Self-Correctable and Adaptable Inference for Generalizable Human Pose Estimation

Zhehan Kan[1] Shuoshuo Chen[1] Ce Zhang[1] Yushun Tang[1] Zhihai He[1,2*]

[1]Department of Electronic and Electrical Engineering, Southern University of Science and Technology, Shenzhen, China

[2]Pengcheng Laboratory, Shenzhen, China

SUSTech Southern University of Science and Technology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Poster Session TUE-PM-134

Highlights

- We demonstrate that it is theoretically possible to learn a feedback-correction network to refine the prediction results of a well-trained network, outperforming the SOTA and SCIO by 1.4%, which is quite significant.

- We have introduced a correction network which is able to correct the prediction error for the test sample guided by the self-referential feedback error. This error was generated by a learned fitness feedback network. We found that this self-referential error is highly correlated with the actual network prediction error.

- Using the self-referential error, we introduced a new loss function to perform quick adaptation and optimization of the correction network during the inference stage.

- We apply the proposed self-correctable and adaptable inference method to human pose estimation and have achieved remarkable performance gain and significant improvement of generalization capability of the pose estimation network.

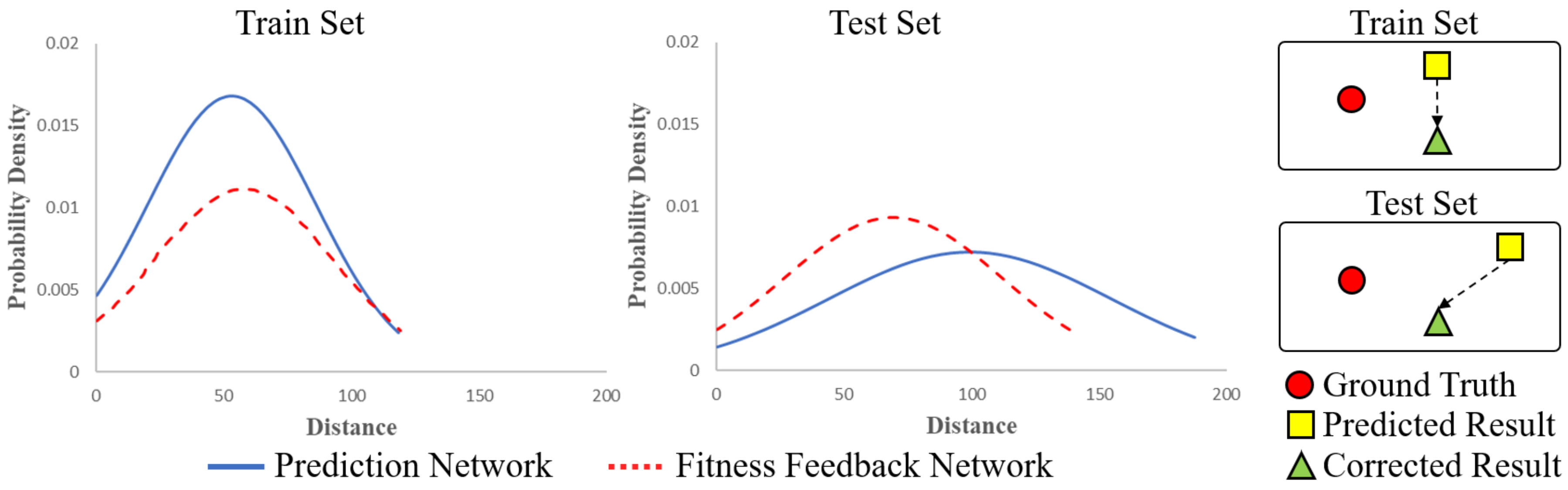
SUSTech Southern University of Science and Technology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Problem Definition

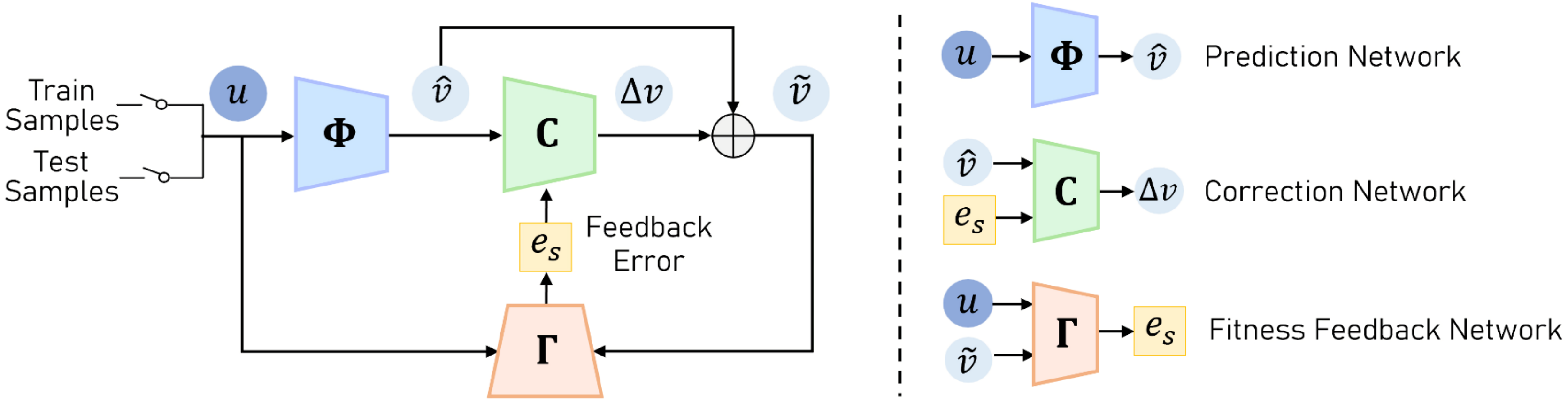
Goal: Human pose estimation aims to correctly detect and localize keypoints, i.e., human body joints or parts.

- How can we tell if the prediction is accurate or not during testing and how to characterize the prediction error?
- How to correct the prediction error based on the specific characteristics of the test sample?



Main Idea

- We propose the construction of a novel Feedback Fitness Network (FFN) that possesses the capability to assess the accuracy of predictions without prior knowledge of the ground truth.
- By leveraging the FFN, we extract self-referential errors that exhibit a strong correlation with actual errors. These errors serve as valuable indicators for training a prediction-correction network, enabling dynamic adjustment of predictions during inference.
- Furthermore, we utilize the self-referential errors in conjunction with the FFN to construct a self-supervised loss function, which facilitates rapid adaptation and optimization of network models during inference, particularly on test samples.



SUSTech Southern University of Science and Technology

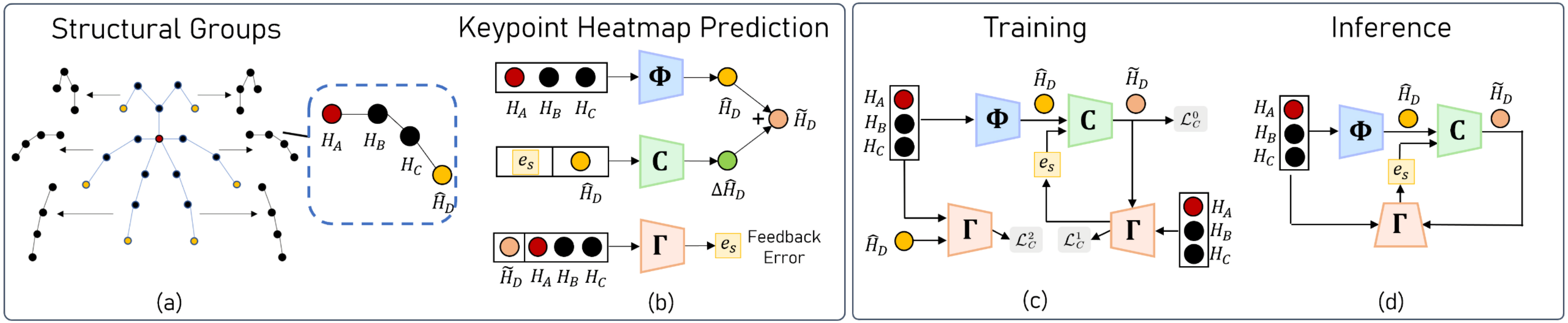
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Method

- Partitioned the body keypoints into 6 structural groups according to their biological parts, and each structural group is further partitioned into two subsets: distal keypoints and proximal keypoints.

- Developed a prediction-FFN network that effectively extracts self-referential errors and facilitates learning of a correction network under the guidance of these errors.

- During the inference phase, the self-referential error is leveraged to dynamically update the correction network by incorporating information from new samples.



SUSTech Southern University of Science and Technology

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Experimental Results

➤ Comparison with the state-of-the-art methods on COCO test-dev.

Method	Backbone	Size	AP	AP^{50}	AP^{75}	AP^{M}	AP^{L}	AR
CFN [14]	-	-	72.6	86.1	69.7	**78.3**	64.1	-
CPN(ensemble) [4]	ResNet-Incep.	384×288	73.0	91.7	80.9	69.5	78.1	79.0
CSM+SCARB [28]	R152	384×288	74.3	91.8	81.9	70.7	80.2	80.5
CSANet [42]	R152	384×288	74.5	91.7	82.1	71.2	80.2	80.7
HRNet [30]	HR48	384×288	75.5	92.5	83.3	71.9	81.5	80.5
MSPN [21]	MSPN	384×288	76.1	93.4	83.8	72.3	81.5	81.6
PoseFix [25]	HR48+R152	384×288	76.7	92.6	84.1	73.1	82.6	81.5
DARK [43]	HR48	384×288	76.2	92.5	83.6	72.5	82.4	81.1
UDP [13]	HR48	384×288	76.5	92.7	84.0	73.0	82.4	<u>81.6</u>
Graph-PCNN [35]	HR48	384×288	76.8	92.6	84.3	73.3	82.7	<u>81.6</u>
SCIO [15]	HR48	384×288	<u>79.2</u>	<u>93.5</u>	<u>85.8</u>	74.1	<u>84.2</u>	<u>81.6</u>
SCAI (Ours)	HR48	384×288	**80.6**	**94.8**	**87.0**	<u>78.1</u>	**84.8**	**83.1**
Performance Gain			**+1.4**	**+1.3**	**+1.2**	**-0.2**	**+0.6**	**+1.5**

Experimental Results

- Comparison with the state-of-the-art methods on CrowdPose test-dev.

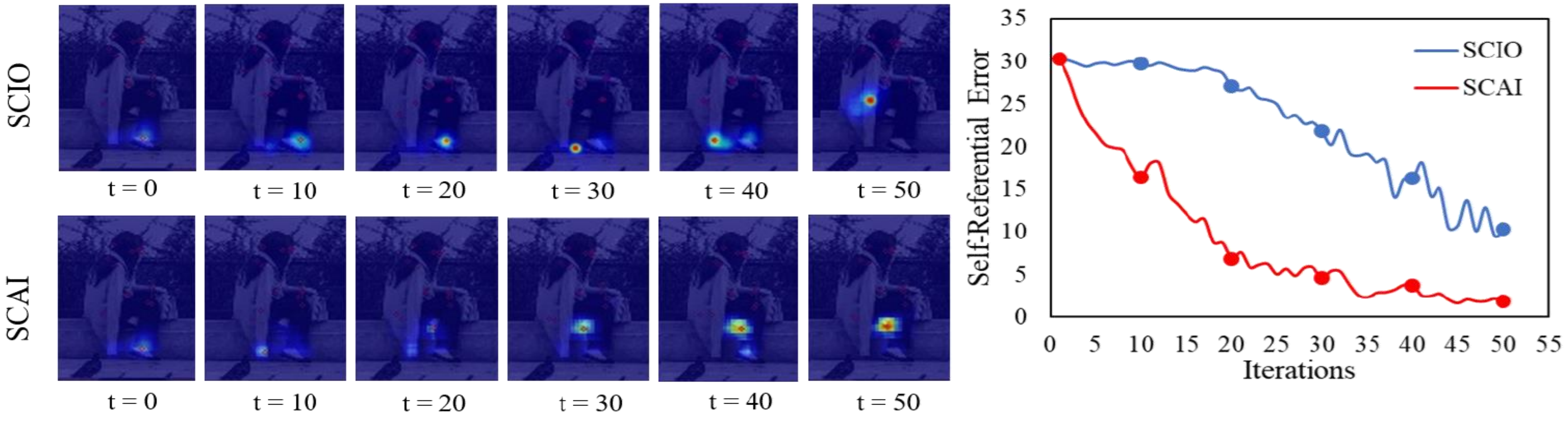
Method	Backbone	AP	AP^{med}
Mask-RCNN [12]	ResNet101	60.3	-
AlphaPose [8]	-	61.0	61.4
OccNet [10]	ResNet50	65.5	66.6
JC-SPPE [20]	ResNet101	66.0	66.3
HigherHRNet [5]	HR48	67.6	-
MIPNet [16]	HR48	70.0	71.1
SCIO [15]	HR48	71.5	72.2
SCAI (Ours)	HR48	**72.4**	**73.2**
Performance Gain		**+0.9**	**+1.0**

- Comparison with state-of-the-art of three baselines on COCO test-dev.

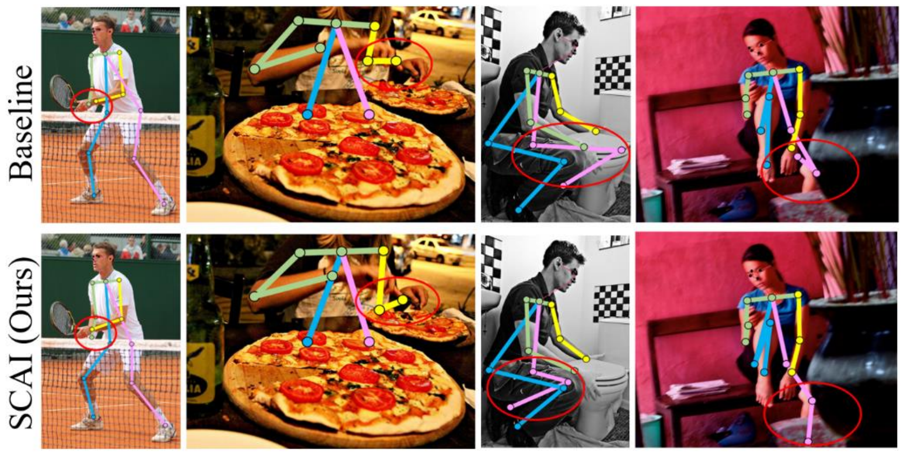
Method	Backbone	AP	AP^M	AR
SimpleBaseline	R152	73.7	70.3	79.0
+ **SCAI**	R152	**78.8**	**75.5**	**82.5**
Performance Gain		**+5.1**	**+5.2**	**+3.5**
HRNet	HR32	74.9	71.3	80.1
+ **SCAI**	HR32	**79.9**	**76.8**	**82.5**
Performance Gain		**+5.0**	**+5.5**	**+2.4**
HRNet	HR48	75.5	71.9	80.5
+ **SCAI**	HR48	**80.6**	**78.1**	**83.1**
Performance Gain		**+5.1**	**+6.2**	**+2.6**

Experimental Results

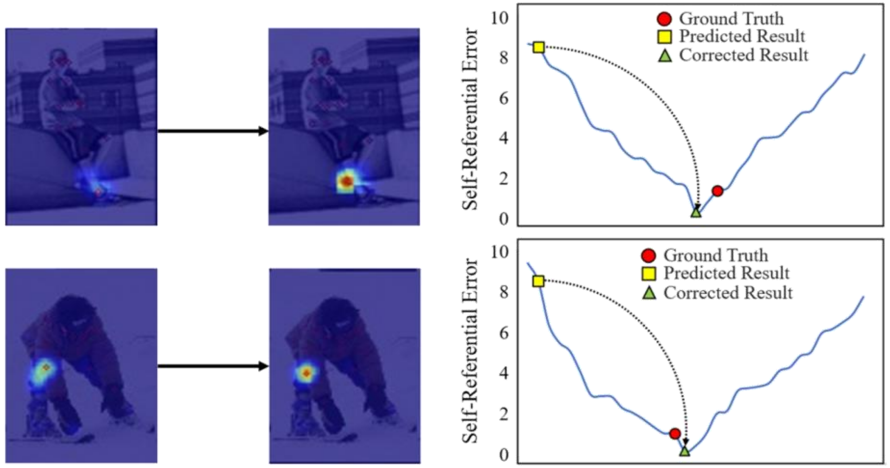
- Comparison between the local search and the correction from SCIO and SCAI.



- Refinement of predicted keypoints.

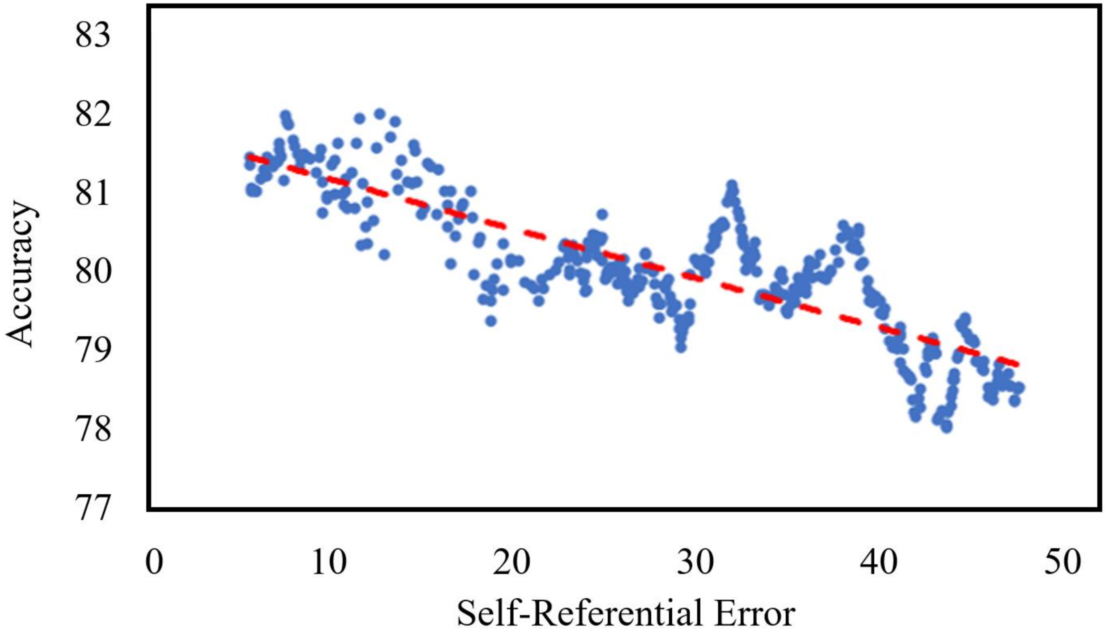


- Two examples of corrected keypoints.

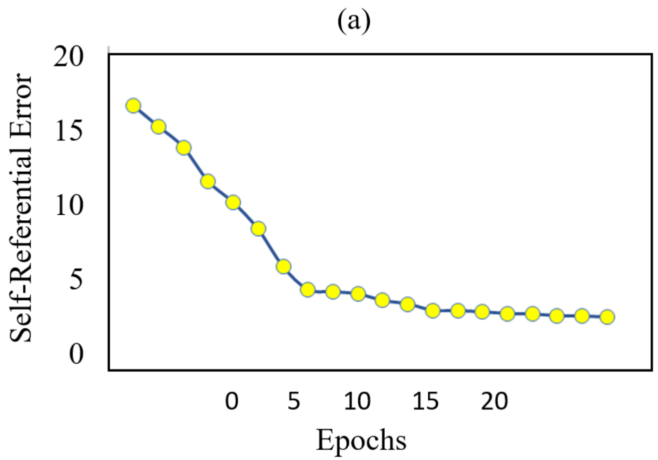


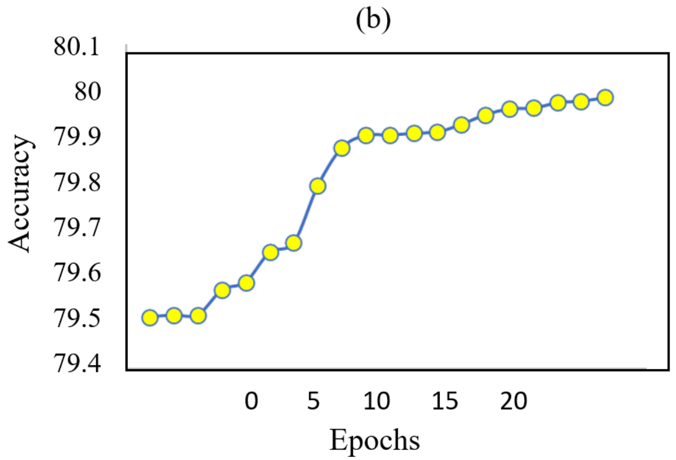
Experimental Results

- Strong correlation between self-referential error and network prediction accuracy.



- (a) is the loss training during inference, (b) is the test result corresponding to each epoch.





Conclusion

- We demonstrate that it is possible to learn a feedback-correction network to refine the prediction results of a well-trained network.
- We propose SCAI, a learning-based feedback-control method for prediction to address the generalization challenge.
- We apply the SCAI method to human pose estimation and achieve SOTA performance, improves the current best method by up to 1.4%.