RHD latent 2.5D heatmap github

Qingfu Wan

strawberryfgalois@gmail.com

Abstract

This memo features

1. Pipeline

1 Generate image index for training (testing). (Refer to layer **GenRandIndex**, **GenSequentialIndex**. **ReadIndexFromFile**)

2 Read image. Cropped training (testing) image is in folder **training/all_hand/crop_image** (**evaluation/all_hand/crop_image**).

3 Read average bone length statistics as in Sec 3.3 of paper [1].

4 Read bounding box of cropped hand in original RGB image.

5 Read camera intrinsic parameters namely focus length and principal point offsets in pixels along the x and y axis respectively. See more details here https://arxiv.org/pdf/1808.00641.pdf.

6 Read 2D ground truth (coordinates in bounding box).

7 Read 3D ground truth in camera frame.

8 Compute the ground truth hand scale, which in paper [1] is in Sec 3.1 Equation (2) and last sentence of Sec 3.2.

9 Get the ground truth of scale-normalized root-relative (minus root depth) depth. Sec 3.1 Equation (2). Normalize this depth number to a reasonable range, say [0, 1].

10 Generate ground truth for 2D heatmap.

11 Generate ground truth for depth map (centered around each 2D keypoint). Sec 4.1 Equation (12) shows this.

12 Forward the network.

13 Slice final prediction into heatmap (21 channel) and depth map (21 channel)

14 Soft-argmax operation on 2D heatmap to get 2D pose **p**. See Fig 1 in paper.

15 Multiply latent depth map with normalized 2D heatmap. (here normalization means probability values add up to 1.0). And then *softargmax* on this multiplication to get scale-normalized depth value $\hat{\mathbf{Z}}^r$. See Fig 1 for explanation.

16 Get 2D prediction in original uncropped image by adding bounding box bbx_x1, bbx_y1, bbx_x2, bbx_y2.

17 Get scale-normalized root depth in camera coordinate. (Sec 3.2) Add this depth to form scale-normalized depth of all 21 joints in camera coordinate.

18 Scale recovery in Sec 3.3.

19 Get final 3D reconstruction result using scale-normalized depth, predicted 2D joints and computed scale.

2. Debugging

Output 2D heatmap For example, it's easy to use layer **DeepHandModelOutputHeatmapSepChannel** to output predicted 2D heatmap to see whether it's reasonable.

Overlay 2D pose on RGB image It's also easy to overlay 2D skeleton prediction on RGB image input.

3. Image Index

Not all images are used for some are corrupted.

3.1. Training

Index file is in **training/all_hand/all_train_index.txt**. 25923 samples in total. Deprecated index file in **training/all_hand/missing_all_new.txt**. (first line is total number) For *left hand only* or *right hand only*, see **missing_left_new.txt** and **missing_right_new.txt**.

3.2. Testing

Index file is in **evaluation/all_hand/all_evaluation_index.txt**. 1507 samples in total. Deprecated index file of *all* or *left hand only* or *right hand only* is similar to above.

4. Conclusion

Good luck!

References

[1] U. Iqbal, P. Molchanov, T. Breuel, J. Gall, and J. Kautz. Hand pose estimation via latent 2.5 d heatmap regression. *arXiv preprint arXiv:1804.09534*, 2018. 1