Classifying Sign Languages and Fingerspellings with Convolutional Neural Networks

Brandon Vu, Richard Chen, Nyle Wong; *Mentor: Haochen Shi*

Department of Computer Science

Introduction

- Sign language originates from the deaf community and is a language where handshapes and gestures are the main modes of communication
- American Sign Language (ASL) is the most studied and has many existing datasets and models developed for it
- There are many low-resource sign languages including Japanese Sign Language (JSL), Irish Sign Language (ISL), and Arabic Sign Language (ArASL)
- Our goal is to build a model that addresses this gap, taking the resources available to develop a robust classifier

Data

Hand Detection: We use the EgoHands dataset and David Lee's ASL dataset

EgoHands Dataset	**15,053 images**	**720 x 1,280 pixels RGB**
David Lee's ASL Dataset	**1,728 images**	**384 x 384 pixels RGB**

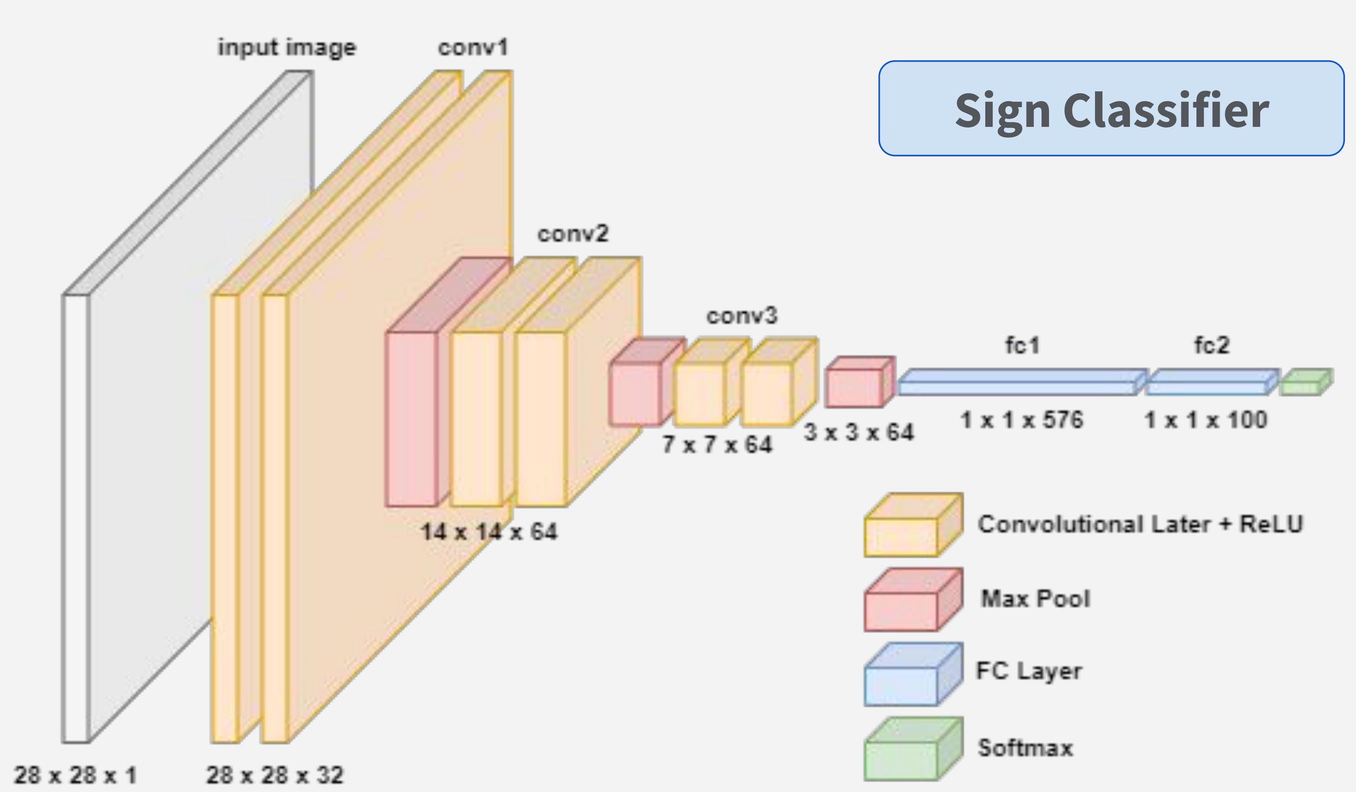
Sign Classification: We use fingerspelling data from four sign languages: American, Japanese, Irish, and Arabic

Preprocessing

Crop hand ⇒ Resize to 28 x 28 ⇒ Grayscale

8,055 images per language	**32,220 images total**	**28 x 28 pixels B&W**

Model Architecture



Methods

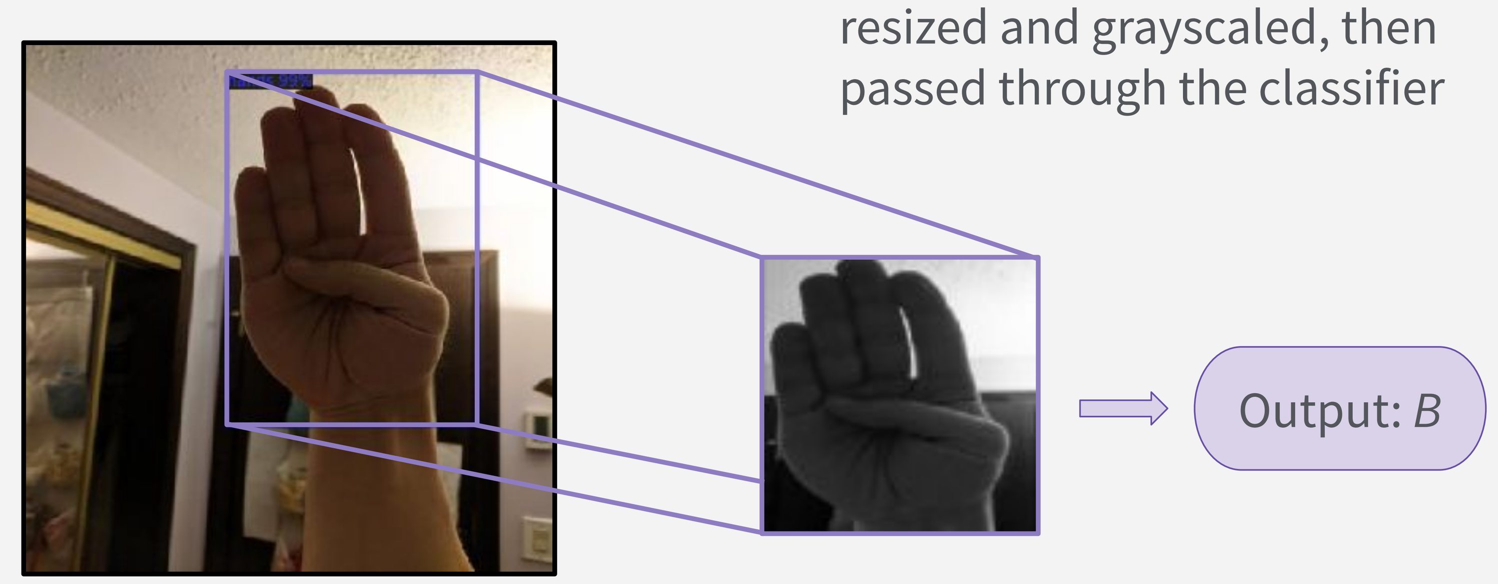
- We separate our approach into two steps: hand-detection and sign classification. Each step has a separately trained model

Hand Detection

- Model trained on data in order to localize hand within image
- Used **Faster-RCNN-R50-FPN-3x** pretrained model
- Fine-tuned on EgoHands and David Lee's ASL Dataset
- Detect bounding box for hand

Sign Classification

- Model is trained on the combined data of all sign languages
- Each sign + language combination is given a unique id
- To predict an image, it's first cropped to the bounding box, resized and grayscaled, then passed through the classifier



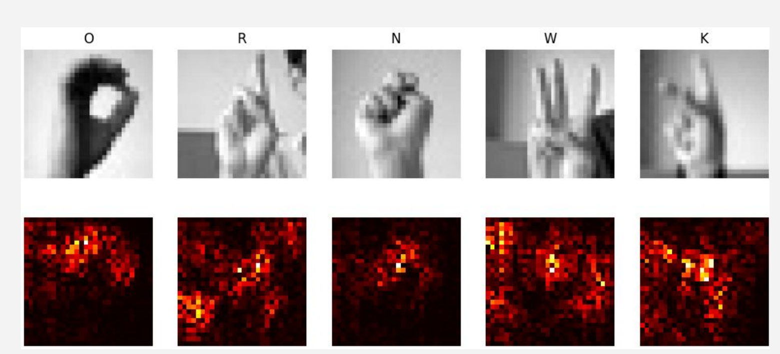
Hand Detection bounding box example

Experiments

Hand Detection: Learning rate: 1e-5, Batch Size: 8, Adam Optimizer, Cross Entropy Loss, 4 epochs

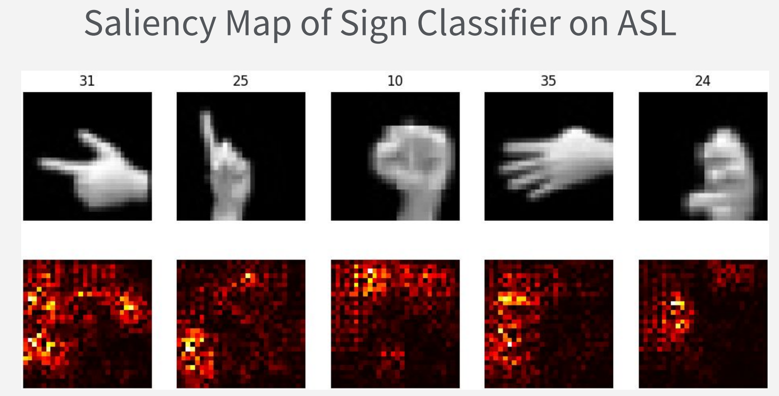
Sign Classification: Learning rate: 1e-4, Batch Size: 10, Adam Optimizer, Cross Entropy Loss, 2 epochs + 2 epochs with 5e-5 learning rate

Saliency Maps



Saliency Map of Sign Classifier on ASL

- ASL saliency map shows model is learning the right features
- Highlighting shape of the hand and distinct patterns for each sign
- Hot spots on characteristic aspects of signs



Saliency Map of Sign Classifier on images with dark backgrounds (JSL)

- Inverse-like saliency map on dark backgrounds
- Highlighting the dark areas, outlining the inverse shape
- Indicates that the background is still a big factor in predictions

Results

Sign Classifier:

Validation Accuracy	Test Accuracy
90.22%	90.25%

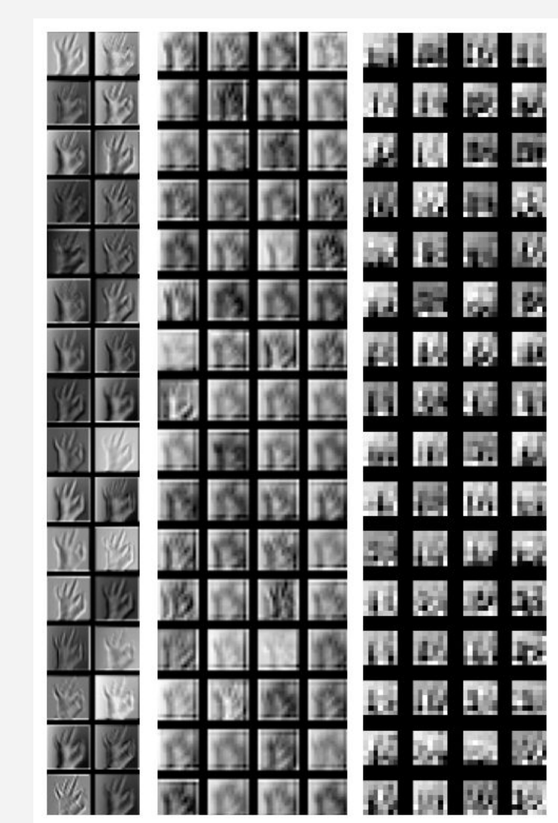
Hand Detector: The metric used for the hand detection model is Average Precision (AP), which combines precision and recall. This takes in a intersection of union (IOU) value, which is in parentheses below

	MAP(0.5:0.95)	AP(0.5)	AP(0.75)
Validation Set:	66.447%	96.071%	80.632%
Test Set:	66.508%	94.661%	82.385%

Time Trials:

	Training Time	Inference Time
Hand Detector	36min	0.036s
Sign Classifier	5min	0.027s

Convolutional Layer Weights Visualization



- As another analysis tool, we display the weights of each convolutional layer, from left to right respectively
- The visualized weights show what we expect to see: different features of the hand and sign are being highlighted in each image
- Detail is lost in later layers
- Model is making general abstractions from the initial image in order to make classifications

Weights for each convolutional layer: 32, 64, and 64 respectively

Conclusion

- With the two-model approach, we were able to develop a method of translating fingerspelling for potentially low-resource sign languages
- Both models were able to achieve high-performing results
- Limitations of our approach is due to data: despite the hand detection step, it's difficult to avoid overfitting due to the artificial nature of the datasets used
- Next steps would involve data augmentation and engineering to make the models more resistant to noise and backgrounds