Keypoint-wise Adaptive Loss for Whole-Body Human Pose Estimation

Bongjo Seong, Hyug Jae Lee, Rokkyu Lee

Motivation

Motivation

1) Whole-body keypoints have **different scales** even for the same person (i.e., **different labeling noise** for each body part).
2) Whole-body keypoints are **mixed dense/coarse keypoints**, but are encoded into the heatmap as a 2D Gaussian distribution with **the same sigma $\boldsymbol{\sigma}$**.
3) The keypoint heatmap has an **imbalance problem** between foreground and background pixels.

Observation

- Replacing MSE loss with AWing loss, we observed estimation performance degradation, except in dense keypoints.

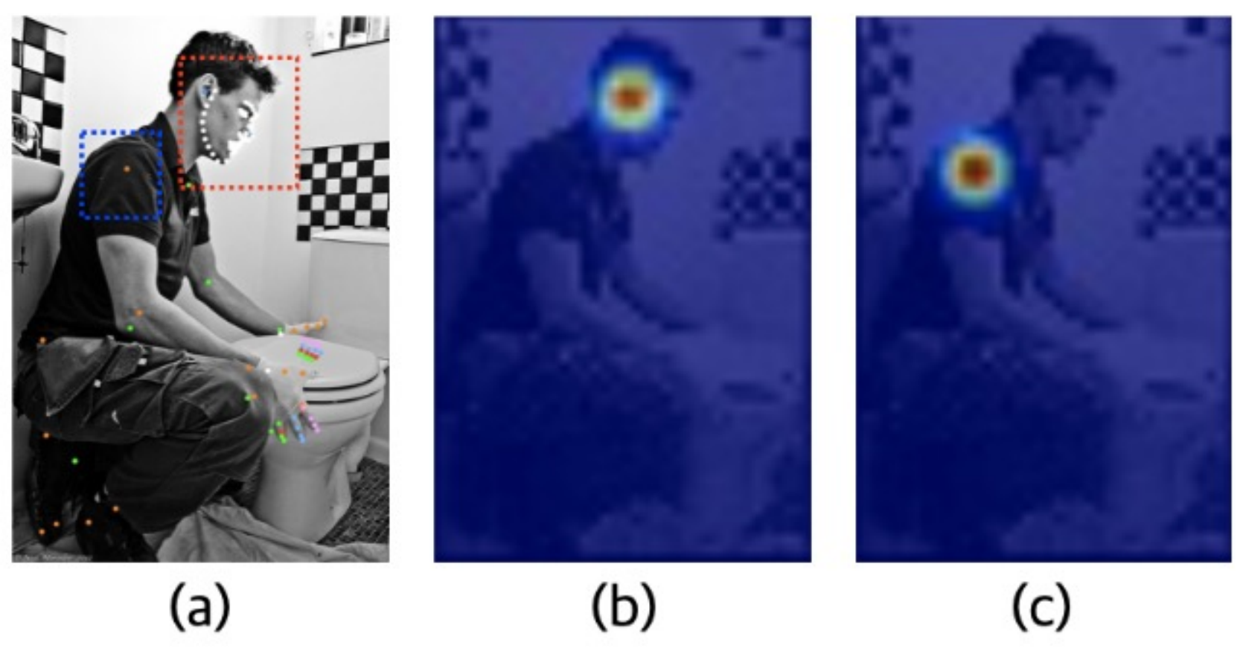


Figure 1: Comparison of a heatmap covering adjacent keypoints. (a) GT coordinates of whole-body keypoints. (b) GT heatmap covering adjacent dense keypoints. (c) Heatmap that rarely covers adjacent coarse keypoints.

Proposed Method

A network architecture of the proposed method

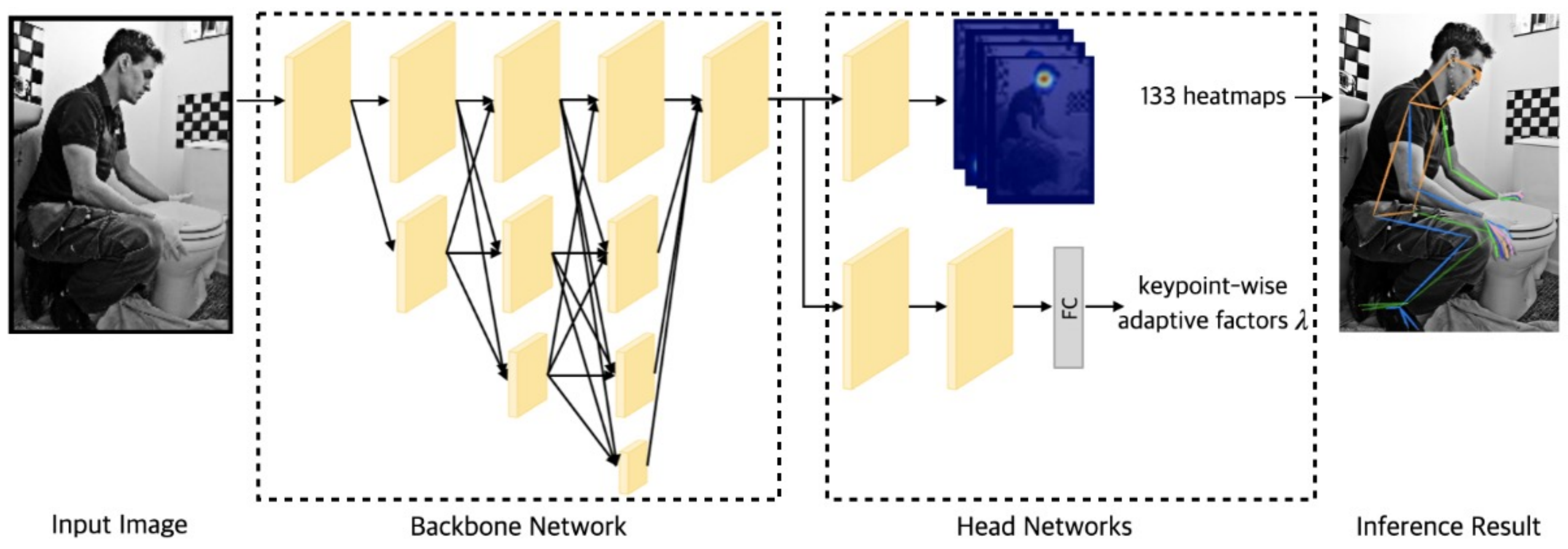


Figure 2: A human pose estimator architecture of the proposed KAL/FoWKAL method.

Heuristic Loss

- We assume that the **dense body parts** have the advantage of focusing on the foreground for accurate predictions, but the **coarse body parts** suffer from label ambiguities.
- Dense keypoints adopt the AWing loss and coarse keypoints adopt MSE loss.

$$L_{heuristic}(P,\hat{P}) = \underbrace{\lambda_{fh} L_{bf}(P,\hat{P})}_{\text{MSE loss}} + \underbrace{\lambda_{bf} L_{fh}(P,\hat{P})}_{\text{AWing loss}},$$

Keypoint-wise Adaptive Loss (KAL)

- Keypoint-wise Adaptive Factor (KAF) controls the extend of the **focus on the foreground in the heatmap.**

$$L_{Adaptive}(P_k,\hat{P}_k) =$$
$$\lambda_k L_{AWing}(P_k,\hat{P}_k) + (1-\lambda_k) L_{MSE}(P_k,\hat{P}_k),$$

- We add a regularization term context of **relationships between body parts.**

$$L_{KAL}(P,\hat{P})$$
$$= \sum_{part}\{\frac{1}{N_{part}} \sum_{k \in part_{kpt}} L_{Adaptive}(P_k,\hat{P}_k)\} + L_{reg}(\lambda).$$

Foreground-Weight Keypoint-wise Adaptive Loss (FoWKAL)

- Foreground-Weight Adaptive Heatmap Regression (FWAHR) lead the **model to focus on relatively harder samples on the foreground pixels in the heatmap.**

$$W(p,\hat{p}) = \begin{cases} p^{\gamma} \cdot |1-\hat{p}| + |\hat{p}| \cdot (1-p^{\gamma}) & if\ \hat{p} \geq 2^{-\frac{1}{\gamma}}, \\ \tau p & otherwise, \end{cases}$$

- When KAL and FWAHR are used together, it is called FowKAL.

$$L_{FoWKAL}(P,\hat{P})$$
$$= \sum_{part}\{\frac{1}{N_{part}} \sum_{k \in part_{kpt}} L_{WAdaptive}(P_k,\hat{P}_k)\} + L_{reg}(\lambda).$$

Experimental Results

Results on the COCO-WholeBody V1.0 dataset

Method	whole-body		body	foot	face	hand
	AP	AR	AP	AP	AP	AP
Bottom-up methods:						
AE (Newell, Huang, and Deng 2017)	27.4	35.0	40.5	7.7	47.7	34.1
OpenPose (Cao et al. 2017)	33.8	44.9	56.3	53.2	48.2	19.8
Keypoint Communities (Zauss, Kreiss, and Alahi 2021)	60.4	-	69.6	63.4	85.0	52.9
Top-down methods:						
ZoomNet† (Jin et al. 2020)	54.1	65.8	**74.3**	**79.8**	62.3	40.1
HRNet-w32 (Sun et al. 2019)	55.3	62.6	70.0	56.7	63.7	47.3
TCFormer (Zeng et al. 2022)	57.2	67.8	69.1	69.8	64.9	**53.5**
HRNet-w32+DARK (Zhang et al. 2020)	58.2	67.1	69.4	56.5	73.6	50.3
HRNet-w32+DARK+FoWKAL (Ours)	**61.6**	**71.1**	72.7	74.2	**73.8**	**53.5**

Table 1: Performance comparisons with the state-of-the-art bottom-up/top-down methods. The results are reported on the COCO-WholeBody V1.0 dataset (Jin et al. 2020). HRNet-w32 and HRNet-w32+DARK results are from MMPose (Contributors 2020). ZoomNet† is trained with the COCO-WholeBody V0.5 training set.

Ablation study

Method	MSE	AWing	KAL	FWAHR	whole-body AP	body AP	foot AP	face AP	hand AP
(a)	√				58.2	69.4	56.5	73.6	50.3
(b)		√			57.9	67.6	52.4	**76.8**	50.9
(c)	√	√			58.7	70.2	58.6	76.5	48.4
(d)			√		58.4	71.8	73.4	69.6	45.8
(e)	√	√		√	61.2	71.1	69.0	76.4	53.2
(f)			√	√	**61.6**	**72.7**	**74.2**	73.8	**53.5**

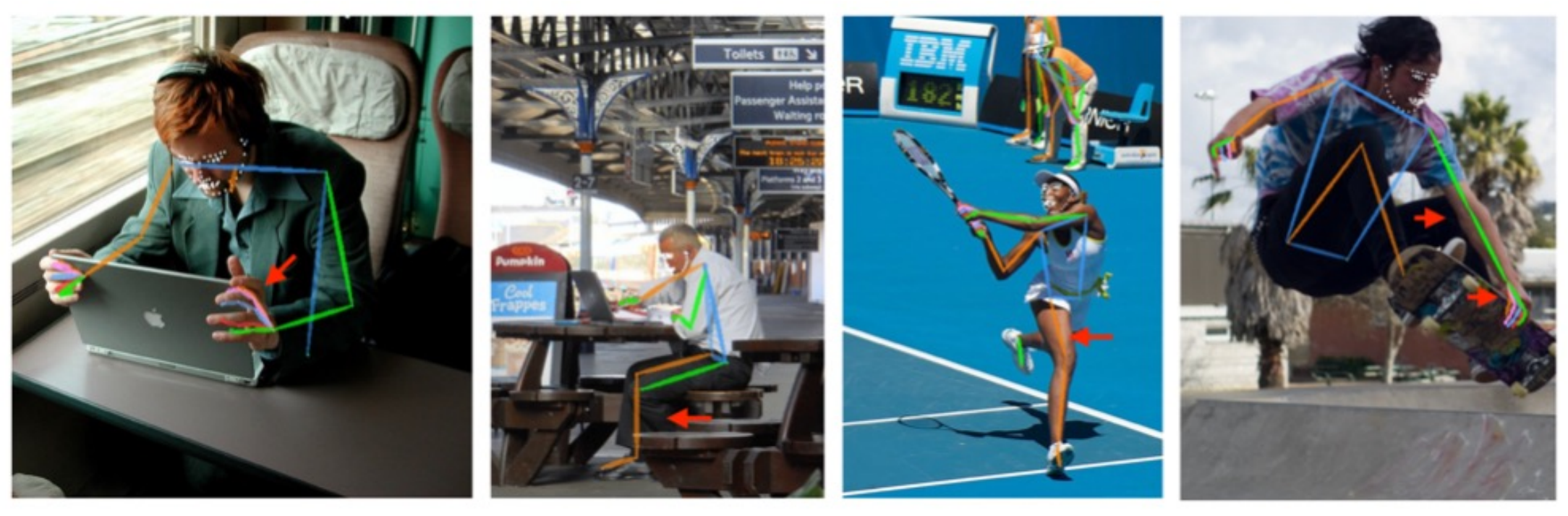
Table 2: Ablation study on Mean Squared Error/Adaptive Wing loss, Keypoint-wise Adaptive Loss (KAL), and Foreground-Weight Adaptive Heatmap Regression (FWAHR), respectively. Method (a) is the baseline with MSE loss, method (b) is the AWing loss, method (c) is the heuristic loss, method (d) is the KAL, method (e) is the heuristic loss and the FWAHR, and method (f) is the Foreground-Weight Keypoint-wise Adaptive Loss (FoWKAL).

Comparison of FWAHR and WAHR

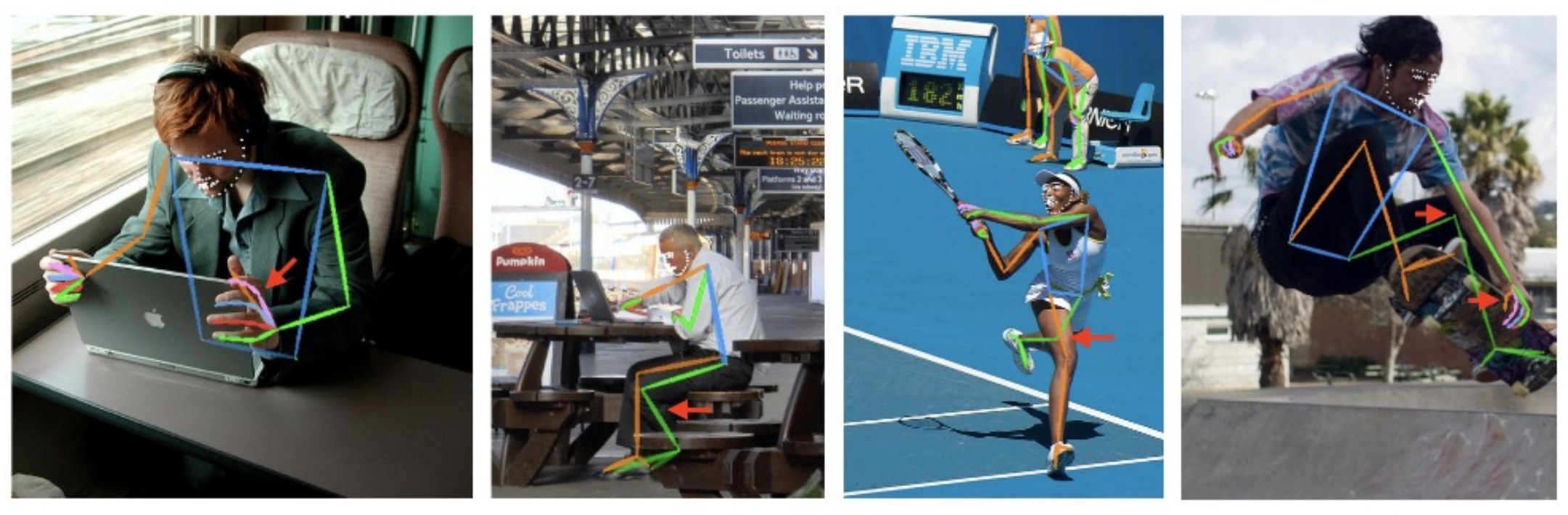
Method	Whole-body AP
Heuristic + WAHR	59.6
Heuristic + FWAHR	61.2
KAL + WAHR	59.5
KAL + FWAHR	61.6

Table 3: Comparison of Foreground-Weight Adaptive Heatmap Regression (FWAHR) and WAHR (Luo et al. 2021) with heuristic or Keypoint-wise Adaptive Loss (KAL).

Qualitative comparison between FoWKAL and HRNet-w32+Dark



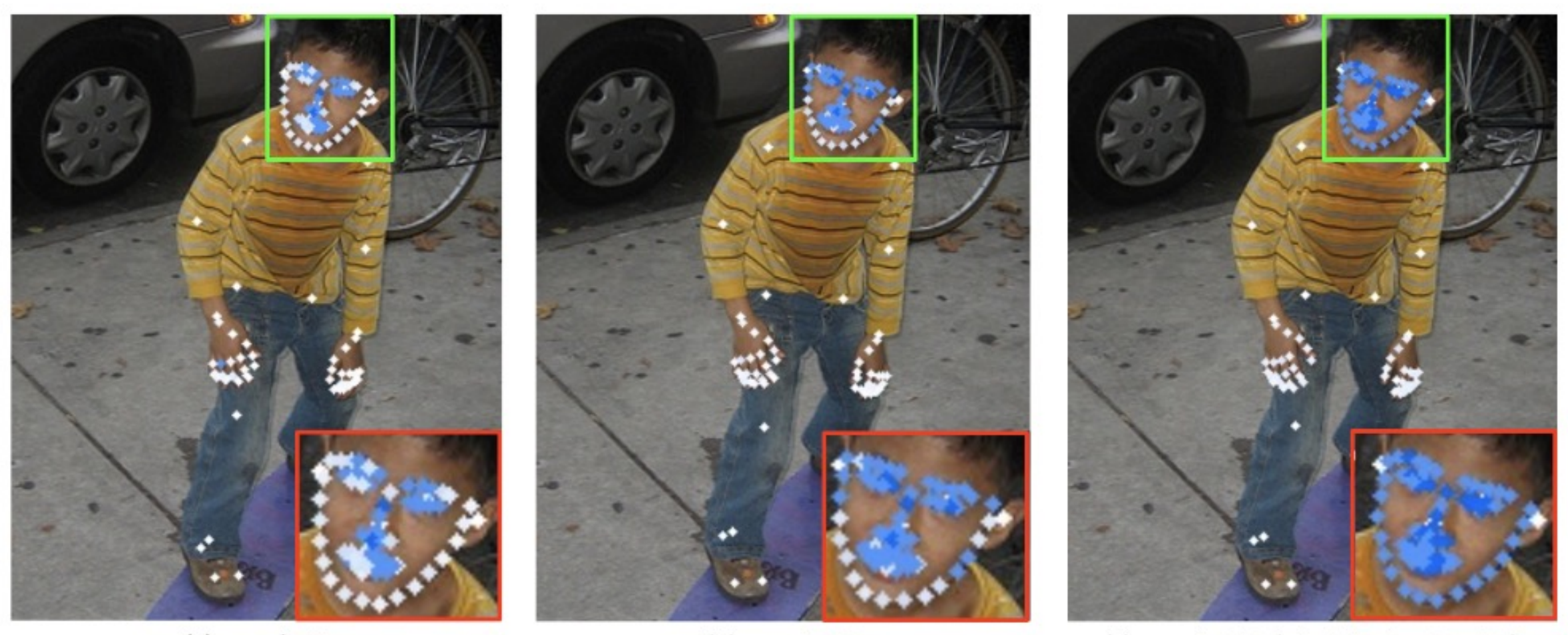
(a) Baseline



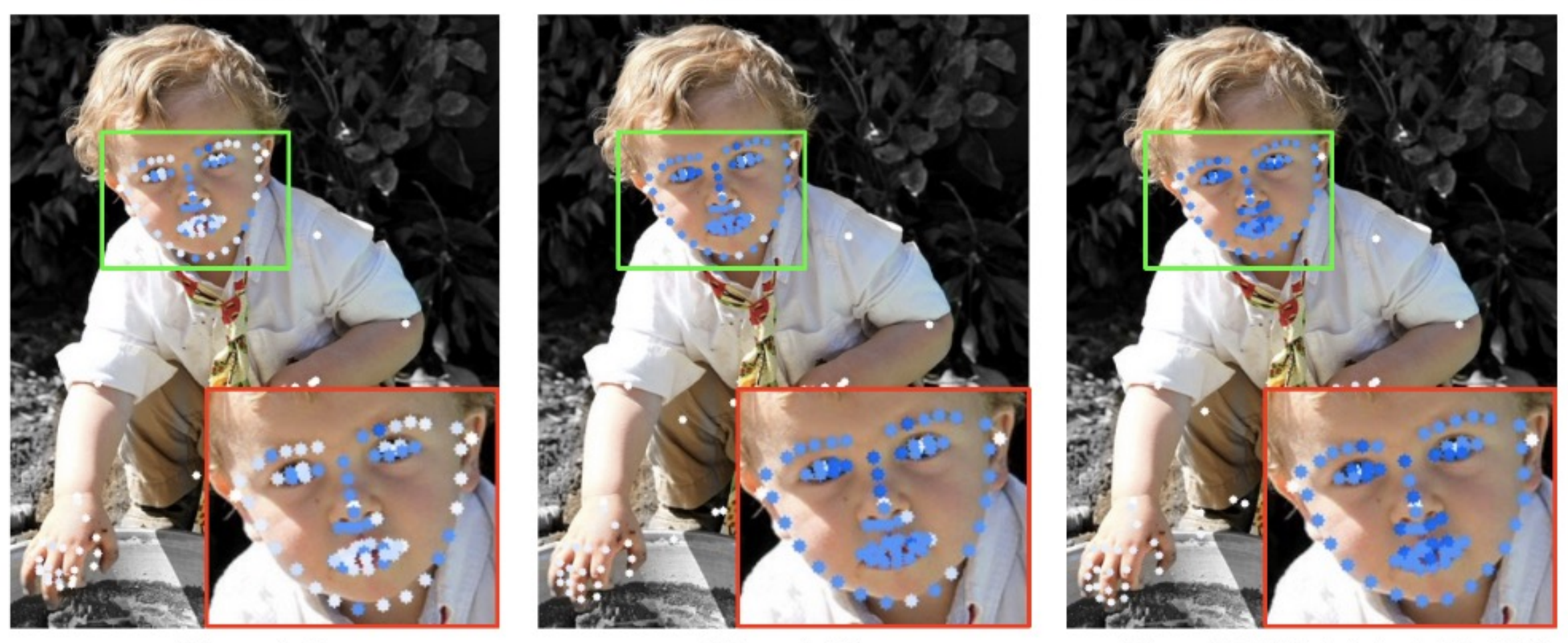
(b) The proposed FoWKAL

Comparison of Keypoint-wise Adaptive Factors (KAF)

- A Whiter point color is closer to MSE, and a bluer point is closer to the AWing loss.



(a) epoch 10 (b) epoch 70 (c) epoch 210 (wholebody AP : 1.0)



(d) epoch 10 (e) epoch 70 (f) epoch 210 (wholebody AP : 0.9)