Contextual Attention for Hand Detection in the Wild
Supreeth Narasimhaswamy*, Zhengwei Wei*, Yang Wang, Justin Zhang, and Minh Hoai
Department of Computer Science, Stony Brook University

Contributions
- A state-of-the-art Hand Detector
- Two large scale annotated hand datasets
- An attention method for object detection

Goal
To improve the hand detection performance of OpenPose and Mask-RCNN on unconstrained images with large variation in appearance of hands (close-up shots, occlusions, motion blur)

OpenPose:

Ours:

Mask-RCNN:

Ours:

Quantitative Comparison:

Datasets
TV-Hand:
- Image frames extracted from video clips of ActionThread dataset, and contains 8.5K images with 9.5K hands.
- COCO-Hand:
- Images from a subset of Microsoft COCO dataset, and has around 26K images with 45K hands.

Quantitative Results
- Benefits of Context:
- Benefits of Data:

Orientation
We extend Mask-RCNN to include an additional network branch to predict hand orientation
Orientation branch shares weights with other branches, so it does not incur significant computational expenses

Orientation Loss:

Total Loss:

Failure Cases

Acknowledgments. This work was partially supported by the VinAI research and NSF IIS-1763981. The authors would also like to thank Tomas Simon for his suggestions about the COCO dataset and Rakshit Gautam for his contribution to the data annotation process.

Code, Data