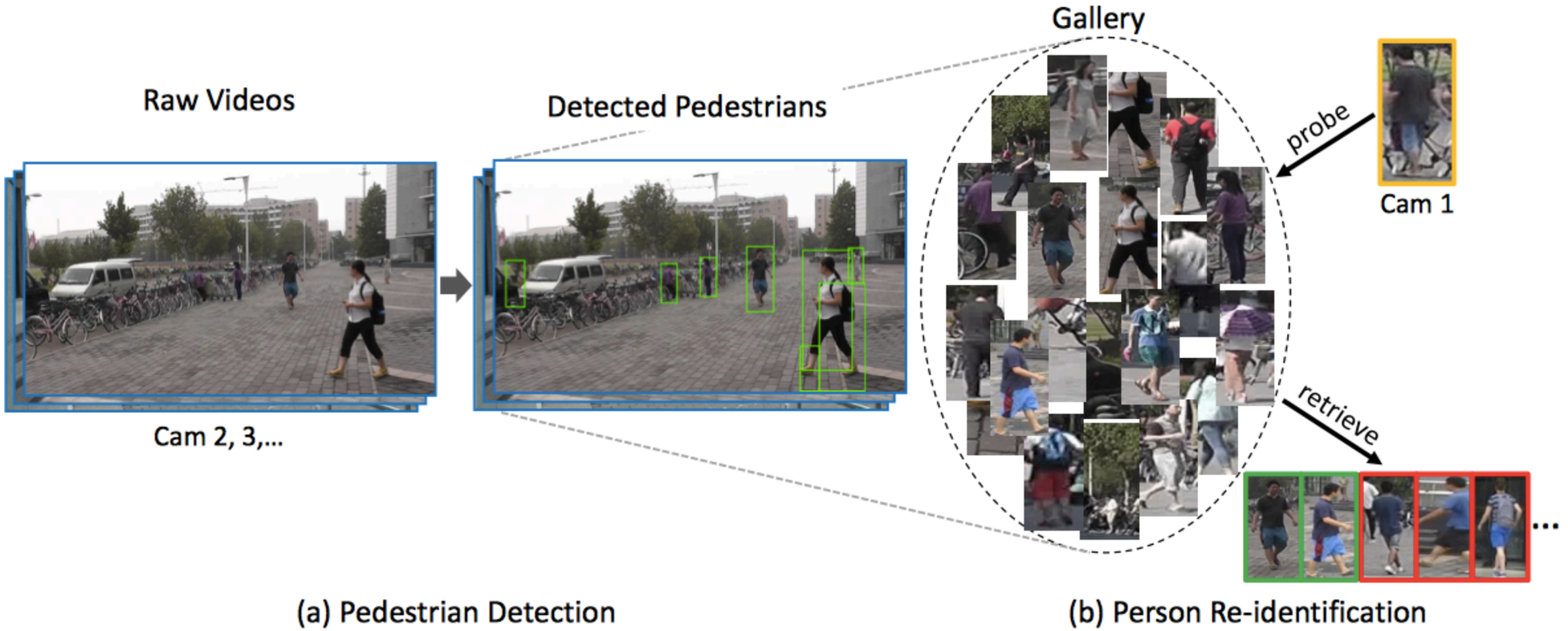


Badcases of COCO

Keypoints of Keypoints Detection

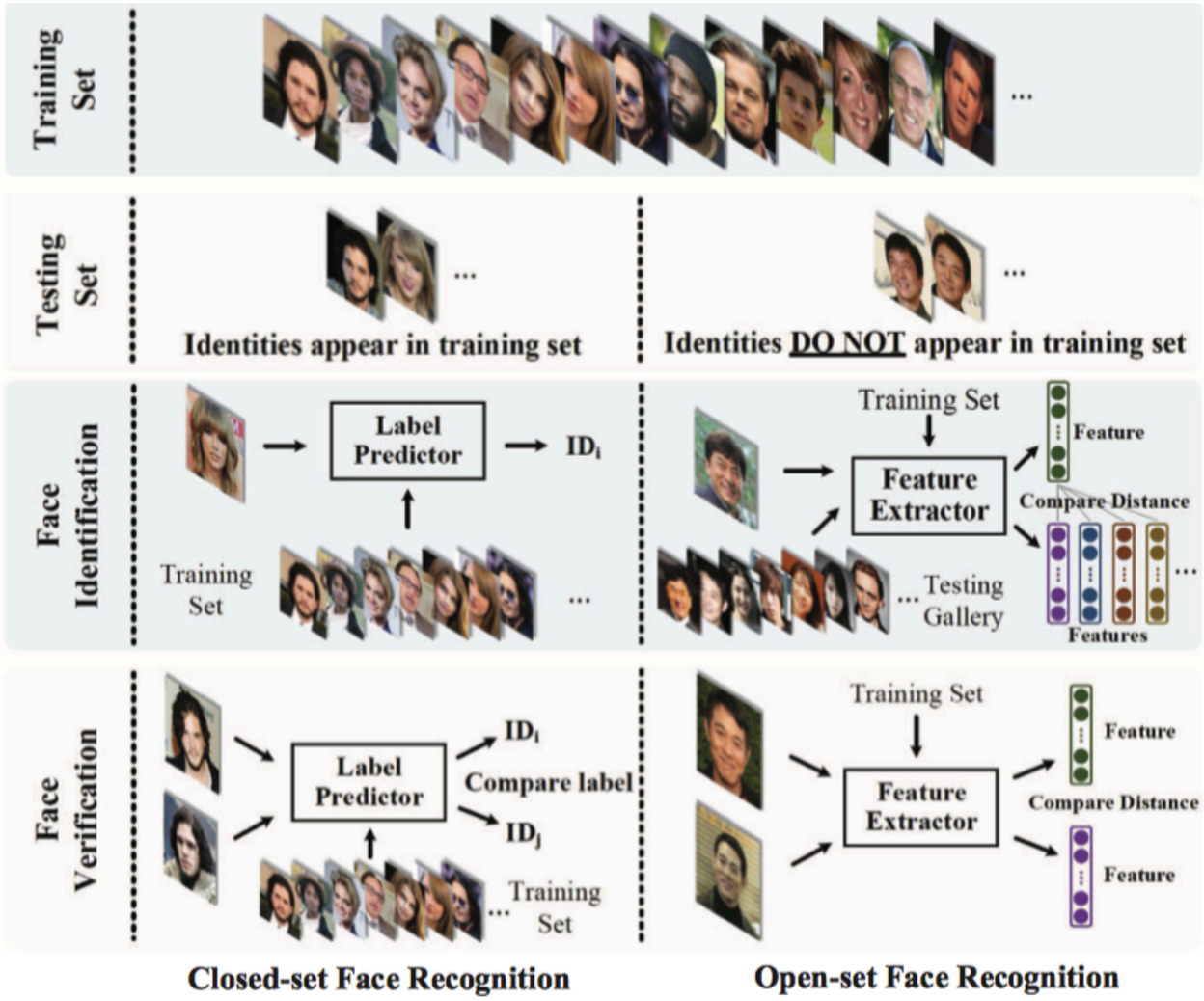
- Instance localization(predominant) affects also keypoints
- Instance scale is an important factor
- Performance can vary greatly across different keypoints sets
- Single model performance is on par with ensembles
- Top-down(detection+single pose estimation)

Re-identification



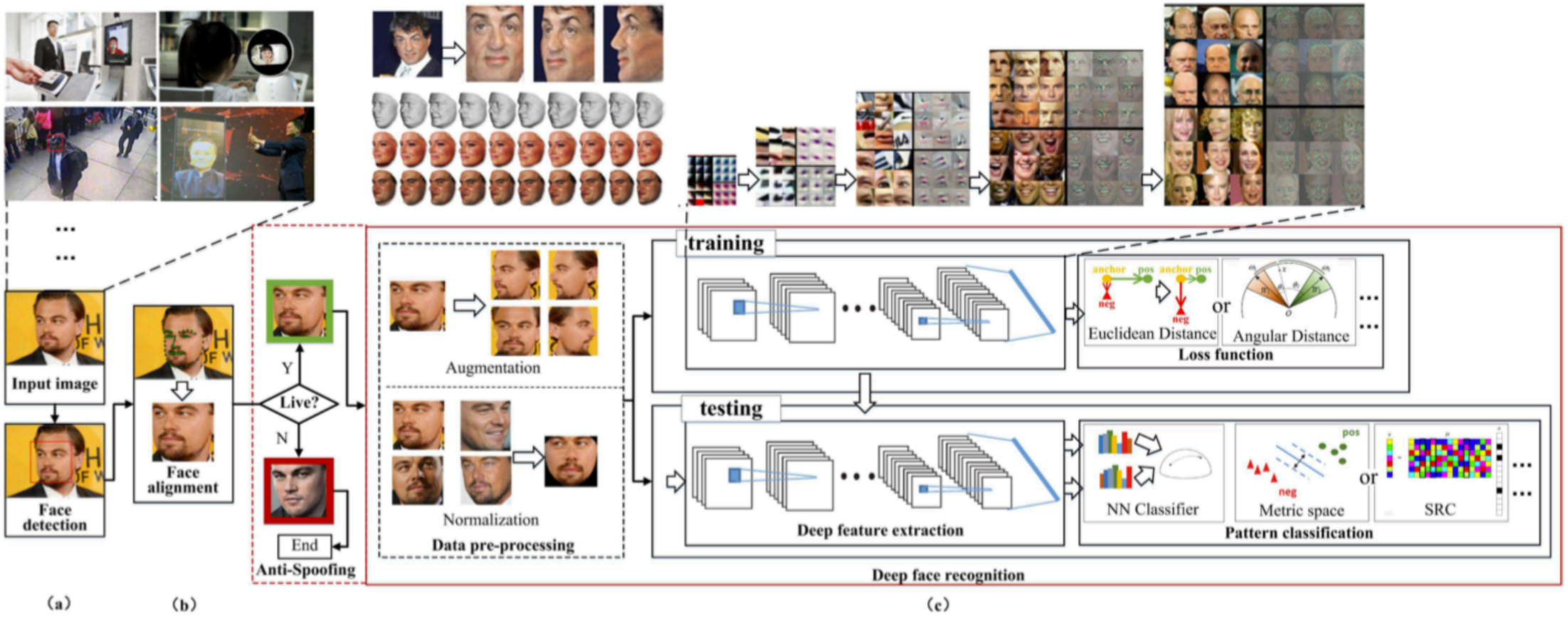
An end-to-end penson re-ID system that includes pedestrian detection and re-identification

Comparison of closed-set and open-set of FR



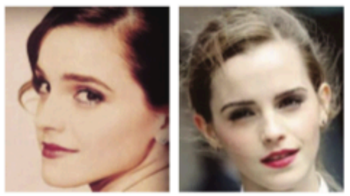
《Deep Face Recognition: A Survey》, 2018

Deep FR system with face detector and alignment



《Deep Face Recognition: A Survey》, 2018

The different scenes of FR



(a) cross-pose

(b) video

(c) cross-age

(d) low-shot

(e) template-based

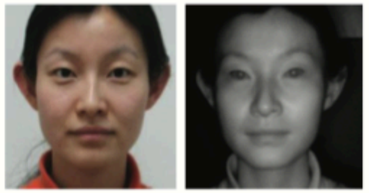
(f) anti-spoofing

(g) 3D

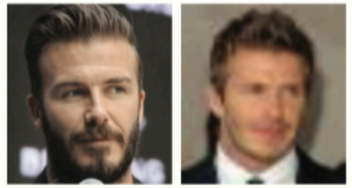
(h) make-up

(i) mobile devices

Heterogeneous face recognition



(j) NIV-VIS



(k) low resolution

(l) photo-sketch

《Deep Face Recognition: A Survey》, 2018