INDIANA UNIVERSITY
School of
Informatics, Computing, and Engineering

All Hands on Desk
Hand Segmentation and Activity Localization in EgoHands Dataset

Chuhua Wang, Sven Bambach, David Crandall
School of Informatics, Computing, and Engineering, Indiana University

INTRODUCTION

- Hand pose and movement contain important information of what we do and what we plan to do, which has not been extensively studied.
- This project is based on the work of Bambach et al. [1], and we improve the deep learning model with Mask R-CNN[2].
- **Two primary goals:**
  - Detecting, segmenting and distinguishing between different hand types
  - Distinguishing and locating different activities based on hand segmentation(s)

EGOHANDS DATASET

- 4000 frames as the training/validation data
- 800 frames as the testing data

Hand Type Detection & Segmentation

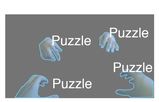
- **Hand Types**
  - Each hand is labelled as 1 of the 4 hand types: 'Myleft', 'Myright', 'Yourleft' and 'Yourright'

Ground Truth for hand segmentation, where color indicates different hand types

Activity Localization

- **Union**
  - All hands are labelled as 1 of the 4 activities: Cards, Chess, Jenga and Puzzle

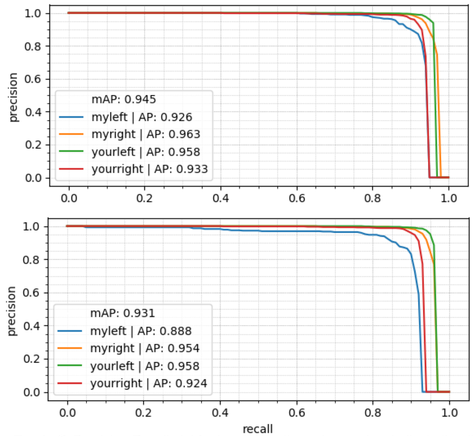
- **Box**
  - Each hand is labelled as 1 of the 4 activities

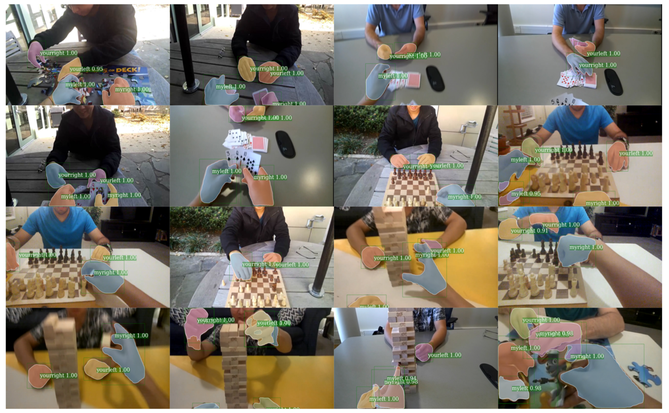


- **Box with Hand Types**
  - Each hand is labelled as 1 of the 4 hand types and activities
  - 4 activities x 4 hand types = 16 classes

HAND TYPE DETECTION & SEGMENTATION

- Model is initialized with ResNet-101 pre-trained model and fine-tuned for EgoHands dataset.
- Detection outperforms the Bambach's et al. work [1]. (**Ours:** 0.945 mAP; **Bambach et al. :** 0.807 mAP)
- Segmentation outperforms recent work by Khan et al. [3]. (**Ours:** 0.931 mAP; **Khan et al. :** 0.879 mAP)

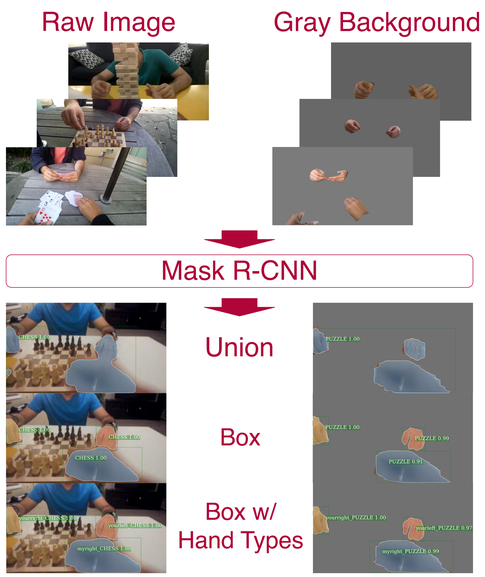


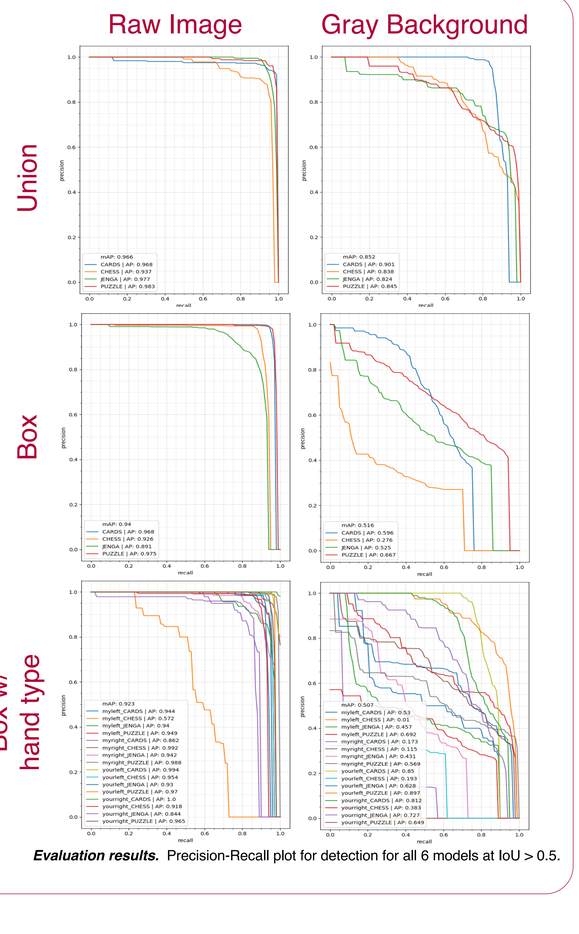


Top Left: Precision-Recall plot for hand detection, where colors indicate the different hand types. ***Right:*** Some random examples
Bottom Left: Precision-Recall plot for hand segmentation. IoU threshold was set to 0.5 for both plots.

ACTIVITY LOCALIZATION

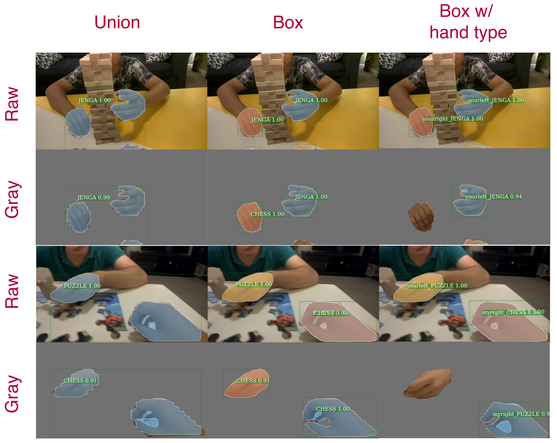
- Created training data with background converted into mean gray
- 3 approaches to train the network: 'Union', 'Box' and 'Box with hand types'
- (Raw image + Gray background image) x 3 approaches = 6 models





Evaluation results. Precision-Recall plot for detection for all 6 models at IoU > 0.5.

ACTIVITY LOCALIZATION



CONCLUSION AND FUTURE WORK

- Our results outperform previous works in both hand detection and segmentation.
- For activity localization, the model learns the spatial information between objects when multiple segmentations are given as 1 instance.
- The model performs well on detection and segmentation tasks without background. The decreasing precision is mostly due to misclassification on different activities. If background is given, the context information helps to classify objects.
- **Future work:**
  - Train new models with limited context information given
  - Add new pose annotations to existing EgoHands dataset

ACKNOWLEDGMENTS

Thanks to Sven Bambach for meeting with me, and thanks him and David Crandall for their expertise. Further, thanks to David Crandall for providing server to train my model.

KEY REFERENCES

[1] S. Bambach, S. Lee, D. J. Crandall, and C. Yu. Lending a hand: Detecting hands and recognizing activities in complex egocentric interactions. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.
[2] K. He, G. Gkioxari, P. Dolla´r, and R. Girshick. Mask R-CNN. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2017.
[3] A. U. Khan and A. Borji. Analysis of hand segmentation in the wild. *arXiv preprint arXiv:1803.03317*, 2018.