Localizing Human Keypoints Beyond the Bounding Box

Soonchan Park [1,2] Jinah Park [2]
[1]Contents Research Division, ETRI
[2]School of Computing, KAIST

Acknowledgement
This research is supported by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: R2020070002)

Motivation

- From partial image, we naturally estimate entire human pose from the clues such as the length of the limbs, bending angle of joints.
- Contrast to human recognition system, the estimation range of the neural networks is restricted by the given image.

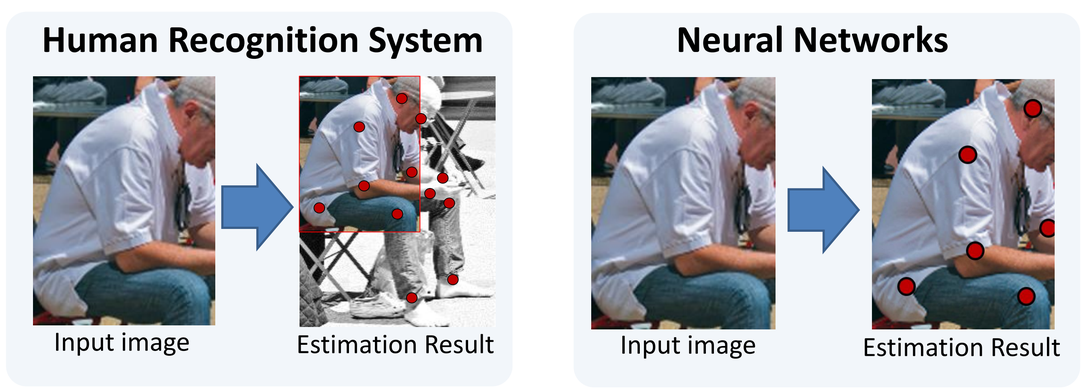


Figure 1. Estimation of human recognition system is performed beyond the image, but the estimation range of the neural network is restricted by given image.

Proposed Method

- Data preparation (i.e., 3x3 cropping) to create the patterns between partial images and entire human pose. (Sec.3.1)
- Position Puzzle Network(PPNet) to refine the bounding box to include entire human body. (Sec.3.2)
- Position Puzzle Augmentation(PPAug) to effectively trains keypoint detector to be able to localize keypoints out of the image. (Sec.3.3)

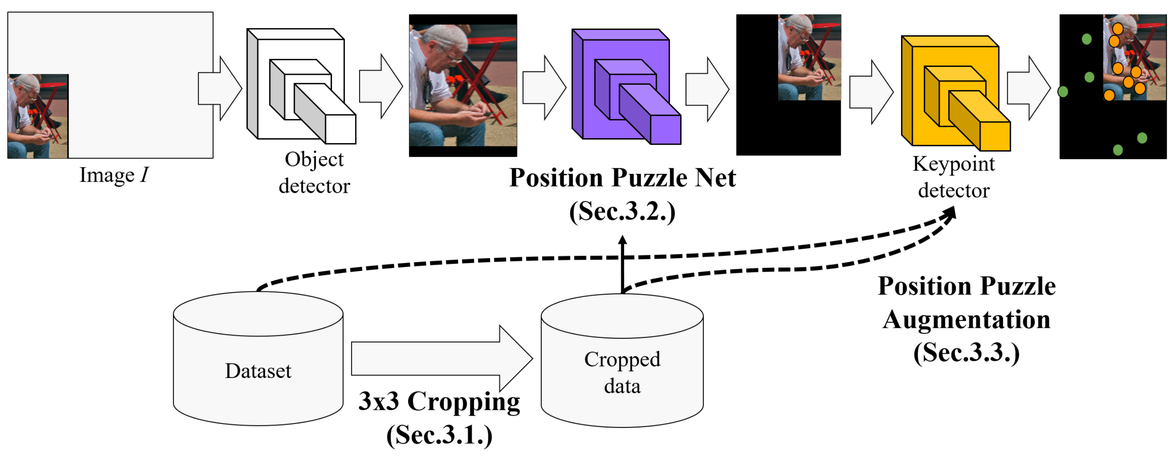


Figure 2. The overview of proposed method for the top-down-based keypoint detection. 3x3 Cropping, Position Puzzle Network(i.e., PPNet), and Position Puzzle Augmentation(i.e.,PPAug) are three methods we proposed in this paper.

Data preparation: 3x3 Cropping

- The method collects human objects which has their full body in the image, and divides them using 3x3 grid.
- The method samples sub-bounding boxes by holding one edge (8 samples) and one corner (16 samples) of 3x3 grid, and lastly, adds the original bounding box itself.

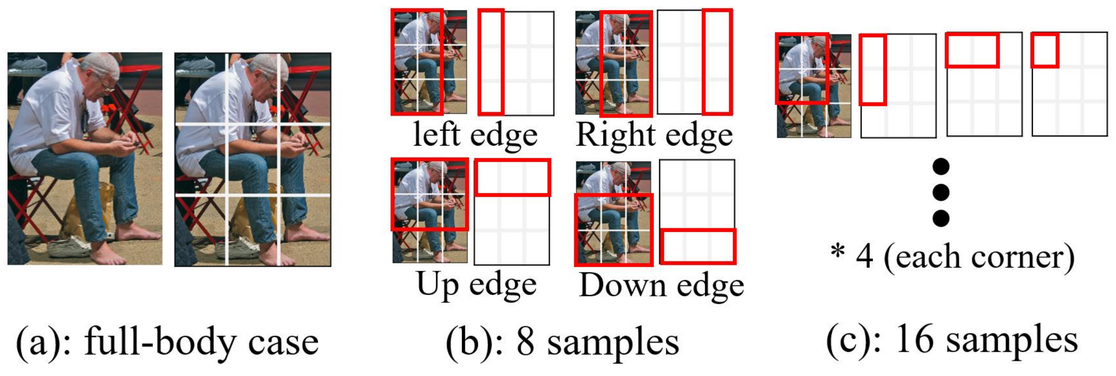


Figure 3. (a) a sample which has full body of the target object. (b) cropped boxes are collected from the 3x3 grid by holding one edge of the bounding box. (c) cropping examples by maintaining one corner of the bounding box.

Position Puzzle Network and Position Puzzle Augmentation

- PPNet and PPAug utilize the dataset collected by the 3x3 cropping.
- PPNet is a function to reconstruct the bounding box using 4D vector so that the box contains entire body of the target. (Fig.4)
- PPAug produces the data having partial image and entire pose as Fig.5 illustrates for training keypoint detector.

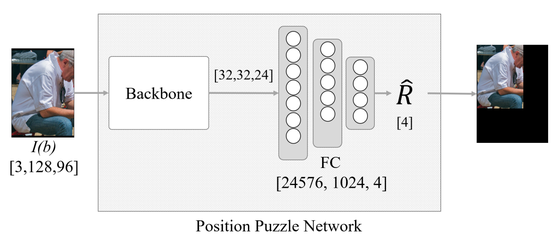


Figure 4. Architecture of PPNet.

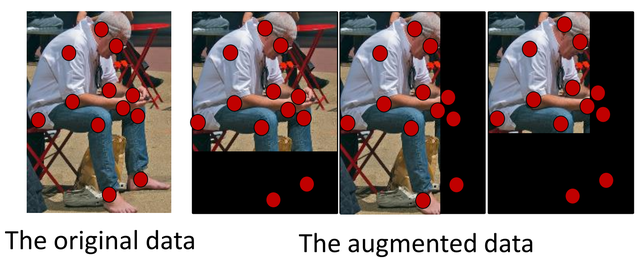


Figure 5. PP Augmentation generates patterns between partially cropped out image and entire human pose.

Evaluation on Crop-COCO dataset

- Another dataset is arranged to evaluate the performance of detecting entire keypoints using the partial images.
- S_{random} = Set of the cropped images by random.
- S_{easy}, $S_{moderate}$, and S_{hard} are Set of 20%, 40%, and 60% erased images, respectively.
- Proposed method remarkably outperforms the baseline by 37.6% and 30.6% in mAP and mAR, respectively.

	S_{random}		S_{easy}		$S_{moderate}$		S_{hard}	
	mAP	mAR	mAP	mAR	mAP	mAR	mAP	mAR
HR-Net[20]	21.3	25.8	69.0	73.3	41.2	46.1	11.4	15.8
PPNet$_{\alpha=0.7}$ +HR-Net$_{aug}$	**29.3**	**33.7**	**78.1**	**81.6**	**55.6**	**60.3**	**21.4**	**25.1**

Table 1. Mean averaged precision(mAP) and mean averaged recall(mAR) comparisons in Crop-COCO dataset.

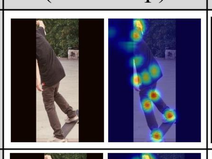
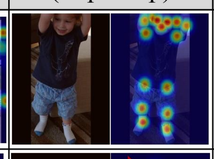
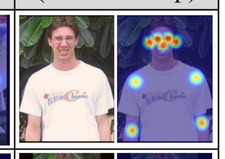
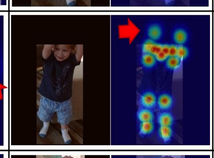
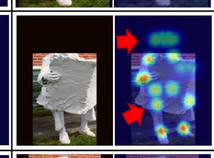
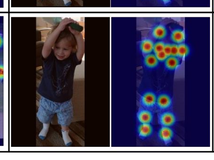
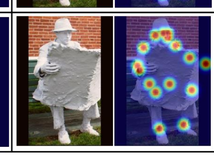
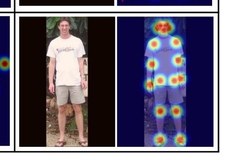
Val. Data # (Crop Dir.)	7977 (Left Crop)	427160 (Right Crop)	60899 (Top Crop)	94326 (Top Crop)	17905 (Bottom Crop)
HR-Net					
PPNet +HR-Net$_{aug}$					
Ground-Truth					

Figure 5. Estimation results from the baseline and the baseline with proposed method. The red arrows indicate the region of enhancement. The 3rd row is the reference image from the original COCO dataset.

Evaluation on the original COCO dataset

- The proposed method slightly enhances the performance in the original COCO dataset although its **evaluation metric focuses on the only keypoints in the image**.
- Proposed method outperforms the baseline by 0.6% and 0.5% in mAP and mAR for validation set, and by 0.8% and 0.9% in mAP and mAR for test set.

		mAP	$AP_{.5}$	$AP_{.75}$	AP_m	AP_l	mAR	$AR_{.5}$	$AR_{.75}$	AR_m	AR_l
COCO-val$_{AP=56.4}$	HR-Net	72.5	88.6	78.8	68.6	**80.2**	78.4	93.0	84.2	73.6	**85.3**
	PPNet$_{\alpha=0.7}$ +HR-Net$_{aug}$	**73.0**	**89.1**	**80.0**	**69.8**	79.6	**78.8**	**93.4**	**85.2**	**74.6**	84.8
COCO-test$_{AP=60.9}$	HR-Net	71.8	91.0	78.8	68.0	**78.3**	77.4	94.5	83.9	73.0	83.6
	PPNet$_{\alpha=0.7}$ +HR-Net$_{aug}$	**72.4**	**91.3**	**80.0**	**69.3**	78.2	**78.1**	**94.9**	**85.0**	**74.0**	**83.7**

Table 2. Performance comparison using validation set and test set of the original COCO dataset.

Contact Information
Soonchan Park
Contents Research Division, KAIST
Web http://sites.google.com/cv-scpark E-mail parksc@etri.re.kr

ETRI Electronics and Telecommunications Research Institute
KAIST School of Computing
Computer Graphics & Visualization Lab