

Signing Off on



Using Deep Learning to recognize hand signs from the
ASL alphabet

Objective: Create a model that can classify hand gestures as ASL Alphabet hand signs.



Data

Kaggle Dataset:

- 77,518 images
- 26 image classes
 - A-Z, plus space, minus J
- ~3,000 per image class

Tools utilized

- Pandas/NumPy
- Tensorflow/Keras
- OpenCV
- MediaPipe
- Matplotlib

Methodology – Preprocessing

kaggle

64x64 px



Methodology – CNN



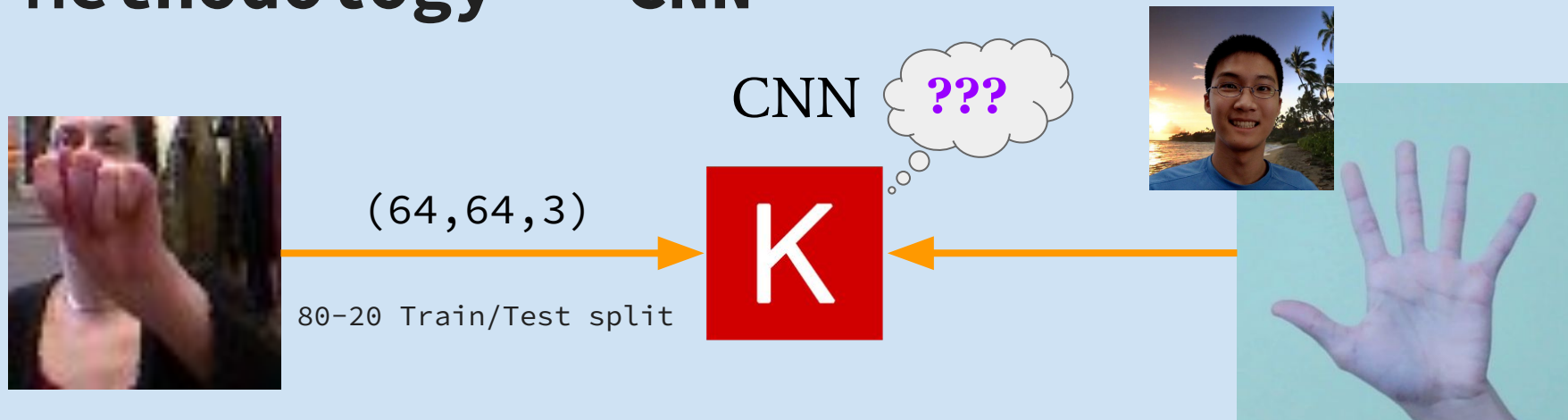
$(64, 64, 3)$

80-20 Train/Test split

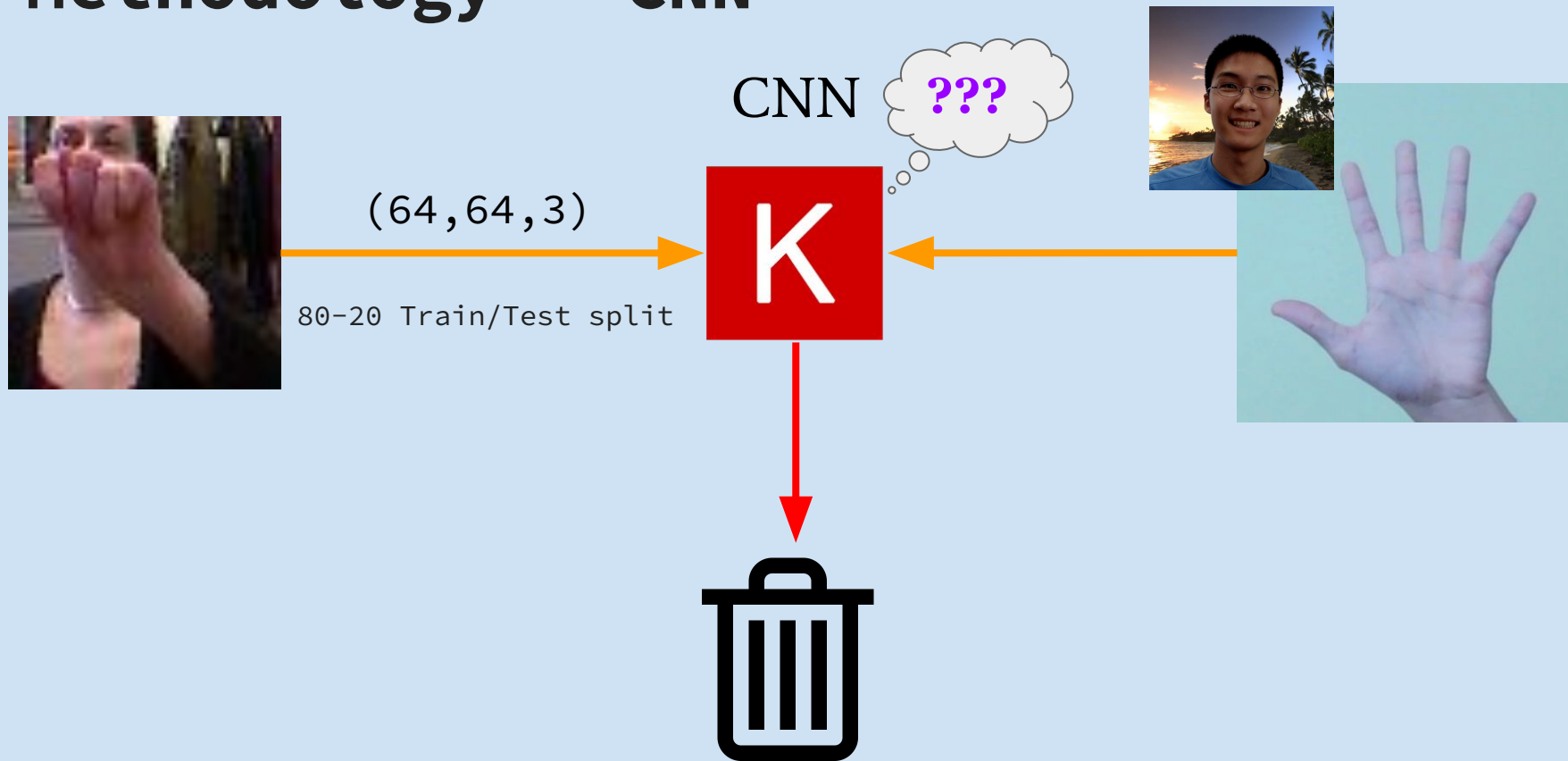
CNN



Methodology – CNN

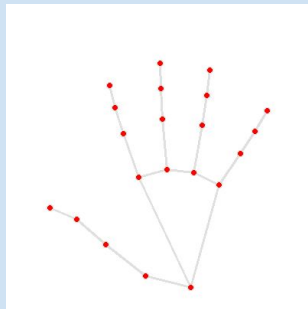


Methodology - CNN



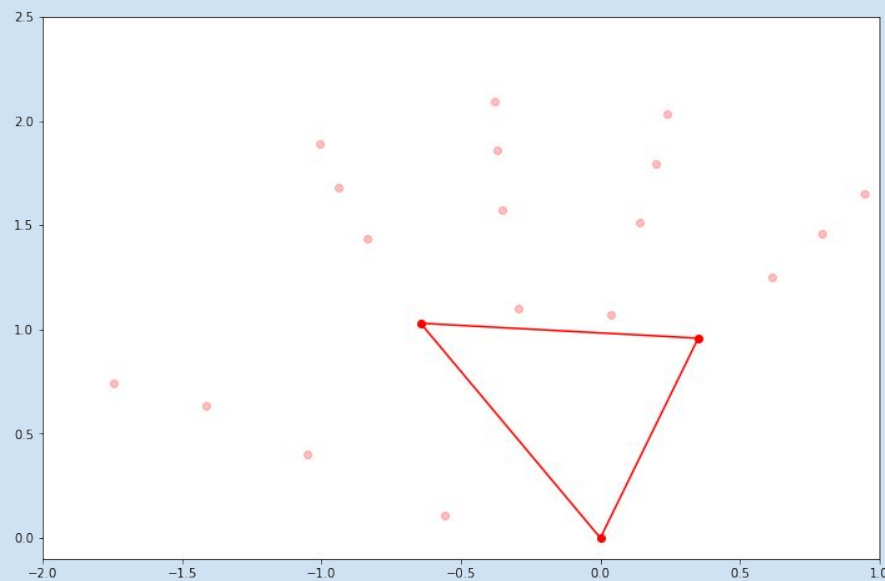
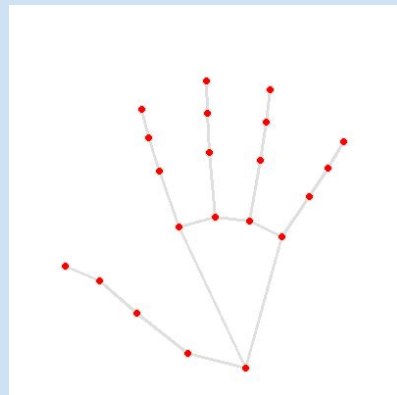
MediaPipe

Google's MediaPipe library can highlight key points in a photo of a hand, allowing for simpler networks and faster modeling

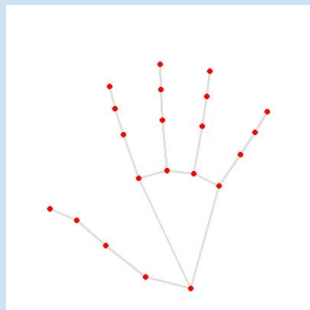


Regularization

In order to prevent overfitting, create a representation of the hand, scaled to the size of the palm.



Methodology - MLP



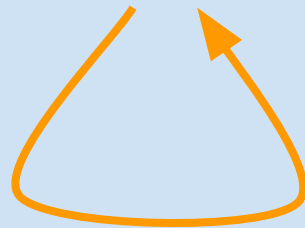
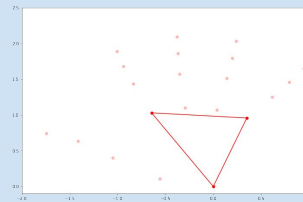
(21, 2)



(42, 1)



MLP



Regularization

 MediaPipe

A highly simplified view of the model

Model

Input Layer: 42

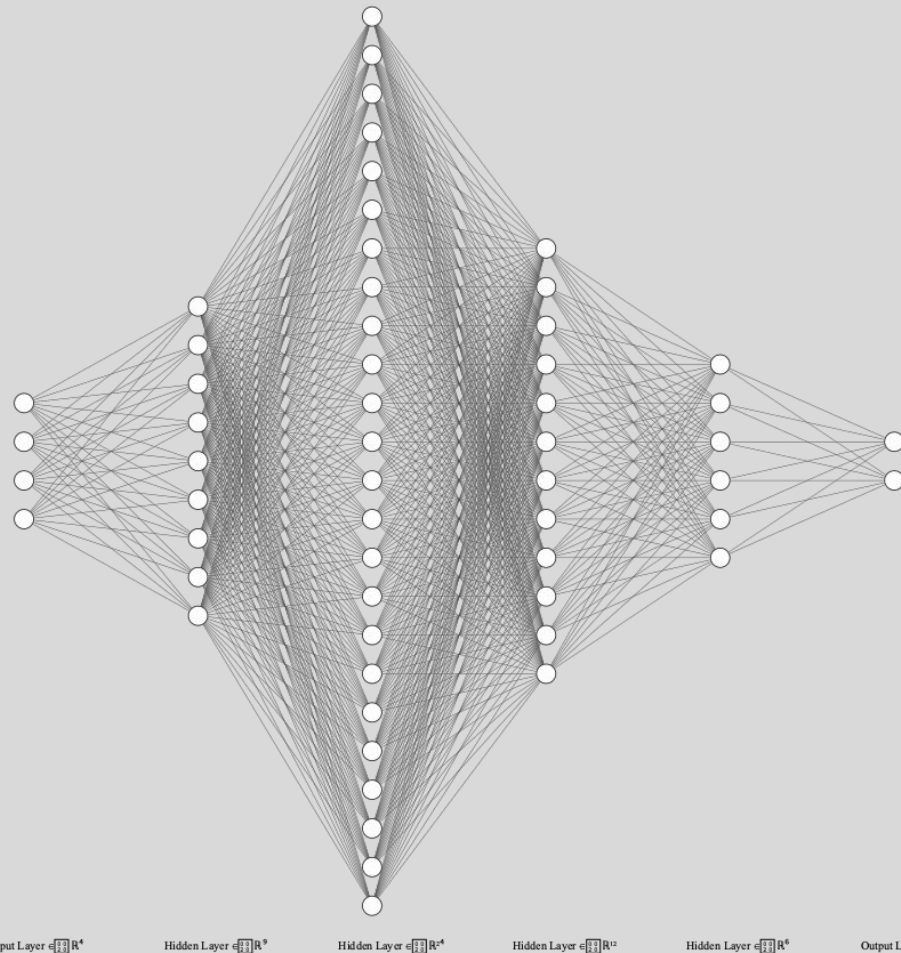
Hidden Layers: 96, 256, 128, 64 (relu)

Output: 26 categories (softmax)

Loss metric: Categorical Crossentropy

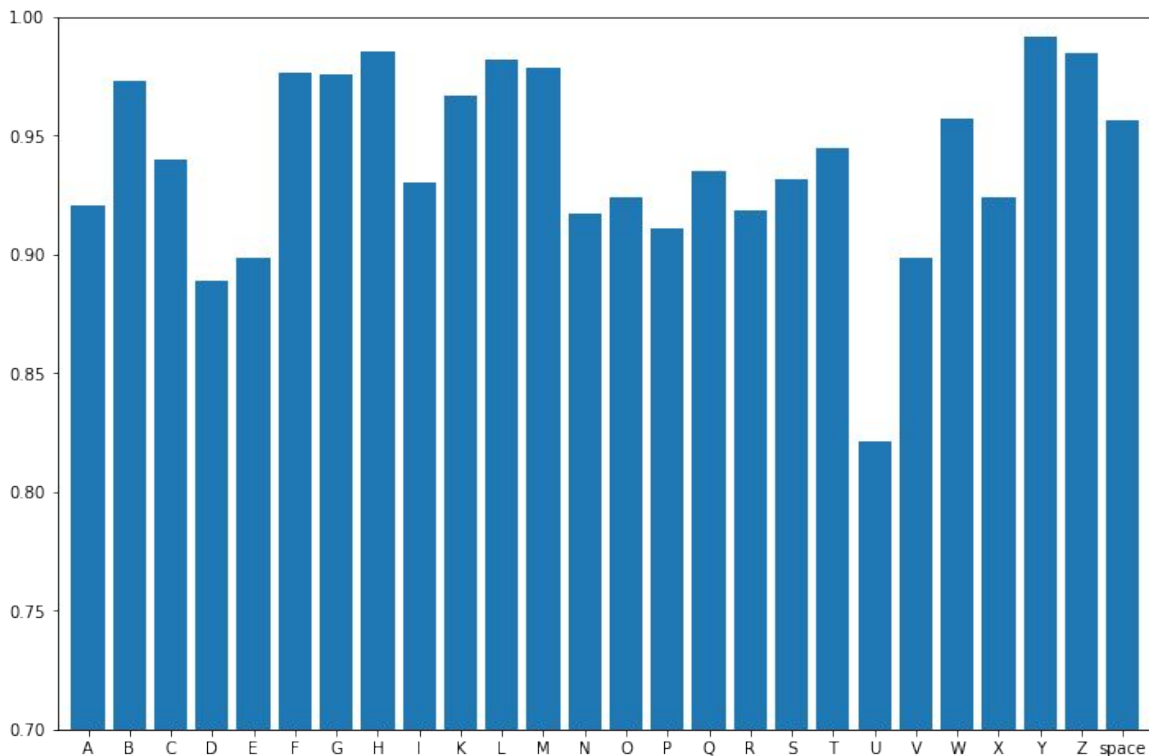
Optimizer: ADAM

Metric: Accuracy (0.942)



Hand Sign Accuracy

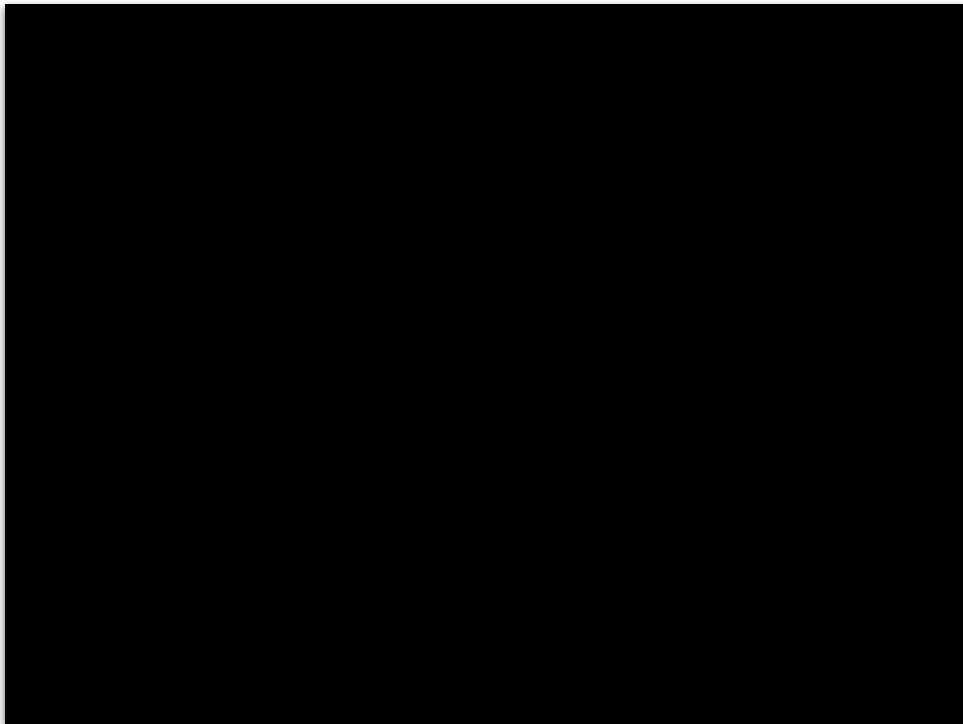
U, V, D, E, Q are among the lowest scoring signs, accuracy-wise.



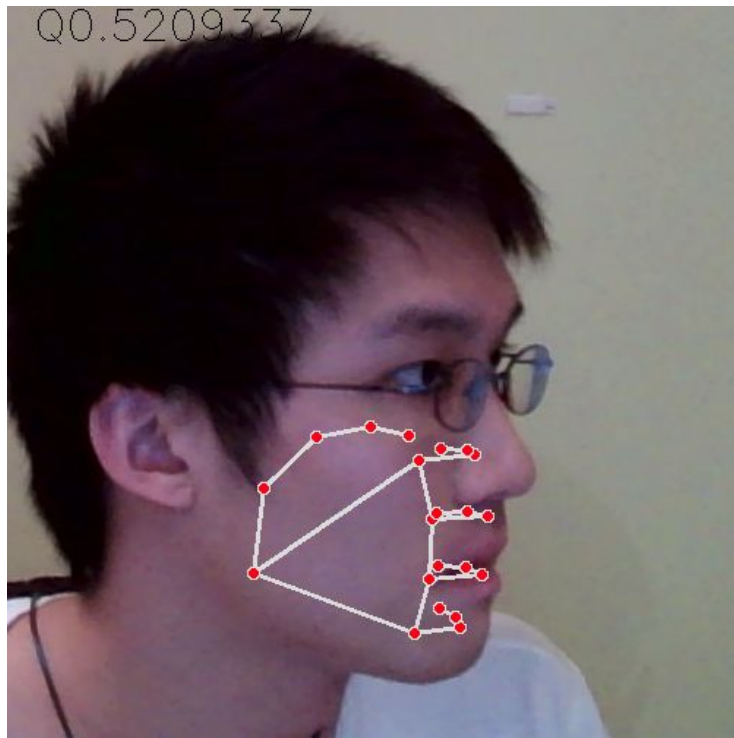
Useful Applications

This model can be used to
translate hand signs in a video.

Right: Matthew making a valiant
attempt to sign a complete
sentence with ASL



Less-useful applications



Data scientists testing their models on literally anything:



('F', 0.9997948)



('X', 0.49064082)

Future Work

- Get more data; lack of varied backgrounds limits CNN effectiveness
- Gather videos of more complex gestures, and use RNN to interpret motion
- Code a feature that builds a sentence from hand signs, using NLP data for auto-correct.
- Use MediaPipe's facial recognition to stop it from recognizing my face as a hand.

Thank you!