EE 541 – Computational Introduction to Deep Learning

Project Proposal

Yuqi Chen, Mengqi Du

November 22, 2022

Project Title: CV for American Sign Language

Topic summary

American Sign Language is a natural language that serves as a predominant sign language for the Deaf people in the United States. If ASL can be recognized and interpreted precisely through a machine learning application, we can bring much convenience to the deaf and hard of hearing community. Before training the model, we will firstly preprocess the dataset with a appropriate way. Then we will test with many different architectures and parameters to get good performance and finetune existing models that perform classifications on the provided gesture images. A successful model will recognize and interpret every gesture with a high level accuracy.

Figure 1: ASL finger alphabet

Dataset description

https://www.kaggle.com/c/sartorius-cell-instance-segmentation

The data set is a collection of images of alphabets from the American Sign Language, separated in 29 folders, each of which corresponds to one single class. 26 classes are for the the letters A-Z and 3 classes for SPACE, DELETE and NOTHING. The training data set contains 87,000 images which are 200x200 pixels. The test data set contains a mere 29 images, to encourage the use of real-world test images.

Estimated Compute Needs

These may include online compute resources with GPU acceleration (e.g., Kaggle or Google CoLab) or personal workstations (e.g. GPU, CPU, RAM).

Likely Outcome and Expected Results

A successful model will recognize and interpret every single gesture from a deaf or hard of hearing person with a high level accuracy. Due to the limit test cases of each class, chances are that our model can not be generalized to the real world gesture recognition.

Primary References and Codebase

We propose to build on the approach used in

- J. Yi, P. Wu, D. J. Hoeppner and D. Metaxas, "Pixel-wise neural cell instance segmentation," 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, pp. 373-377, doi: 10.1109/ISBI.2018.8363596.
- Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. InInternational Conference on Medical image computing and computer-assisted intervention 2015 Oct 5 (pp. 234-241). Springer, Cham.
- Danielle Bragg1, Oscar Koller2, Mary Bellard2, Larwan Berke3, Patrick Boudreault4, Annelies Braffort5, Naomi Caselli6, Matt Huenerfauth3, Hernisa Kacorri7, Tessa Verhoef8, Christian Vogler4, Meredith Ringel Morris1, Sign Language Recognition, Generation, and Translation: An Interdisciplinary Perspective.
- Madeleine S. Durkee, Rebecca Abraham, Junting Ai, Jordan D. Fuhrman, Marcus R. Clark, Maryellen L. Giger, "Comparing Mask R-CNN and U-Net architectures for robust automatic segmentation of immune cells in immunofluorescence images of Lupus Nephritis biopsies," Proc. SPIE 11647, Imaging, Manipulation, and Analysis of Biomolecules, Cells, and Tissues XIX, 116470X (5 March 2021).
- Github codebases: Attentive Neural Cell Instance Segmentation, marshuang80/cell-segmentation