
Gesture Recognition with Transfer Learning

— By Thomas Matson and Rui Yuan Shum —

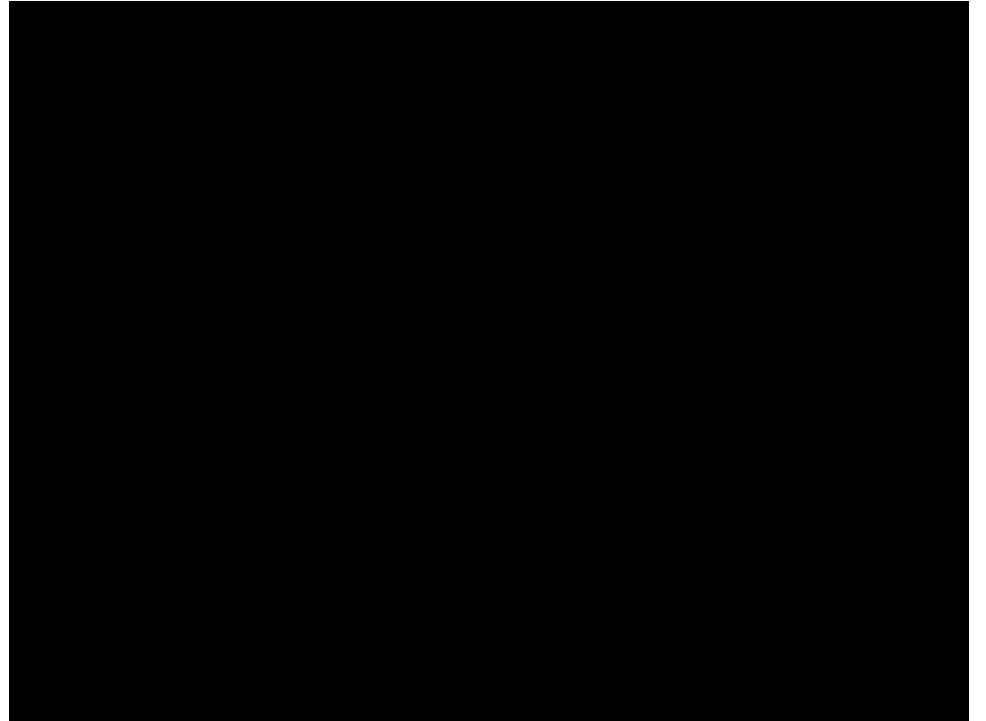
Problem Statement

Cozmo is supposed to be a smart robot at home, but to conventionally train it with custom gestures, you typically would need:

- Knowledge about implementing machine learning
- Lots of labelled data
- A way to train cozmo (that is easy and isn't terribly slow)

Inspiration

Inspired by Google's Teachable Machine Web App, which allows for training of a gesture recognizer in minutes. We want to implement this on Cozmo, so we can give him commands via the camera.



Approach

Transfer learning using pretrained VGG 16.

1. Load VGG16, removing last layer
2. Use that model to extract features from training data
3. With those features, train a smaller, custom model to classify the images

Giving Cozmo Data

Uses polling and Cozmo's camera to get data for each label

Takes about 600 images for each of 4 classes in about 4 minutes



Demo

Results

Finished training in less than a minute

It then collects 5 images every 2.5 seconds, outputs the mode prediction, and then performs an action based on that prediction.

It achieved ~87% accuracy on unseen data.

What we learned

Simplifying task by providing good user interface

Different ways to perform transfer learning

Recognising limitations of current models, and from there developing a better network architecture:

- Preventing overfitting (Decreasing number of hidden units, adding dropout)
- Tuning hyperparameters to escape local minima and converge smoothly (adding momentum and a decaying learning rate)

What's next?

Data augmentation and model tuning for better accuracy.

Experimenting with a NULL class, so Cozmo knows when to do nothing.

Sign-to-speech:

- Sign language (alphabet) interpretation
- Transfer learning to customize for users

Model Architecture

For our simple custom classifier, that takes in the 4096 outputs from VGG:

Input layer

Dropout layer, with ~30% dropout

10 node fully connected layer (ReLU activation)

Output layer, with 4 class softmax activation.

VGG Architecture

