

Video Human Pose Estimation based on DEKR

Bi Chunhao, Yang Hongdi, Wu Qinman, Li Jiaxuan



上海科技大学
ShanghaiTech University

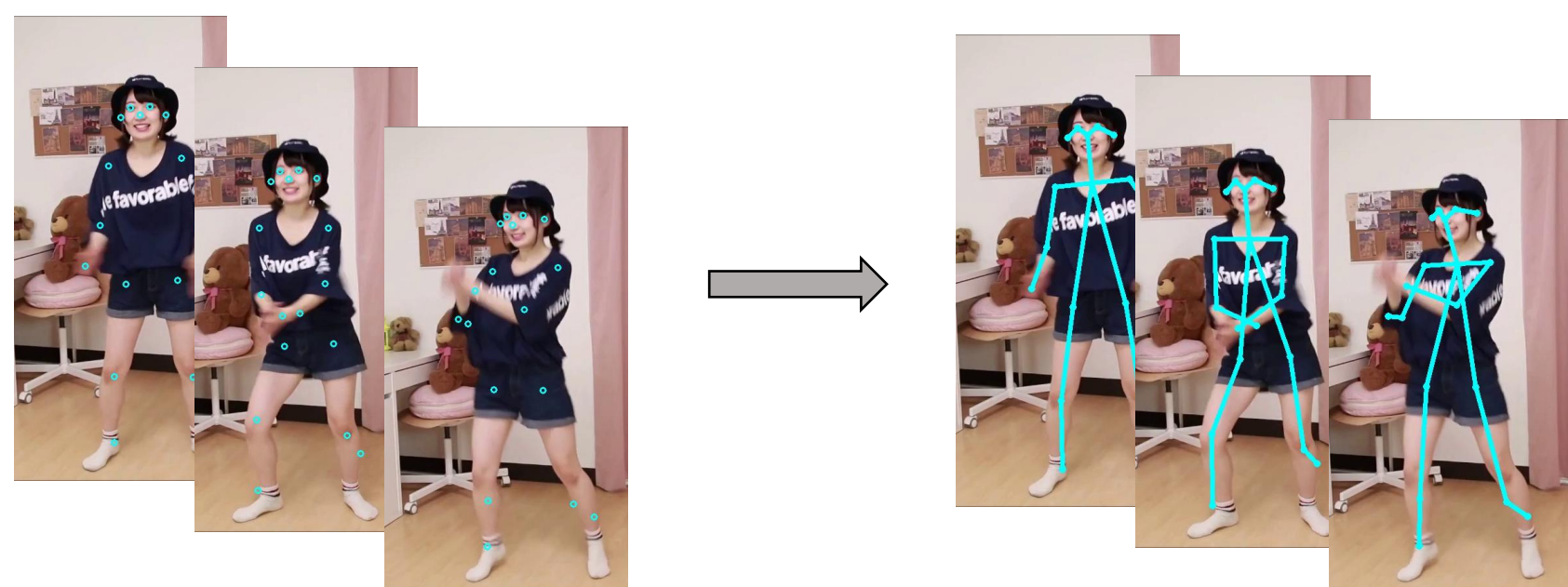
What is DEKR?

DEKR is a cutting-edge algorithm introduced in CVPR 2021 by Geng et al. It mainly focus on multi-person pose estimation on images. Some results are as follow:



What we are trying to do?

We focused on using frame information to achieve better performance in single-person pose estimation in videos.



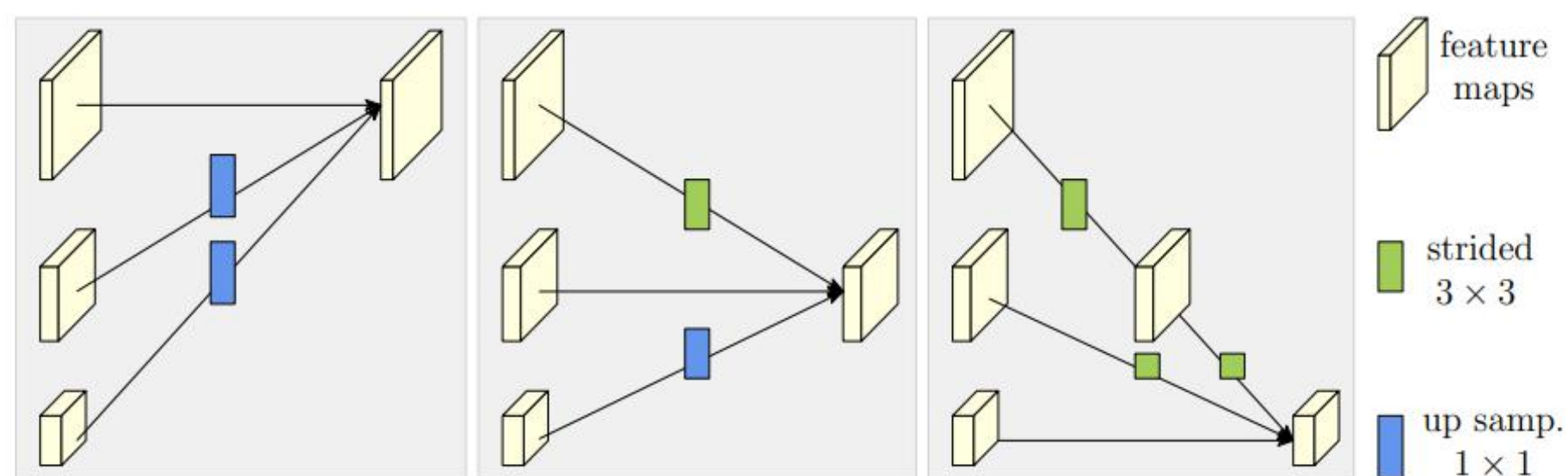
We propose a network structure based on DEKR, with a GRU unit added to attach temporal dependencies to the existing network.

Network Structure

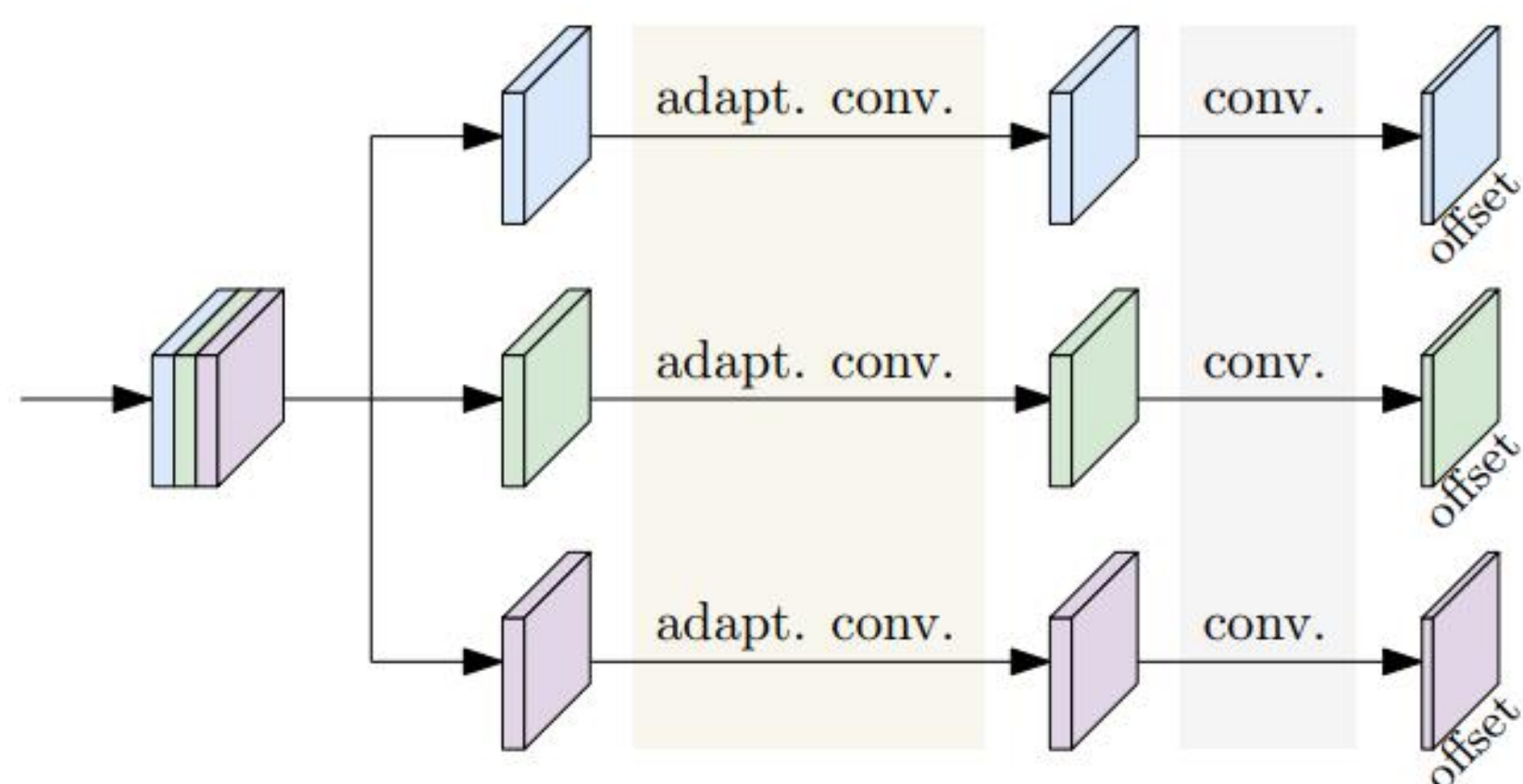
●Fusion layer from HRNet

The fusion layer fuses feature maps in each branch in parallel, mainly including two ways:

- ◆Fusion from lower resolution to higher resolution
- ◆Fusion from higher resolution to lower resolution



●Adaptive Convolution

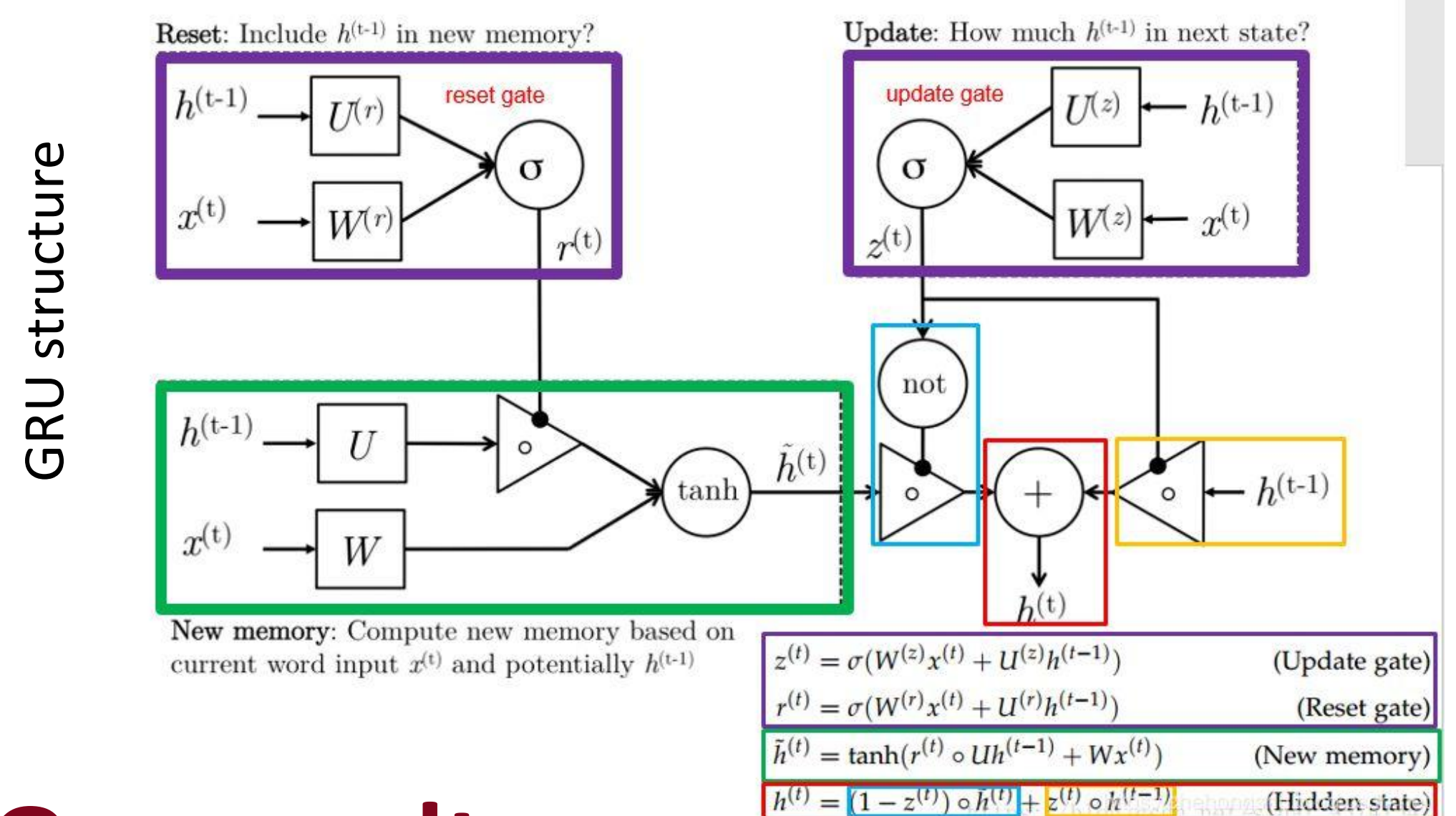


The adaptive convolution is a modification of normal convolution, which is not limited on the region around center pixel but focus around keypoints :

$$y(q) = \sum_{i=1}^9 W_i x(g_{si}^q + q).$$

●Dealing with sequence--GRU

To extract temporal information, we added a module called GRU, which is a popular technique in RNN.



Our result

We test our model on videos with person who face front most of the time and person who face side most of the time.



our prediction results on person faces front



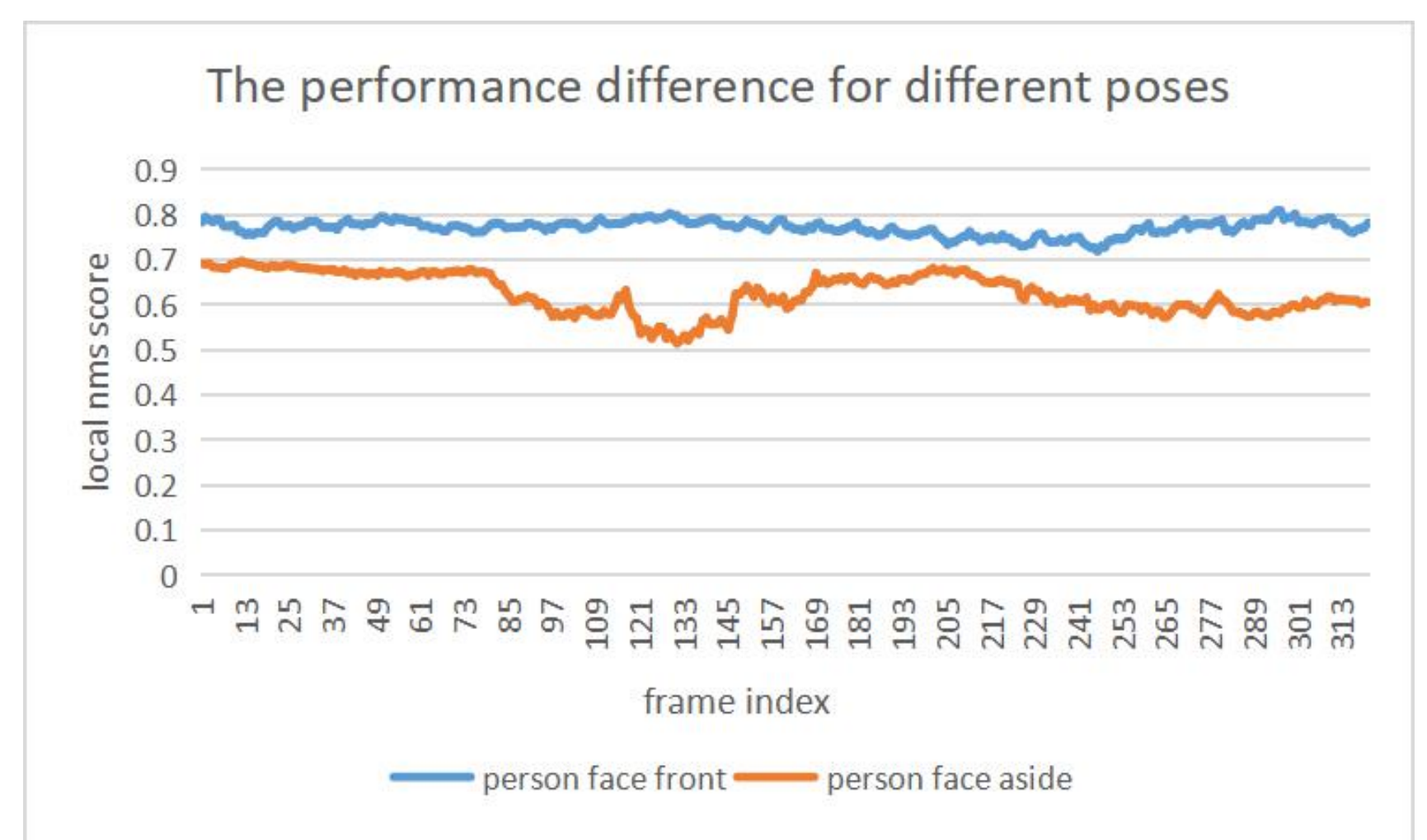
our prediction results on person faces aside

Two testing datasets containing people face front or face aside are used to test the OKS score and AP.

Table 1. testing result on two different test sets

test set	local NMS score	OKS score	AP
person front-faced	77.15	65.97	61.2
person side-faced	62.62	55.93	53.4

And we record the change of local NMS score along with the change of frames.



References

1. Zigang Geng, Ke Sun, Bin Xiao, Zhaoxiang Zhang, and Jingdong Wang. Bottom-up human pose estimation via disentangled keypoint regression, 2021.