Team OKS Keypoints Detection

Joint Workshop of the COCO and Places Challenges at ICCV 2017

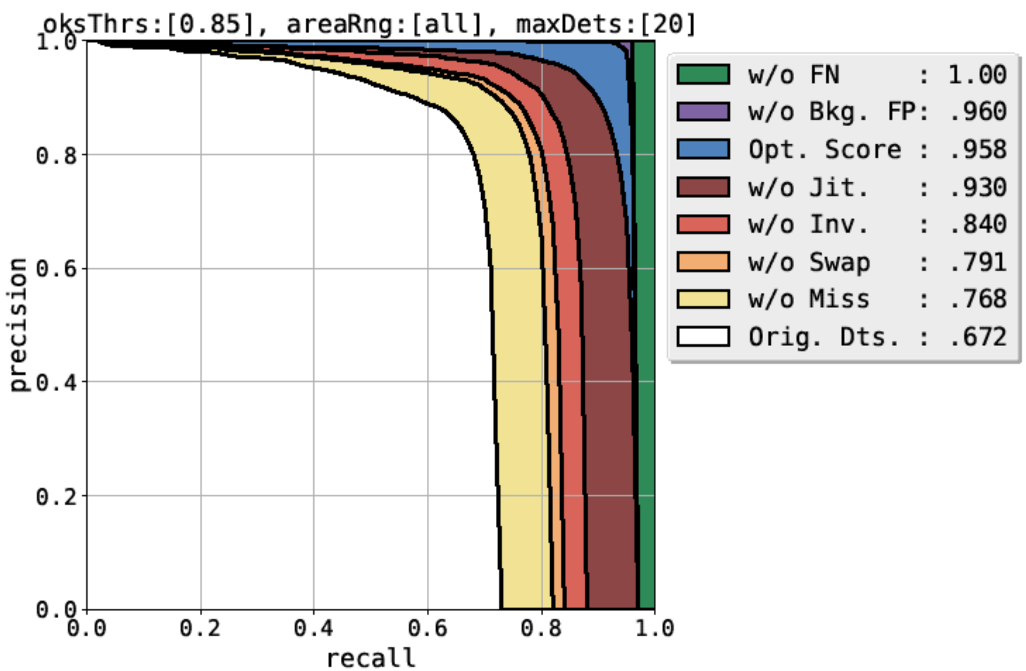
Speaker: Cheng Li

Yujie Wang[1*], Changbao Wang[1*], Quanquan Li[2*] ,
Biao Leng[1], Zhoujun Li[1], Junjie Yan[2]
[1]Beihang University [2]Sensetime Group Limited

(*Equal contribution. This work was done when Yujie Wang and Changbao Wang were interns at Sensetime Group Limited)

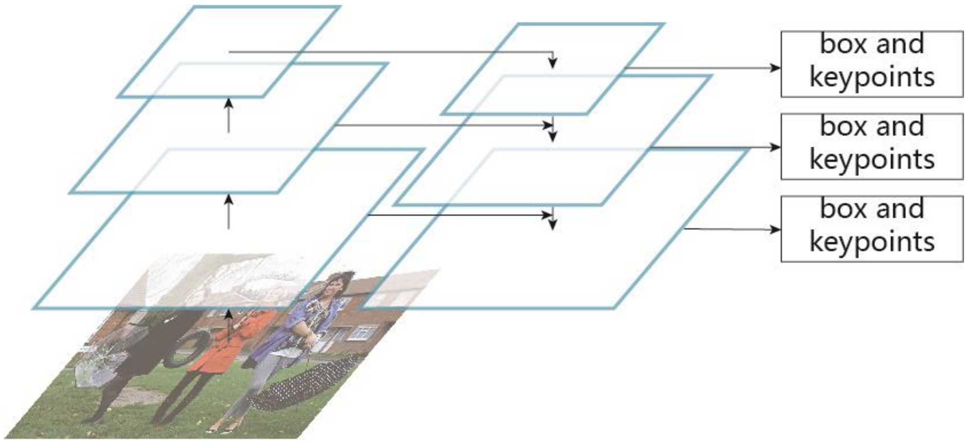
Outline

- Top-Down method
  - Person detection
  - Pose estimation
- Inference
  - Box Proposal Rescoring
  - OKS-NMS
- AP of our submission
  - 72.0 (test-dev)
  - 71.4 (test-challenge)



Person Detection

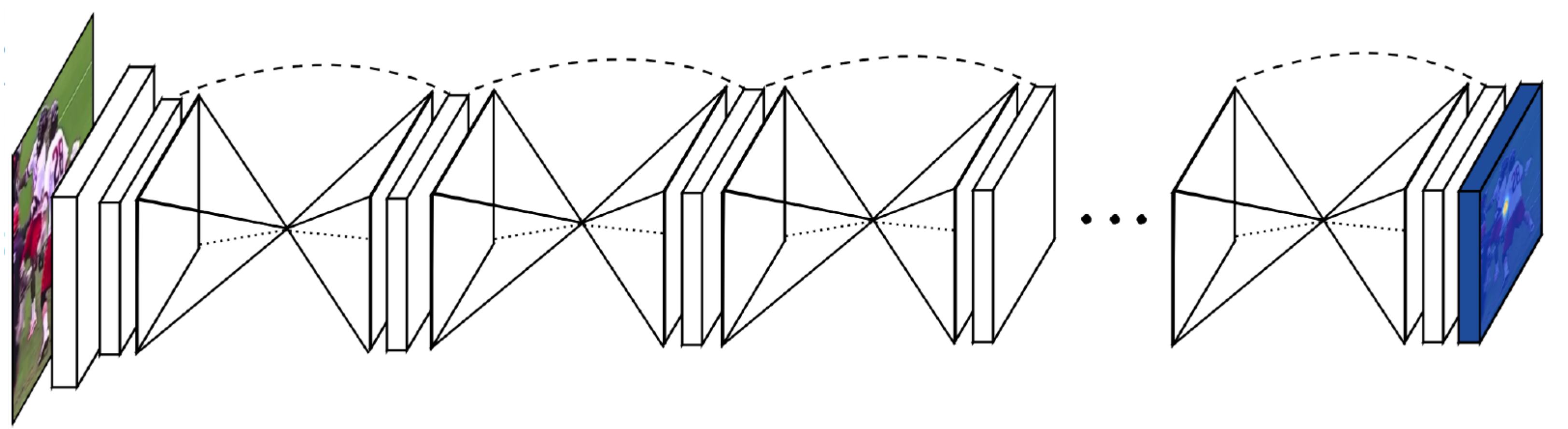
- Re-implement FPN + Mask-RCNN
  - Backbone: ResNet-50
  - Data: COCO only
  - Top 20 boxes
- Performance
  - COCO keypoint validation set
  - Box AP (person) 52.1
  - Box AR (person) 61.3



Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[J]. arXiv preprint arXiv:1612.03144, 2016.
He K, Gkioxari G, Dollár P, et al. Mask r-cnn[J]. arXiv preprint arXiv:1703.06870, 2017.

Pose Estimation Network

- Stacked Hourglass (v1) 8 stacks
  - Input size: 256x256
  - Supervision: Gaussian with std 1
  - Only backpropagate the loss of annotated keypoints



Newell A, Yang K, Deng J. Stacked hourglass networks for human pose estimation[C]//European Conference on Computer Vision. Springer International Publishing, 2016: 483-499.

Can we make the pose network better?

Explore new architecture

- Hourglass is good, but is it the best?

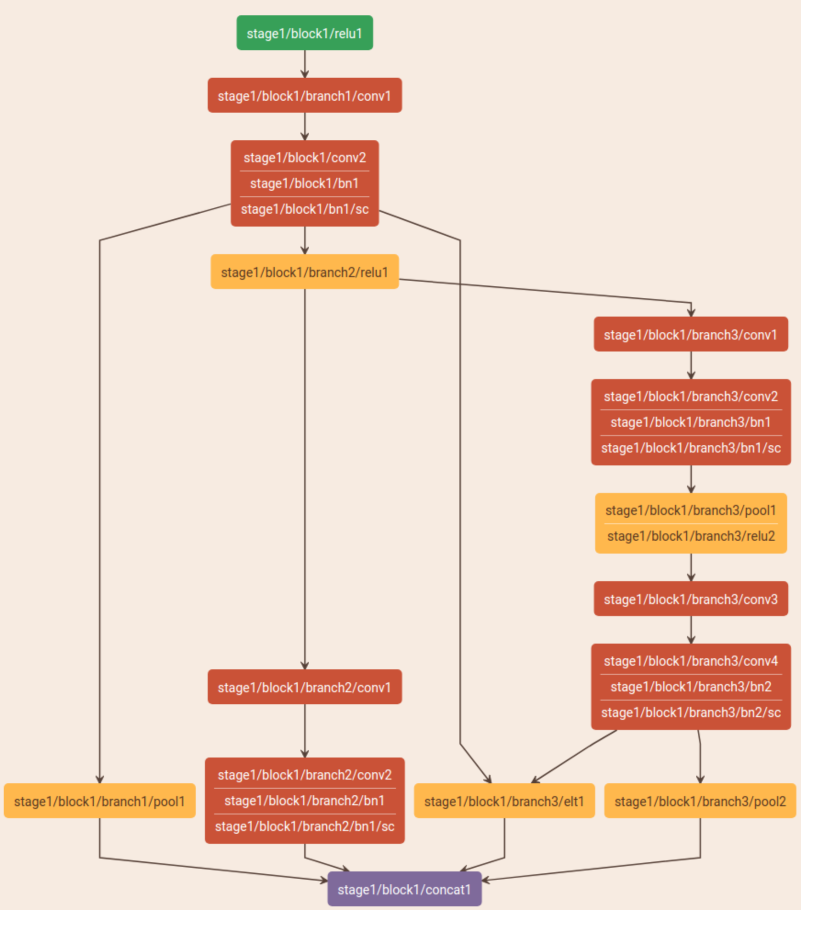
Method	AP (validation, ground truth box)
Hourglass 8 stacks	73.4
Inception ResNet V2*	69.4
ResNet-269*	69.7

- Can we design more effective and efficient architecture?

*These two networks have same stride with hourglass

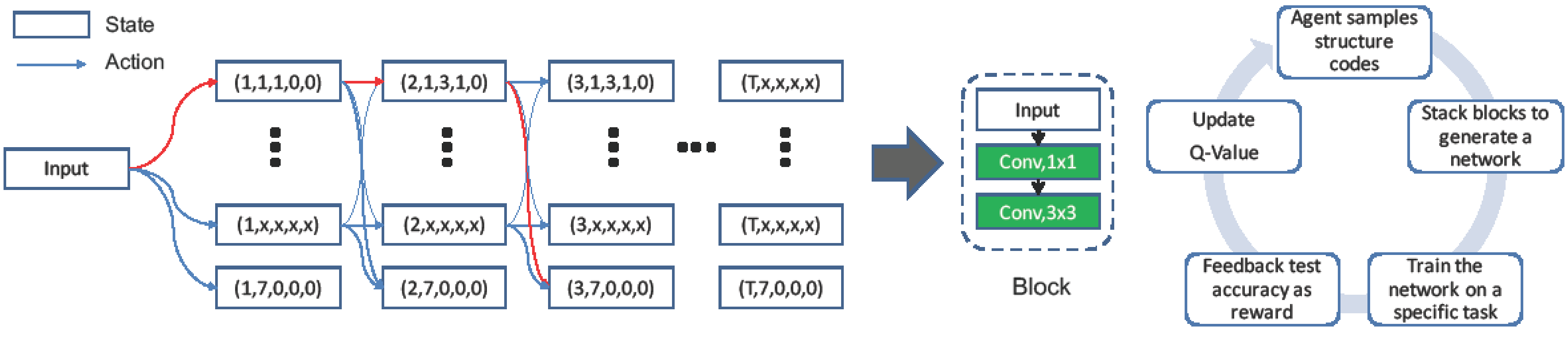
Explore new architecture

- We use automatic neural network design approach BlockQNN to generates optimal model on keypoints task
- We search the best model on MPII dataset and transfer it to coco challenge.
- It costs 5 days to complete the searching process with only 32 GPUs.

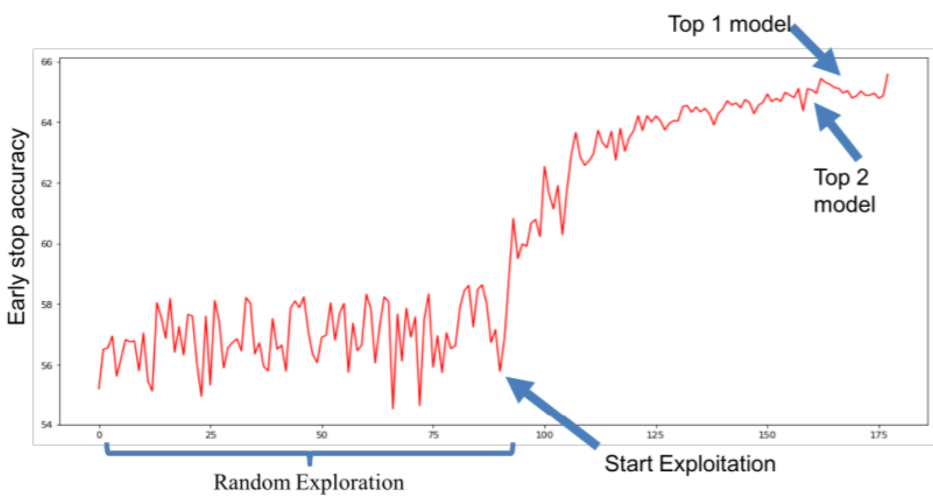


Zhong Z, Yan J, Liu C L. Practical Network Blocks Design with Q-Learning[J]. arXiv preprint arXiv:1708.05552, 2017.

Design Network Blocks by Q-learning



(a) (b)



Accuracy of Q-Learning in Exploration and Exploitation

Network	Error Rate on CIFAR10	Error Rate on CIFAR100
VGG	7.25%	
ResNet	6.61%	
DenseNet	3.74%	19.25%
Network Search from Google	3.65%	
Our method	3.60%	18.64%

Comparison with state-of-the-art methods on CIFAR-10 and CIFAR-100

Z. Zhong, J. Yan, and C. Liu “Practical Network Blocks Desgin with Q-Learning,” arXiv: 1808.5552, 2017

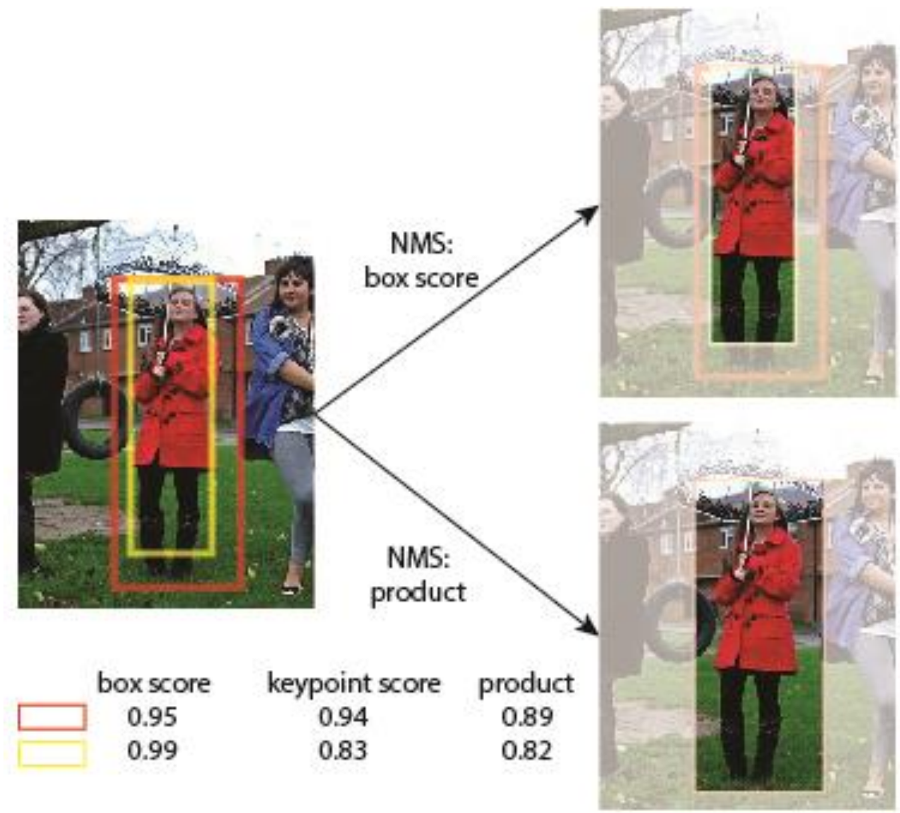
Explore new architecture

- Due to time limited, we only verify the result of hourglass 2 stacks and the generated network
- The generated network has less number of parameter
- We evaluate them on validation set with ground truth box

Method	AP (validation)	Parameter Number
Hourglass 2 stacks	70.1	19M
The generated network	70.5	17M

Box Candidates Rescoring

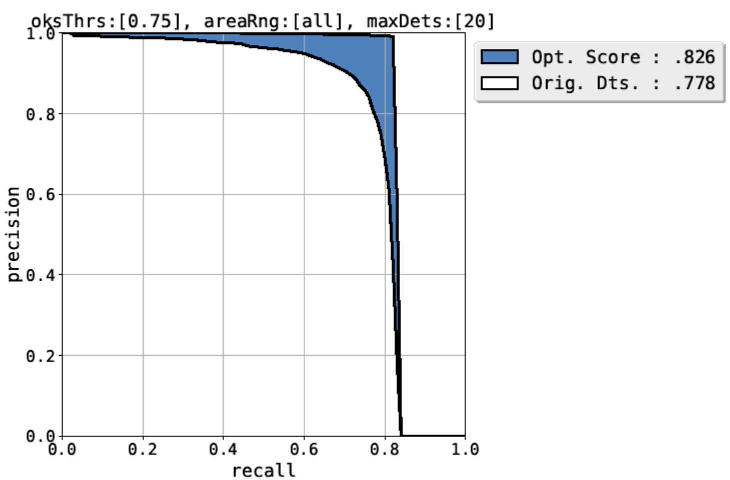
- Traditional method
  - Sort box candidates by box score
  - Select top k boxes as the result
- Our method
  - Sort box candidates by the product of box score and keypoint score
  - Select top k boxes as the result



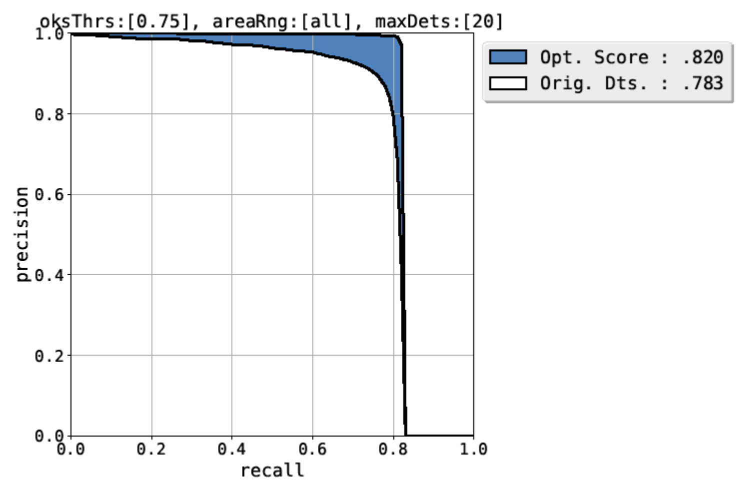
Box Candidates Rescoring

- Comparison among different rescoring criterion

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



Box score



Rescoring

OKS-NMS

- Object Keypoint Similarity (OKS)

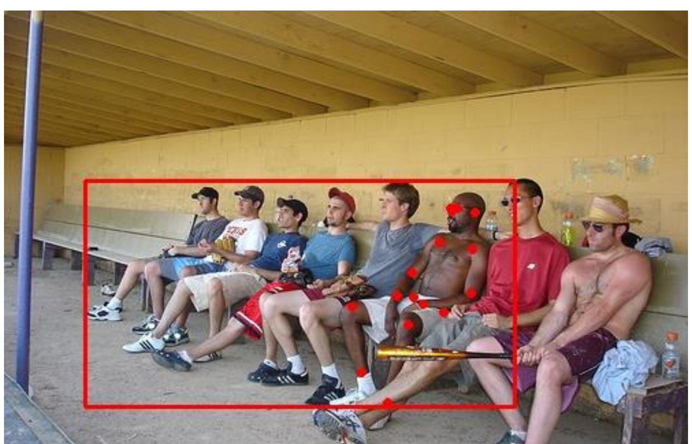
$$OKS_p = \frac{\sum_i exp\left\{-d_{pi}^2/2s_p^2\sigma_i^2\right\}\delta\left(v_{pi}=1\right)}{\sum_i \delta\left(v_{pi}=1\right)}$$

- Can be seen as “IoU” in keypoint detection to perform NMS
- OKS-NMS fails to suppress proposals with high IoU
- Combine IoU-NMS and OKS-NMS:
  - Apply 0.6 IoU-NMS first, then perform 0.5 OKS-NMS (best practice)

Papandreou G, Zhu T, Kanazawa N, et al. Towards Accurate Multi-person Pose Estimation in the Wild[J]. arXiv preprint arXiv:1701.01779, 2017.

Data Selection

False annotations in COCO dataset

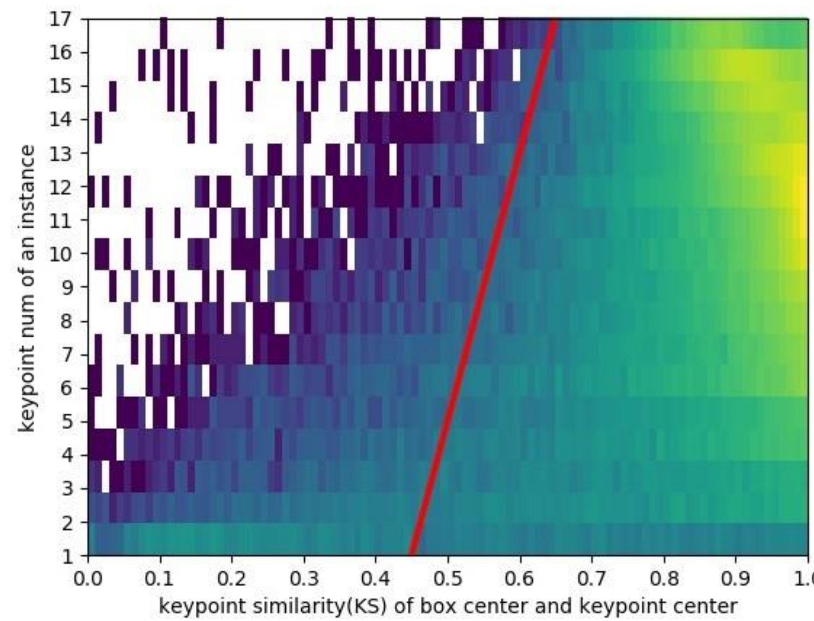


Data Selection

- We statistic the joint distribution of keypoint similarity(KS) (between box center and keypoint center) and keypoint number of an instance

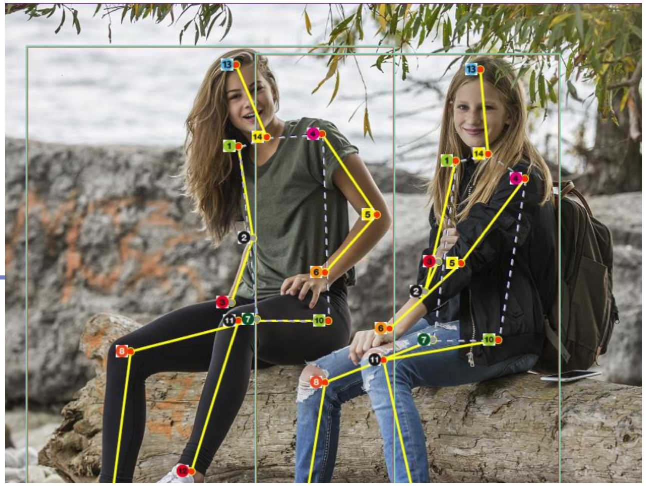
$$ks(c_{box}, c_{kps}) = e^{\frac{-||c_{box} - c_{kps}||_2^2}{2 * area_{box}}}$$

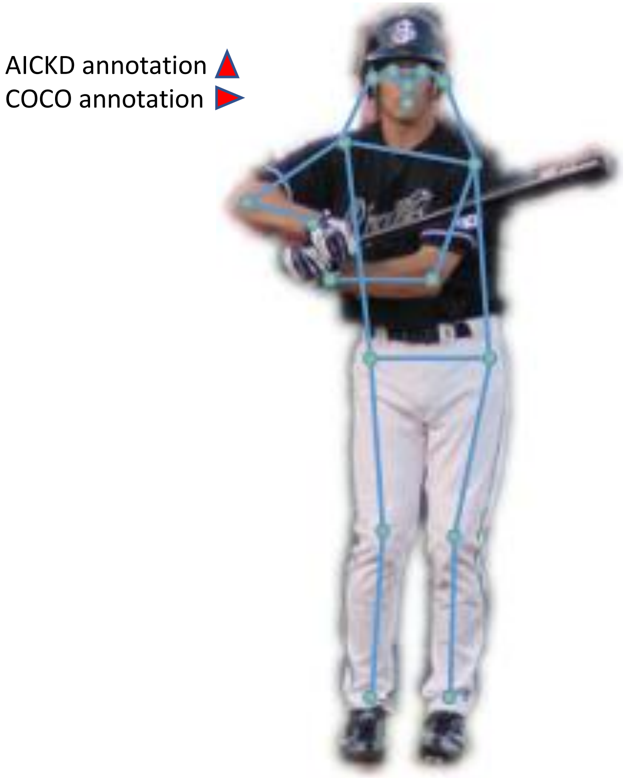
- We only keep the data right of the line (0.45,1) – (0.65,17)



External data

- We use the [AI Challenge Keypoint Dataset(AICKD)] for joint training
  1. Train a hourglass 8 stacks with COCO only data
  2. Use the model above to select hard examples in AICKD
  3. Joint train with COCO data and hard examples of AICKD
- We only backpropagate the loss of common annotations with COCO for AICKD data





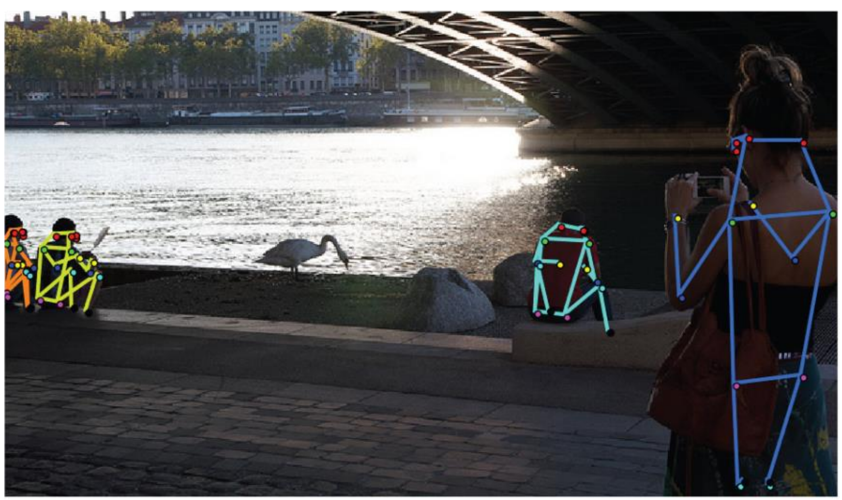
Experiment Results

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

Final submission	AP (test-dev / test-challenge)
Ours (single model, COCO + external data)	72.0 / 71.4
GRMI (COCO + external data)	68.5 / NA

Papandreou G, Zhu T, Kanazawa N, et al. Towards Accurate Multi-person Pose Estimation in the Wild[J]. arXiv preprint arXiv:1701.01779, 2017.

Results Visualization

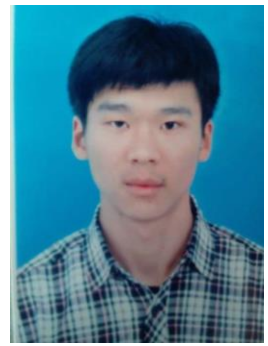


What we learned?

- For performance improvement of 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. We expect better results from automatic designed networks in the future.

Our team


Yujie Wang


Changbao Wang


Quanquan Li


Biao Leng


Zhoujun Li


Junjie Yan

Email: liquanquan@sensetime.com


Thank you!