Towards Human Gesture Recognition AI Models

Macarena Peralta[1], Purva Tendulkar[2], and Carl Vondrick[2]

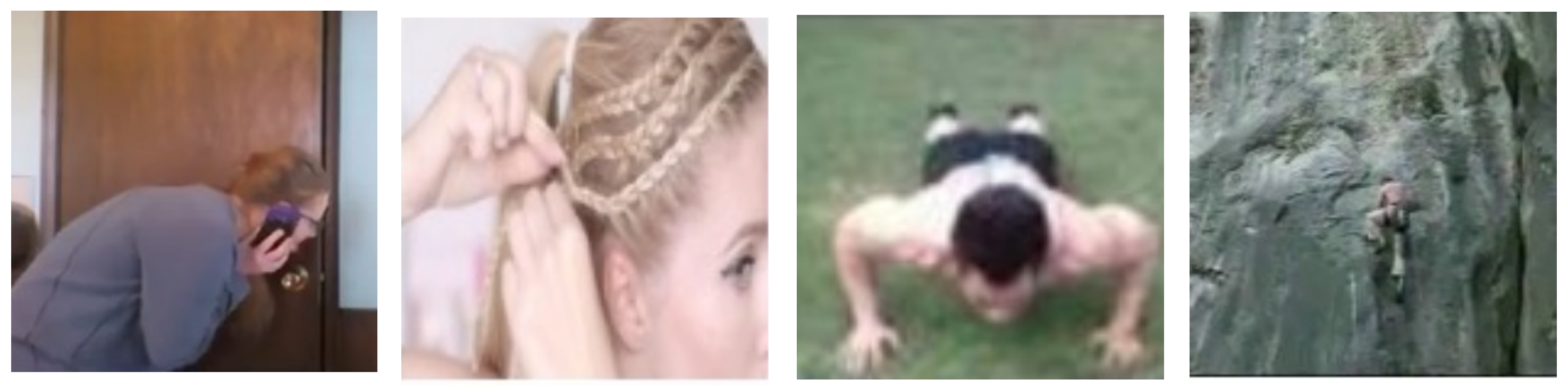
1 Department of Computer Science, University of Michigan, Ann Arbor, MI
2 Department of Computer Science, Columbia University, New York, NY

Motivation

- Understanding the complex ways in which humans use their bodies to express themselves is important for meaningful communication.
- The focus of this project is to create a supervised classification model to recognize videos of humans performing gestures, and categorize them into one of ten categories (e.g., shrugging, scratching head, shushing, etc.).

Related Datasets

- Human action recognition is already an active area of research.
- Existing datasets include Charades [1], Kinetics [2], UCF101 [3], and HMDB51 [4].
- However, none of these have been curated for the purpose of studying communicative gestures.



Holding a Phone — Braiding Hair — Push Ups — Climb

Figure 1: Examples from Charades, Kinetics, UCF101, and HMDB51 respectively.

Data Collection

- We begin by identifying ten distinct categories via crowdsourcing.
- We query Tenor API [5] to obtain 1500 GIFs per category.
- We use Detectron2 [6] object detection software to remove the GIFs that do not contain exactly one person or have an object obstructing the view of the person.

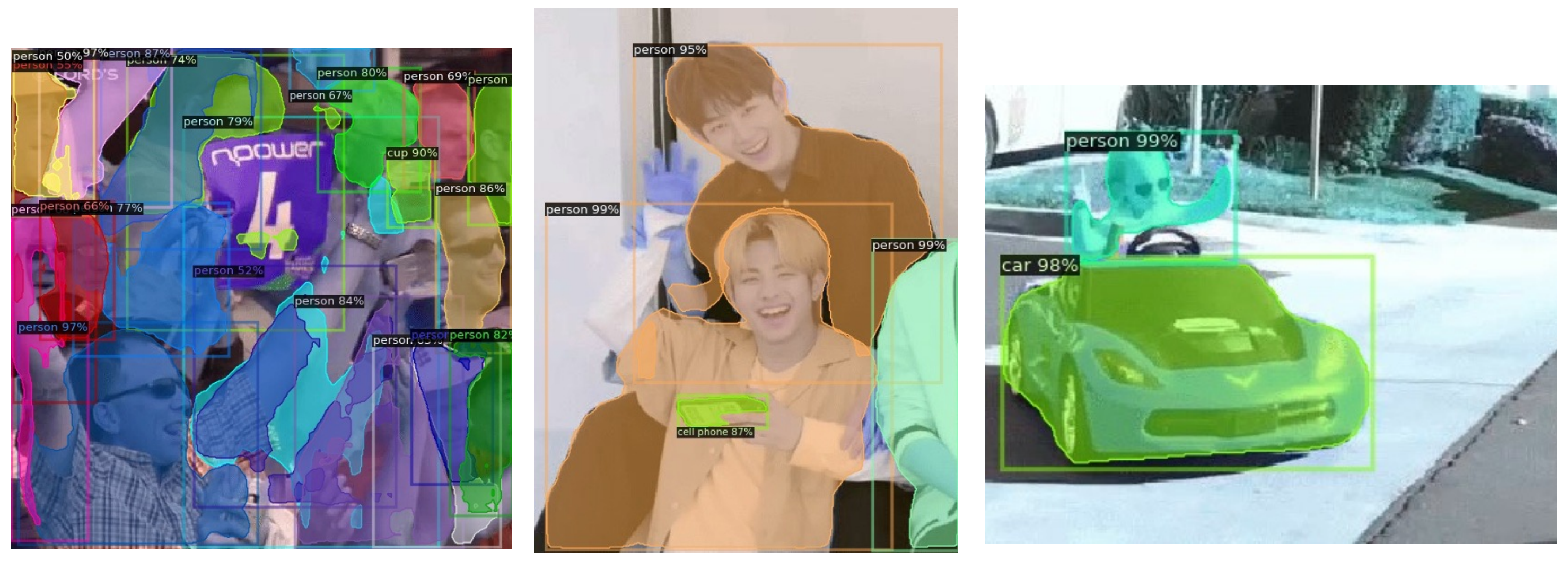


Figure 2: Examples of frames from GIFs that were cleaned out due to not meeting the criteria listed above.

Data Processing

- After cleaning the data, we apply the AlphaPose [7] model (pre-trained on Halpe dataset [8]) which detects 136 keypoints for the body, hands, and face on the GIFs.
- We prepare vector visualizations by removing the backgrounds and comparing the raw video and the AlphaPose keypoint videos.
- Our curated data visualizations can be accessed via the QR code at the bottom of this poster.

Figure 3: Shrugging.

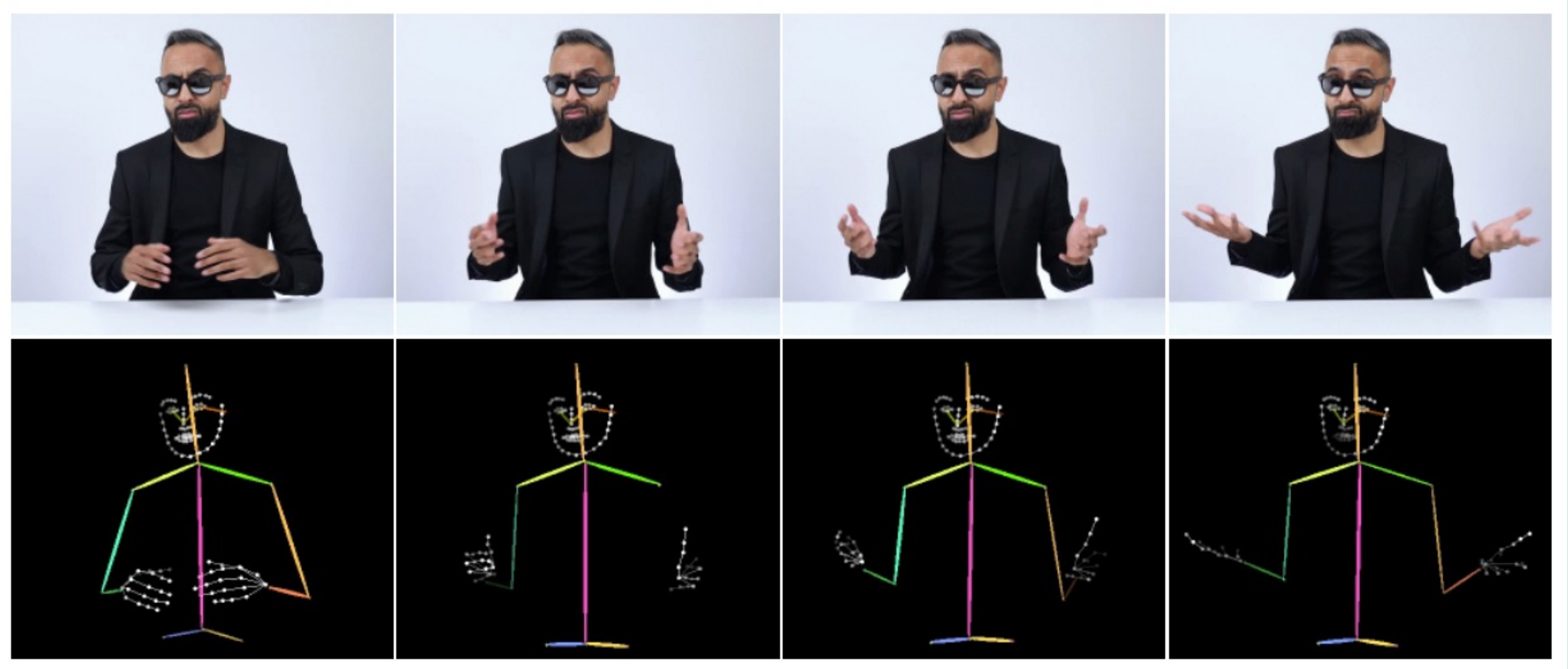


Figure 4: Mind Blown.

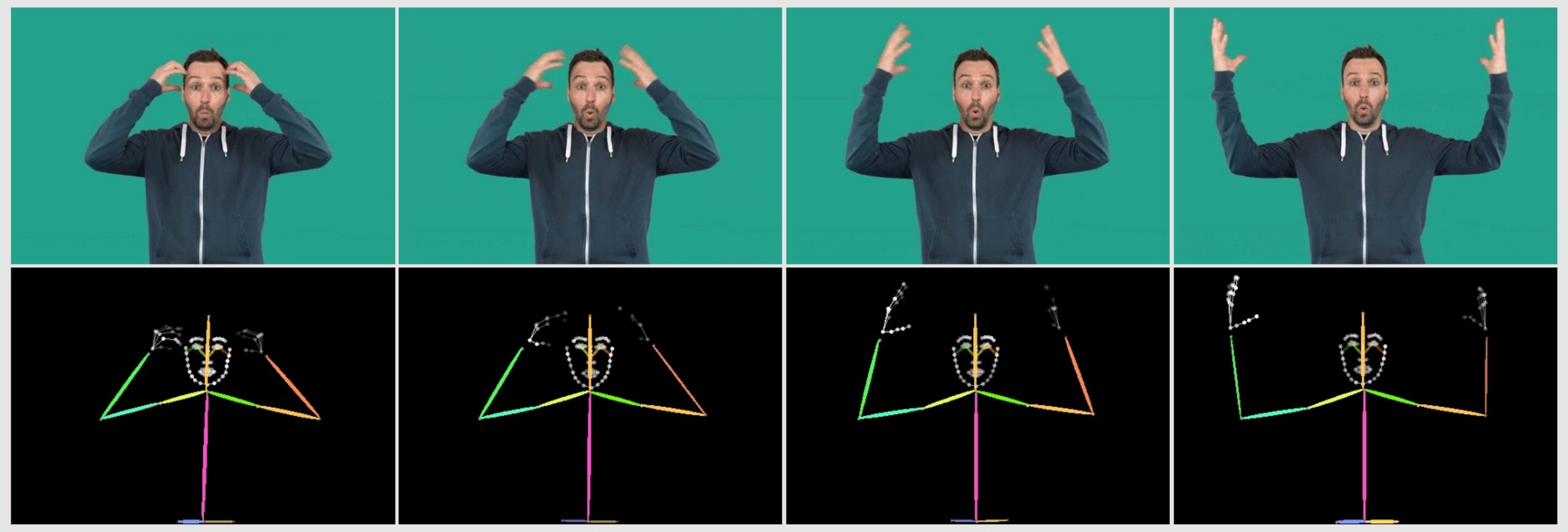
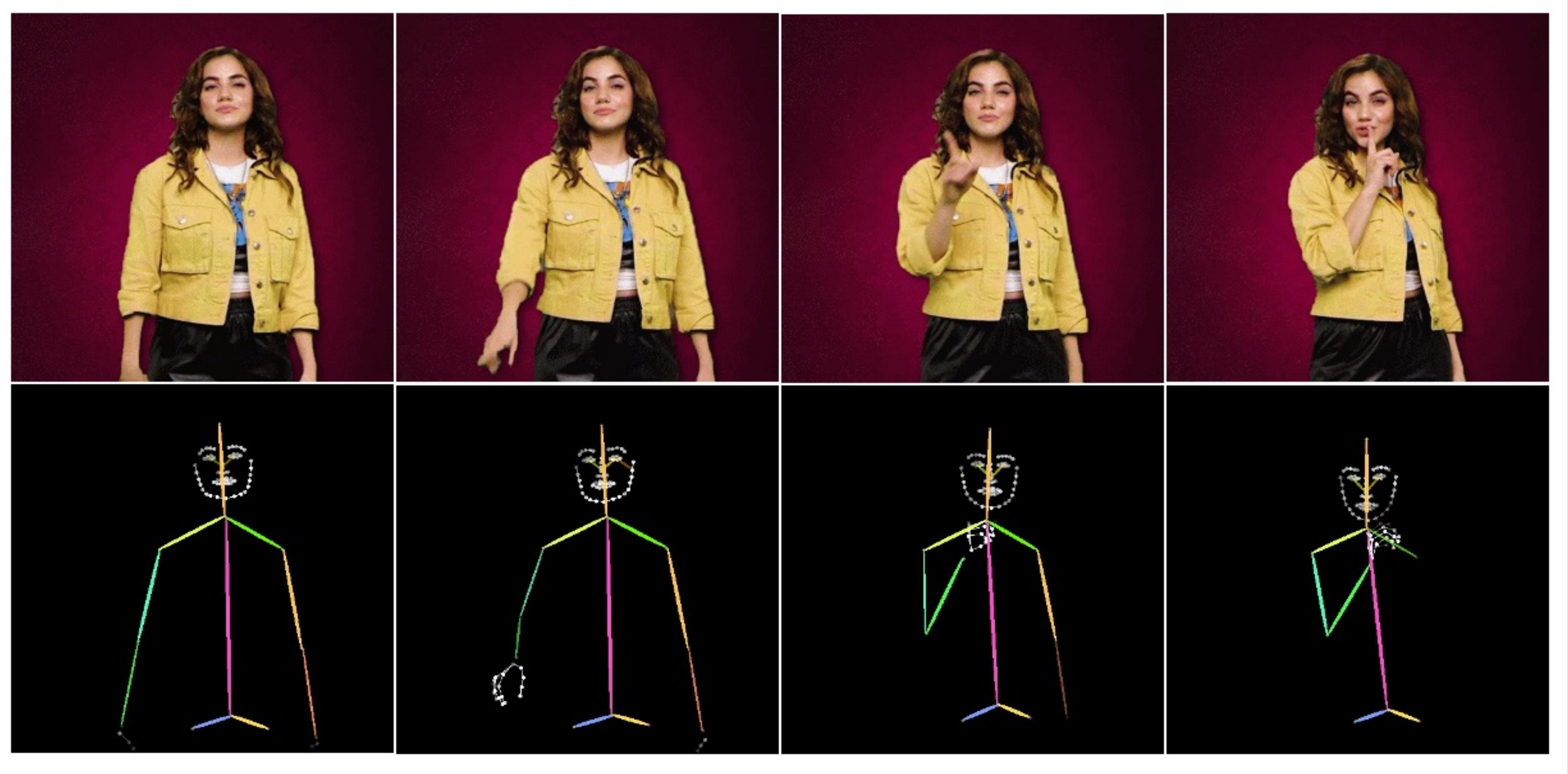


Figure 5: Shushing.



Figures 3-5: In-order frames of selected human gesture examples together with their keypoint visualizations.

Data Statistics

- We have gathered and curated a dataset for human gesture classification containing ten gestures, which is divided into an 80/10/10 data split for training, validation, and testing, respectively.

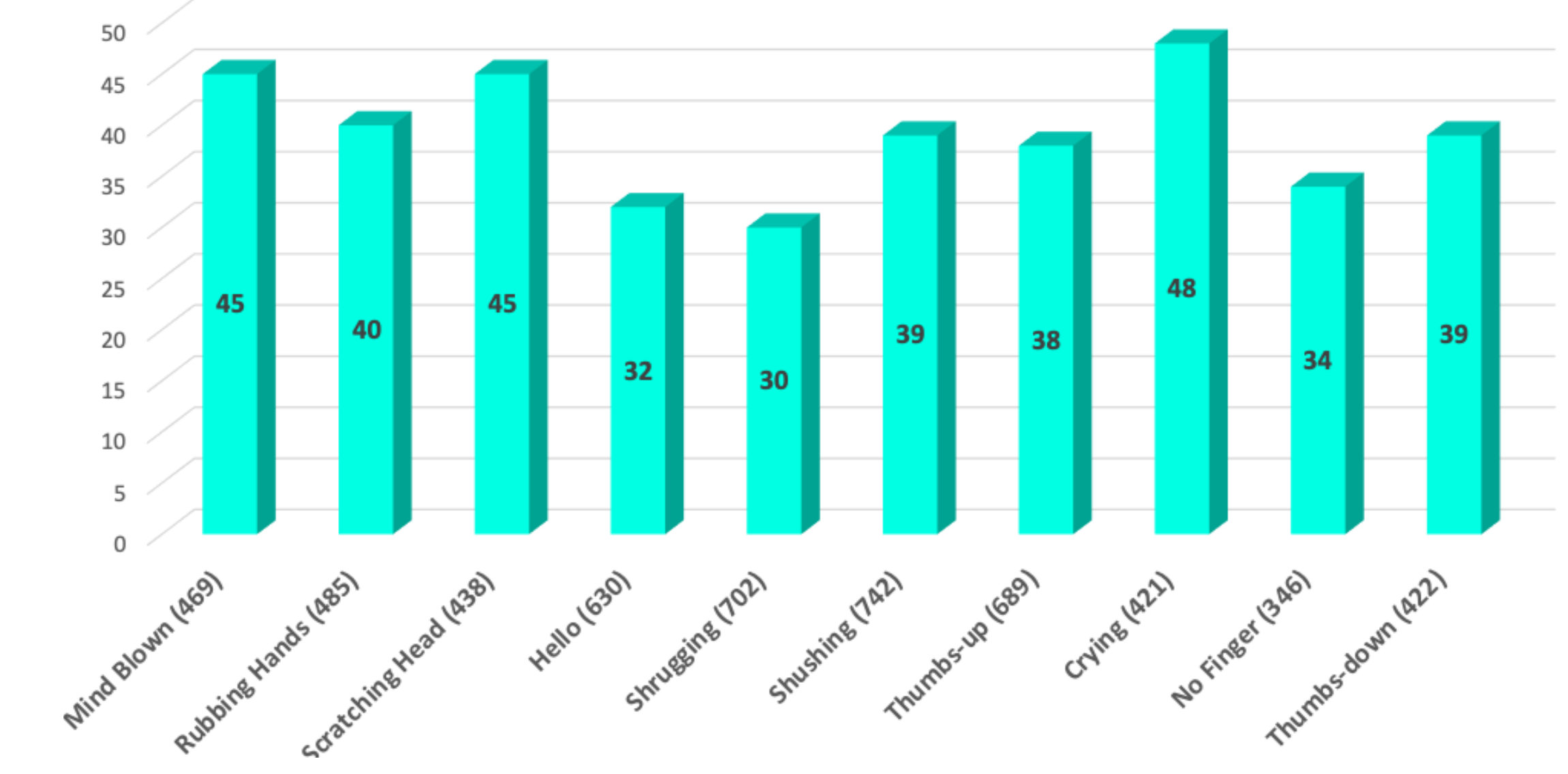


Figure 6: Average number of frames for GIFs in each of the ten gesture categories.

Summary and Next Steps

- With this curated dataset we can train, validate, and test the accuracy of supervised machine learning (ML) models for these ten gestures.
- We have setup the data processing pipeline in a way that can be easily updated to contain more gesture categories in the future, thereby creating a robust framework for human gesture classification.
- Besides human gesture classification, we can also design a conditional generative model that can generate keypoint sequences given a gesture prompt.

References

[1] Gunnar A. Sigurdsson, Santosh Divvala, Ali Farhadi, & Abhinav Gupta (2017). Asynchronous Temporal Fields for Action Recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[2] Carreira, J., Noland, E., Hillier, C., & Zisserman, A. (2019). A Short Note on the Kinetics-700 Human Action Dataset (Version 1). arXiv. https://doi.org/10.48550/ARXIV.1907.06987

[3] Khurram Soomro, Amir Roshan Zamir & Mubarak Shah. “UCF101 - Action Recognition Data Set.” *UCF Center for Research in Computer Vision*, Nov. 2012, https://www.crcv.ucf.edu/data/UCF101.php.

[4] H. Kuehne, H. Jhuang, E. Garrote, T. Poggio and T. Serre, "HMDB: A large video database for human motion recognition," 2011 International Conference on Computer Vision, 2011, pp. 2556-2563, doi: 10.1109/ICCV.2011.6126543.

[5] “GIF API - Better, Faster & Free: Get Your Gifs with Tenor.” Tenor. https://tenor.com/gifapi.

[6] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, & Ross Girshick. (2019). Detectron2. https://github.com/facebookresearch/detectron2

[7] Hao-Shu Fang, Shuqin Xie, Yu-Wing Tai, Cewu Lu, Jiefeng Li, Haoyi Zhu, Yuliang Xiu, and Chao Xu. (2017). Alphapose.
https://github.com/MVIG-SJTU/AlphaPose

[8] Mvig-Sjtu. “Alphapose/MODEL_ZOO.MD at Master · MVIG-SJTU/Alphapose.” *GitHub*, https://github.com/MVIG-SJTU/AlphaPose/blob/master/docs/MODEL_ZOO.md.

Acknowledgments

Thank you to the Columbia Summer Undergraduate Research Experience (SURE) program, Amazon, graduate mentor Purva Tendulkar, and PI Carl Vondrick for the opportunity to participate in this research project.