Supplementary material for "Weakly-supervised Semantic Feature Refinement Network for MMW Concealed Object Detection"

Shuiping Gou, *Senior Member, IEEE,* Xinlin Wang, Shasha Mao, *Member, IEEE,* Licheng Jiao, *Fellow, IEEE,* Zhen Liu, and Yinghai Zhao

Contents

I The keypoint detection method for structural region division

Keypoint detection is an effective method to locate key object parts, which has been used for pose estimation, face detection, etc. To verify whether the human keypoint detection method would help to improve detection performance, we use the latest pose detection model launched by Google Research, MoveNet [1] to detect keypoints of the AMMW image, and thus obtain structural regions based on keypoints. MoveNet is trained for detecting 17 keypoints of natural images, shown in Fig.1.

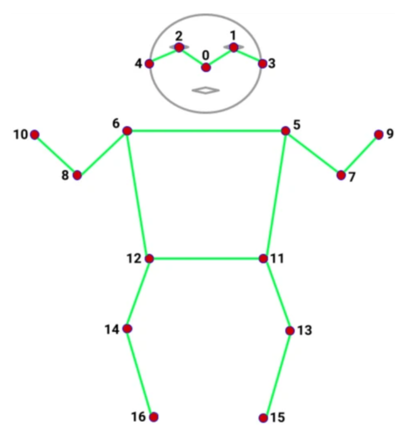

Fig. 1. The diagram of the 17 keypoints of the human body detected by Movenet.

MoveNet is directly used to detect keypoints in AMMW images since there are not keypoint annotated images for model retraining. However, owing to the difference between natural images and AMMW images, some keypoints detected in the AMMW human images are not accurate, especially keypoints in the head and arms, shown in Fig. 2. Therefore, we select the AMMW images with accurate keypoints at the shoulder, crotch, and knee for better body part detection. Specifically, we calculate the mean value of the vertical coordinates of the left and right keypoints at the shoulders, crotches and knees respectively to determine the position of each part. The human body is divided into six parts including above the shoulder, the back, the left and right thighs, and the left and right shanks, shown in Fig.3.

II The experimental analyses of keypoint detection-based body detection method

There are 10599 AMMW images obtaining accurate keypoints at the shoulder, crotch and knee. To validate the impact of body detection methods on the detection performance of concealed objects, we use these data to conduct experiments, where the 80% data are used for training, and the remaining 20% for testing.

Shuiping Gou, Xinlin Wang, Shasha Mao, and Licheng Jiao are with school of artificial intelligence, Xidian University, Xi'an 710071, Shaanxi, China (e-mail: shpgou@mail.xidian.edu.cn; xlwang_3@stu.xidian.edu.cn; ssmao@xidian.edu.cn; lchjiao@mail.xidian.edu.cn.)
Zhen Liu and Yinghai Zhao are with Beijing Huahang Radio Measurement Research Institute email: (xinllzheng@163.com; 513043512@qq.com.)

[1]https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html

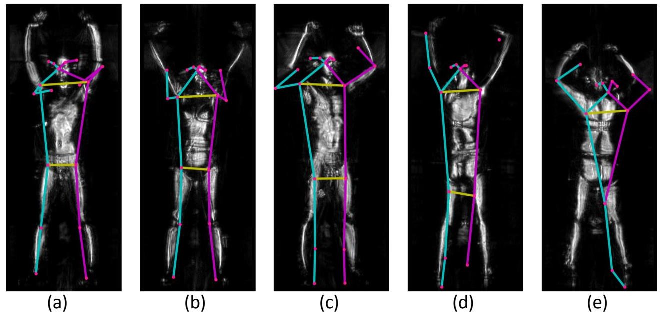

Fig. 2. Examples of MoveNet for human keypoints detection in AMMW images. (a) and (b) are successful cases and (c)-(e) are failure cases without considering the keypoint detection of the head and arms.

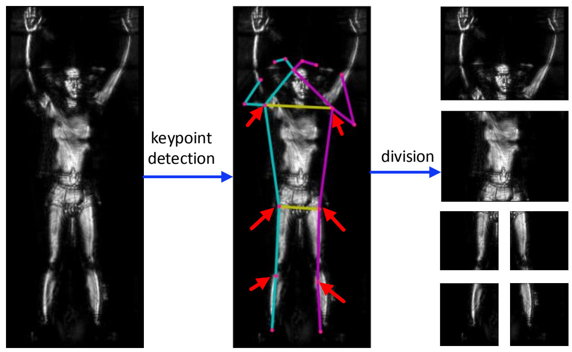

Fig. 3. The body detection based on the human keypoints detected by MoveNet. The body is divided into six parts according to the keypoints of the shoulder, crotch, and knee. The keypoints in the head and arms are generally inaccurate, which are discarded.

Table I shows the AP under different IoU thresholds using different body detection methods, where ’KP-detection’ indicates the keypoint-based body division and ’proportion-detection’ is our proportion-based body division. It is obvious that our proportion-based division method performs a little better. Moreover, Table II compares the Recall, Precision, and FAR of the two division methods under different classification scores with IoU=0.5. It illustrates that compared with the keypoint-based division method, our proportion-based division method slightly improves Recall and Precision, and decreases FAR. Based on the above analysis, the proportion-based division method is adopted to divide the body.

TABLE I
COMPARISON OF THE OVERALL DETECTION PERFORMANCE USING DIFFERENT BODY DETECTION METHODS.

Methods	AP@.1	AP@.2	AP@.3	AP@.4	AP@.5	Avg AP
KP-detection	91.26	90.80	89.08	86.70	78.80	87.33
proportion-detection	**91.82**	**91.34**	**89.46**	**86.71**	**79.11**	**87.69**

TABLE II
COMPARISON OF FALSE ALARM REDUCTION UNDER A GIVEN SCORE.

Metrics	Methods	score≥0.3	score≥0.4	score≥0.5	Avg gains
Recall	KP-detection	80.74	76.01	70.06	
	proportion-detection	81.05	76.15	70.28	↑ **0.22**
Precision	KP-detection	73.54	78.68	82.61	
	proportion-detection	74.91	79.83	83.28	↑ **1.06**
FAR	KP-detection	26.46	21.31	17.39	
	proportion-detection	25.09	20.16	16.72	↓ **0.56**