Improving Detection of Hand Joints in RGB Images using Maximum Confidence Value

Khin Sabai Htwe [*1] Takaaki ISHIKAWA [*2] Hiroshi WATANABE [*1, 2]
[*] Waseda University

Abstract In this paper, we proposed an approach to detect hand joints by updating the best bounding box of the hand area based on a confidence value in OpenPose on egocentric views. The result shows that our approach can reduce processing time and improve the accuracy of hand joint detection on egocentric views.

1. Introduction

Hand joint detection plays an important role in hand pose estimation research field such as hand-hand interaction, hand-object interaction and hand action recognition. In order to detect hand joints on egocentric view correctly, firstly the hand area is needed to be localized. In this paper, we emphasize how to choose the best bounding box to know where is the hand area in the scene.

2. Related Work

Since many hand interaction applications emerge, hand pose estimation becomes a popular research topic in computer vision field. Similarly, hand keypoints detection becomes a part of essential role in hand pose estimation. For hand area detection, authors in [1] have analyzed some hand segmentation approaches by using RefineNet on many datasets. However, some hand segmentation results were not successful in their analysis when the input image is blurred by motion, skin appearance may be the same with other body parts, or lighting is changed. Recently in [2], a hand keypoint detector is proposed using multiview bootstrapping method. They use multiple cameras to create large dataset of annotated hand keypoints. It requires a method to generate the bounding box for the hand area. Therefore, they directly use body pose estimation models in [3] to approximate the hand location from wrist and elbow positions. In this situation, their proposed method depends on the hand area bounding box from [3] to detect the hand keypoints. In this paper, we consider how to detect the hand area without elbow and shoulder information in the scene, especially on egocentric views.

3. Proposed Method

3.1. Hand Bounding Box Detection

We apply fine-tuned You Only Look Once (YOLO) model using pre-trained weights [7] to train hand dataset. It can localize the hand area using weight values with the best Intersection over Union (IoU). IoU is defined as the overlapping area between two region bounding boxes, which are ground truth and predicted area.

3.2. Hand Joint Detection

After obtaining initial bounding box by the approach described in Section 3.1, we use that bounding box to specify the hand area and it is provided to OpenPose that can detect hand joints in the scene. OpenPose can give coordinates (x,y) and confidence value for the hand joint of RGB input images. If the confidence value of hand joints is greater than 0.5, existing bounding box at that confidence is updated as a new bounding box for the next joint detection. Otherwise, we find the location of the best bounding box by indexing the location of the maximum confidence value for all frames to update as a new bounding box. Fig. 1 shows the flow of hand joint detection.

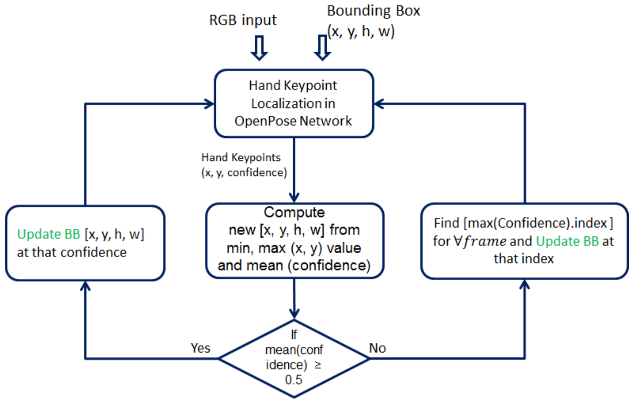

Fig. 1. Flow of Hand Joint Detection

4. Experiment

In this experiment, we set the ground truth of the hand area and trained YOLO on two datasets: ego-Dexter dataset (640x480) [4] and Hand Action dataset (320x240) [5]. The total frames are 6747 images for training and 764 images for testing.

4.1. Hand Bounding Box Detection Results

We show some results in Table 1 and Fig. 2 and measure the accuracy of hand detection results by calculating IoU and mean Average Precision (mAP) to choose the best weight value which is used to predict the bounding box in test. IoU and mAP are formulated as:

$$IoU = \frac{Area(b_p \cap b_t)}{Area(b_p \cup b_t)} \quad , \quad mAP = \frac{all\ (AP)}{classes}$$

where, $AP = \frac{\sum_{Recall} Precision(Recall)}{11}$.

Table 1. mean Average Precision of Hand Detection

Method	Hand Weight Values	Only Ego-Dexter Dataset [6]		Ego-Dexter and Hand Action Dataset	
		IoU (%)	mAP (%)	IoU (%)	mAP (%)
Tiny Yolo fine-tuned	10000	79.08	90.91	77.43	90.91
	20000	82.65	90.91	80.57	99.99
	30000	81.75	90.91	83.80	90.91
	40200	79.05	90.88	79.54	90.80

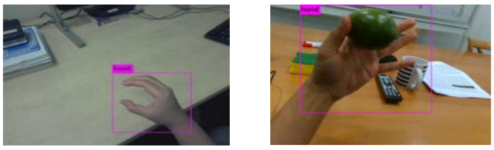

Fig. 2. Hand Bounding Box Detection Results

4.2. Hand Joint Detection Results

We detect the hand joints by updating new bounding box based on the maximum confidence value of joint coordinates. We show some experimental results in Fig. 3 and Table 2 by comparing mean value of hand joint detection results in [2] and [6].

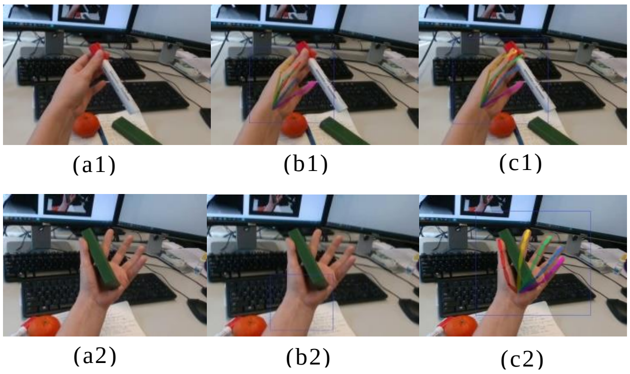

Fig. 3. Joint Detection Results on (a) without bounding box based on [2], (b) and (c) with bounding box based on [6] and our approach, respectively

Table 2. Mean and Standard Deviation Value of Confidence of Hand Joint

Input Video	Method in [2]		Method in [6]		Proposed Method	
	Mean	STD	Mean	STD	Mean	STD
Desk scene	0.0	0.0	0.1079	0.0093	0.4727	0.0113
Fruit scene	0.0	0.0	0.1493	0.0093	0.4122	0.0106

5. Conclusions

We provide bounding box to specify the hand area of OpenPose to get hand joint coordinates and confidence value in this experiment. Moreover, we update the best bounding box by finding the index of the maximum mean value of confidence to detect joints for next frames. As a result, we get better mean value for accuracy and reduce some processing time per frame. It is 28ms by our approach while 41ms by OpenPose. For further work, we will consider how to detect hand joints when hand is occluded by other hand or object from a single view by using some canonical transformation methods.

REFERENCES

[1] A. U. Khan and A. Borji, "Analysis of Hand Segmentation in the Wild", The IEEE Conference on CVPR, pp.4710-4719, June 2018.

[2] T. Simon et al.: "Hand Keypoint Detection in Single Images using Multiview Bootstrapping", The IEEE Conference on CVPR, pp. 4645-4653, July 2017.

[3] Z. Cao, T. Simon, S. Wei, and Y. Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields", The IEEE Conference on CVPR, pp.7291-7299, July, 2017.

[4] EgoDexterDataset: https://handtracker.mpi-inf.mpg.de/projects/OccludedHands/EgoDexter.hm

[5] HandActionDataset: https://web.bii.a-star.edu.sg/archive/machine_learning/Projects/behaviorAnalysis/handAction_ECHT/index.htm

[6] K. S. Htwe, T. Ishikawa, and H. Watanabe, "Hand Joints Detection using Initial Hand Detector on Egocentric Views", p-3-03 (poster), PCSJ/IMPS Conference, November, 2018.

[7] Darknet Library: https://github.com/pjreddie/darknet

*1 Waseda University, Graduate School of Fundamental Science and Engineering, Department of Communications and Computer Engineering

*2 Waseda University, Global Information and Telecommunication Institute

〒169–0072 Tokyo, Shinjuku, Okubo, 3-14-9, Shillman Hall-401

Phone: 03–5286–2509

E-mail: khinsabaihtwe@toki.waseda.jp