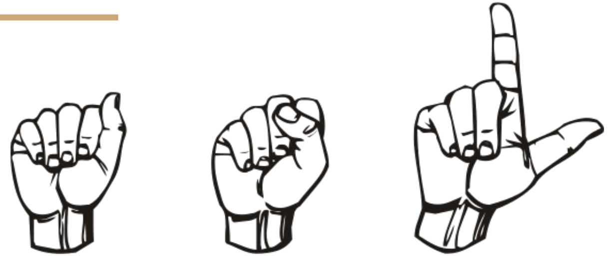


ASL Complete Alphabet Recognition

Group #6
Dima Al Dulaijan
Faisal Mushayt
Myles Cork

BOSTON UNIVERSITY

Related Work

Paper 1: Real-time American Sign Language Recognition with Convolutional Neural Networks

- Interpret alphabet signs excluding j and z
- Model
  - GoogLeNet
  - Loss: Log Loss
  - Optimizer: Not Mentioned, although most likely SGD
  - Pre-trained, but reinitialized last 2-3 layers with Xavier initialization
  - Re-trained on 2 datasets

Paper 1: Real-time American Sign Language Recognition with Convolutional Neural Networks

- Results
  - 72% accuracy on a-y excluding j and z
- Conclusion
  - Produced a somewhat robust model for a-e, but would need more data for whole alphabet
- Future Work
  - Other models
  - More image preprocessing
  - Language model

Paper 2: Using Deep Convolutional Networks for Gesture Recognition in American Sign Language

- Interpret alphabet + digits
- Model
  - Custom deep CNN
  - Loss: categorical cross entropy
  - Optimizer: Not mentioned
  - 1 Dataset
- Tested classifying two sets, a-z and 0-9

Paper 2: Using Deep Convolutional Networks for Gesture Recognition in American Sign Language

- Results
  - 82.5% accuracy on found dataset
  - 67% accuracy on made dataset
- Conclusions
  - Need more training data
- Future Work
  - Predict text based on sequential context

Paper 3: American Sign Language Character Recognition using Convolution Neural Network

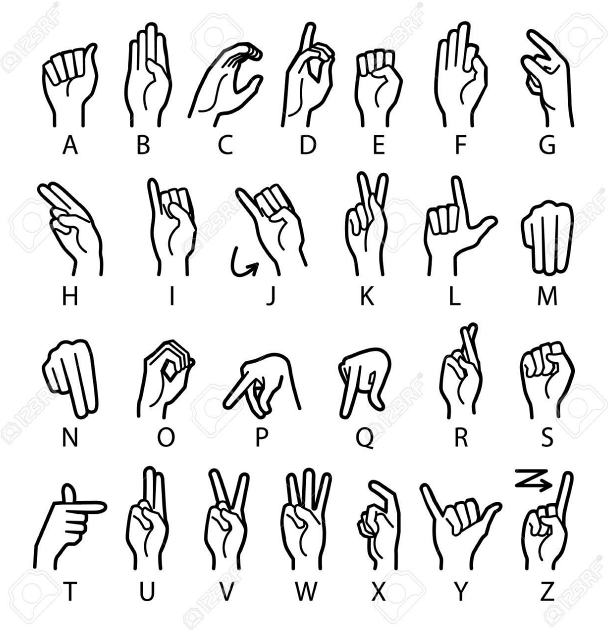
- Model
  - VGG16
  - Loss: Didn’t mention
  - Optimizer: SGD with “momentum”
  - Pretrained on imagenet, retrained last layer on ASL dataset.
  - 1 Dataset

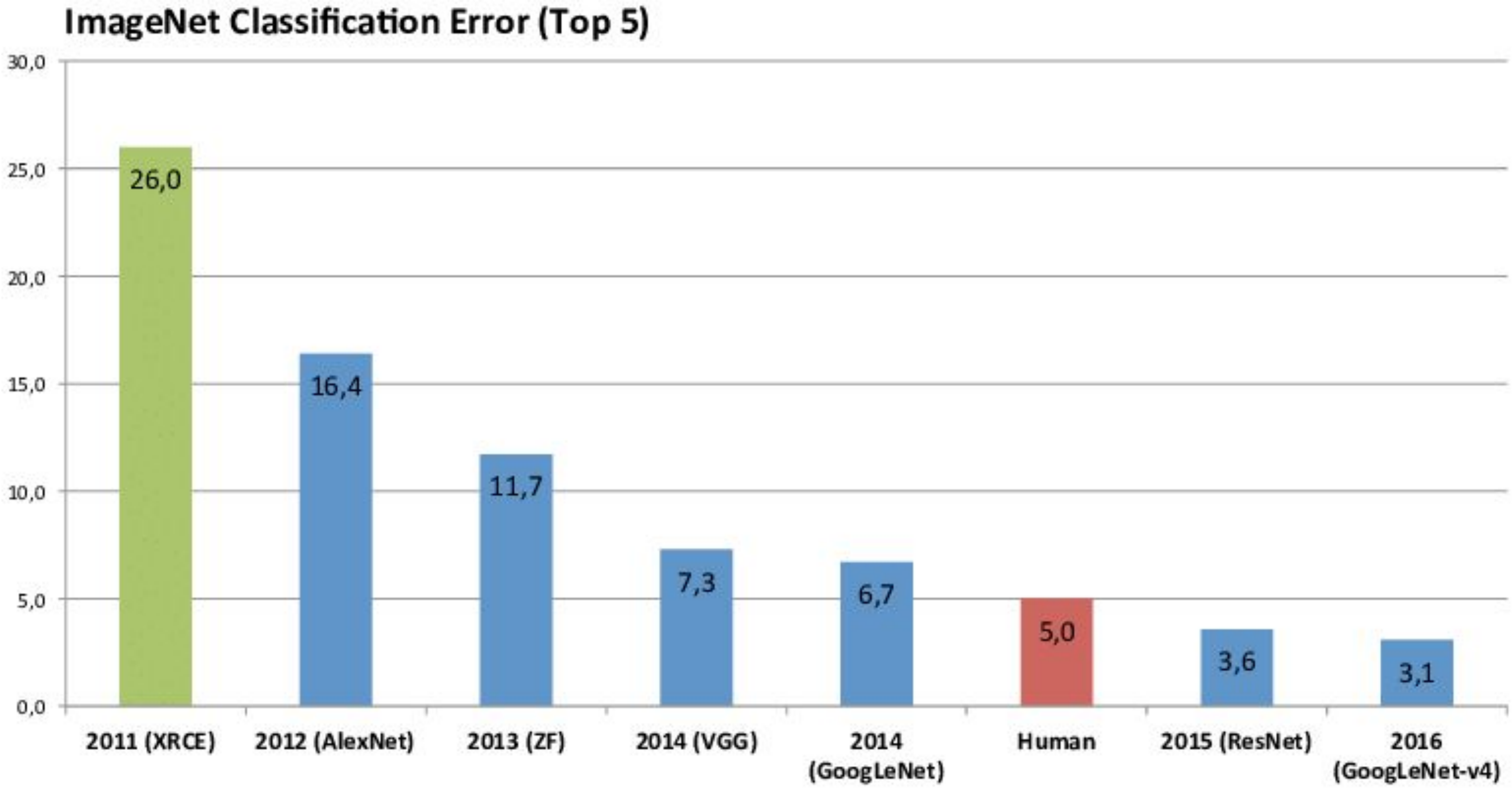
Paper 3: American Sign Language Character Recognition using Convolution Neural Network

- Results
  - 95.54% accuracy
- Conclusion
  - CNN’s are remarkably good at identifying sign language.
- Model overfitted

Our Goals

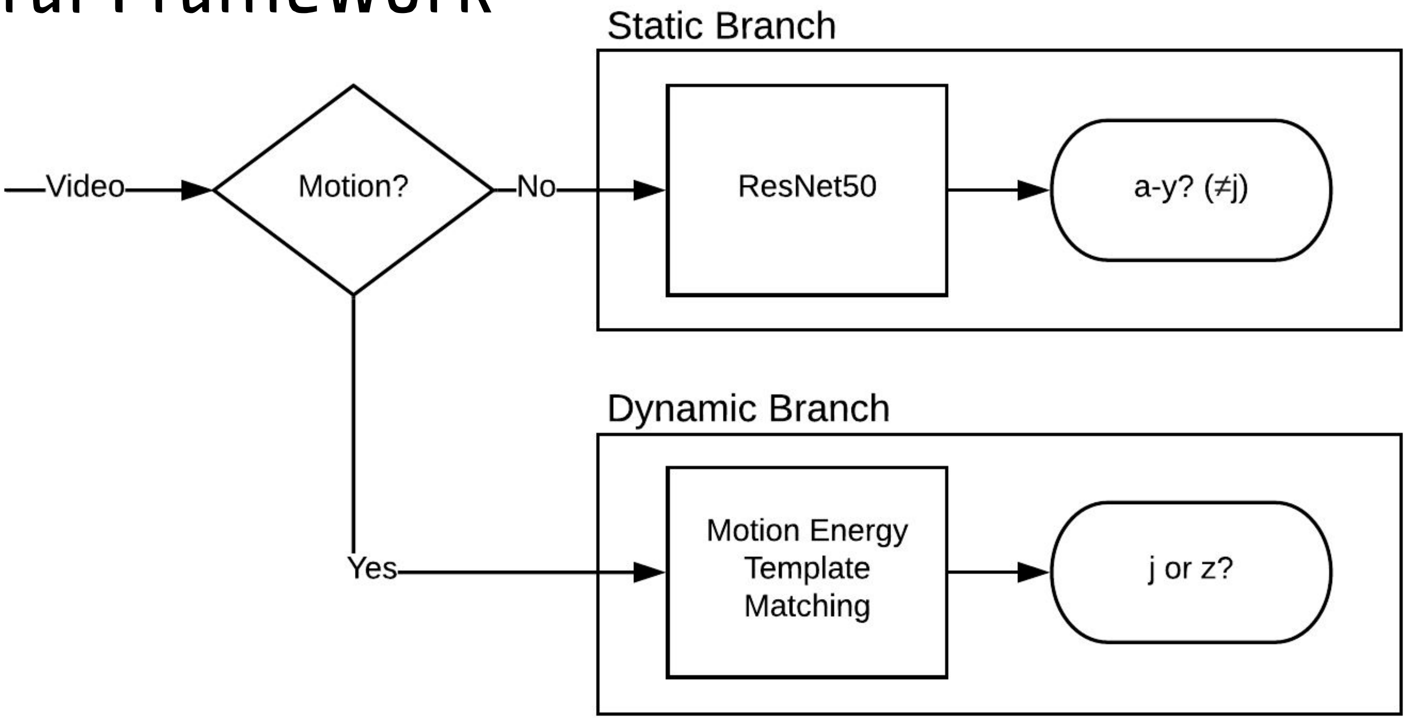
- Interpret both static and dynamic alphabet signs
- Use CNN to determine letters excluding j and z
- Use another classifier to determine if j or z





ImageNet Classification Error (Top 5)
30,0
25,0
20,0
15,0
10,0
5,0
0,0
26,0
16,4
11,7
7,3
6,7
5,0
3,6
3,1
2011 (XRCE)
2012 (AlexNet)
2013 (ZF)
2014 (VGG)
2014 (GoogLeNet)
Human
2015 (ResNet)
2016 (GoogLeNet-v4)

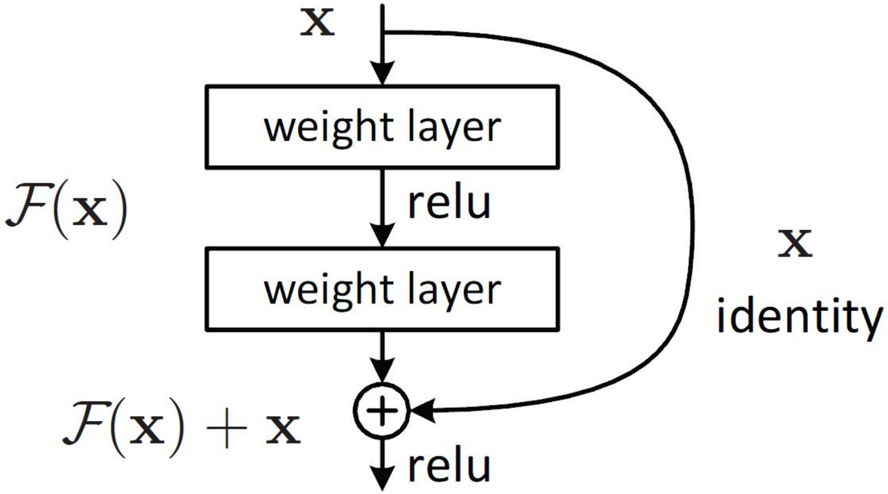
General Framework

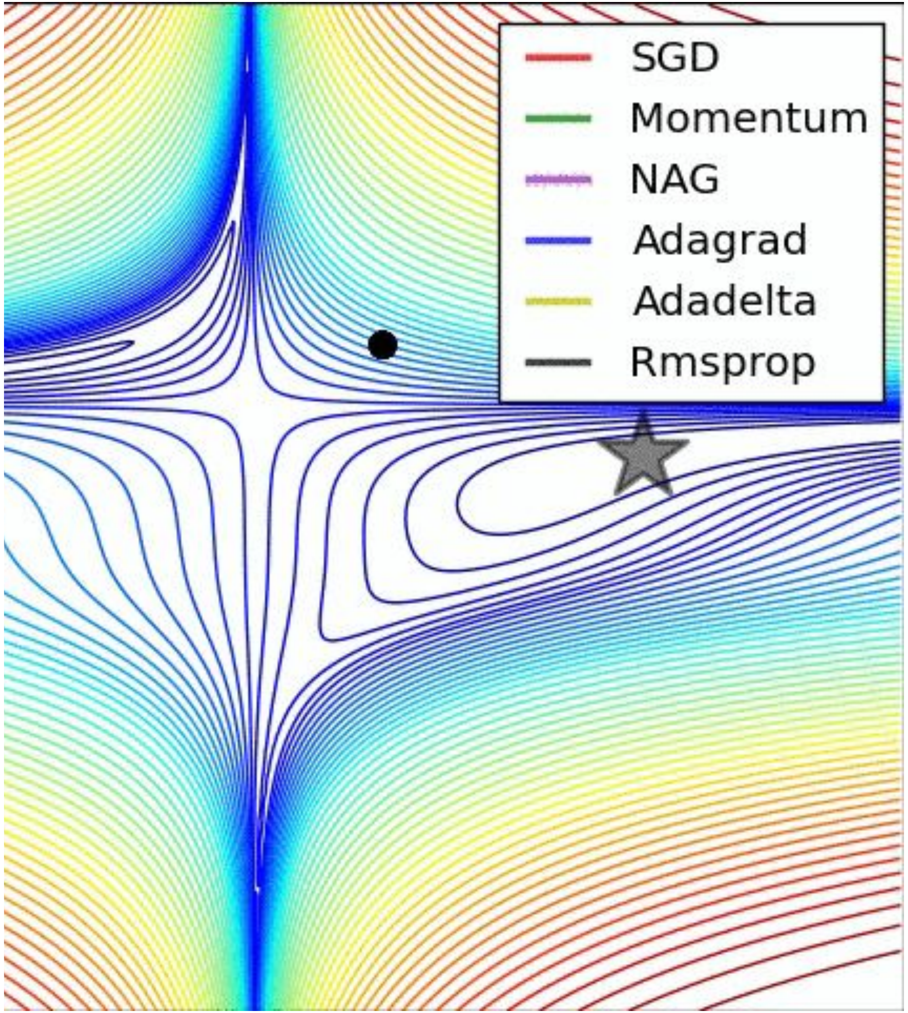


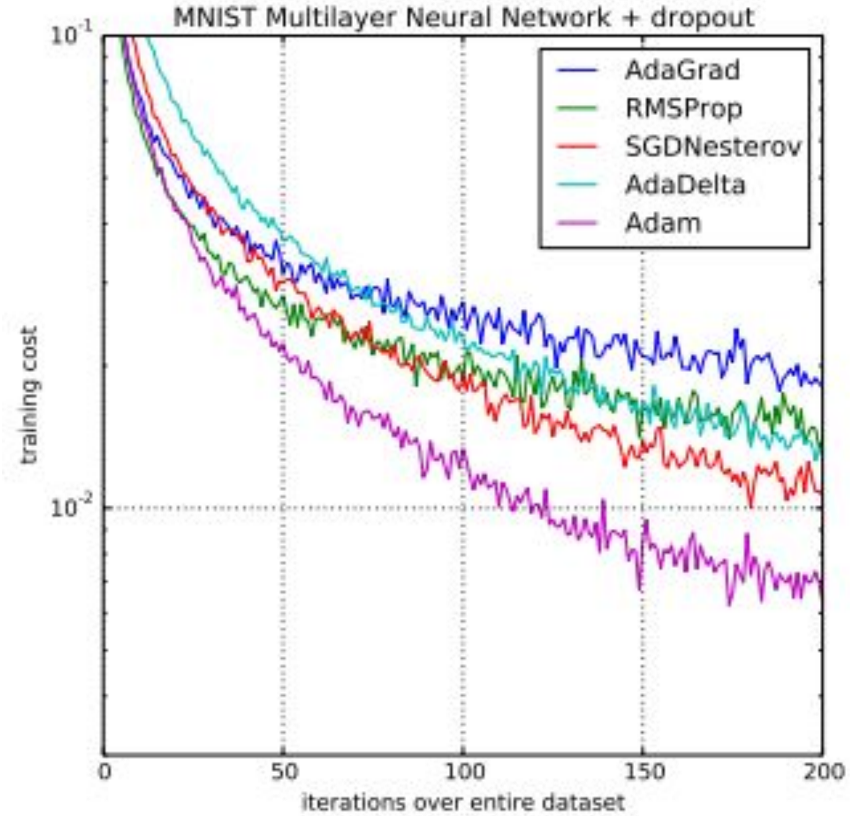
Static

- ResNet50
  - Predicts based on a single frame
  - Training
    - Loss: categorical_crossentropy
    - Optimizer: Adam
    - 50 Epochs
    - Modified regularization

Keras




SGD
Momentum
NAG
Adagrad
Adadelta
Rmsprop

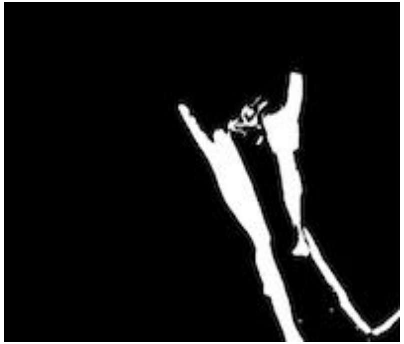

MNIST Multilayer Neural Network + dropout
10^{-1}
10^{-2}
training cost
0
50
100
150
200
iterations over entire dataset
AdaGrad
RMSProp
SGDNesterov
AdaDelta
Adam

Dynamic

Classifying Dynamic Gestures

- J and Z
- Motion Energy History
- OpenCV Template Matching

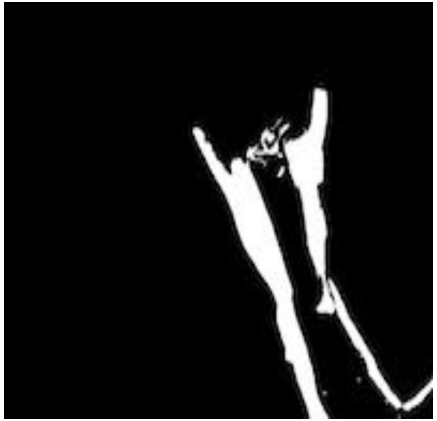
OpenCV



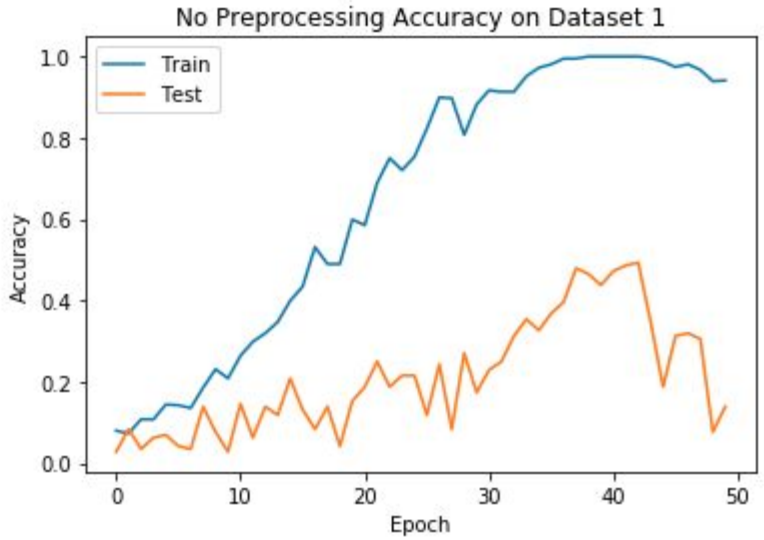
Results and Conclusions

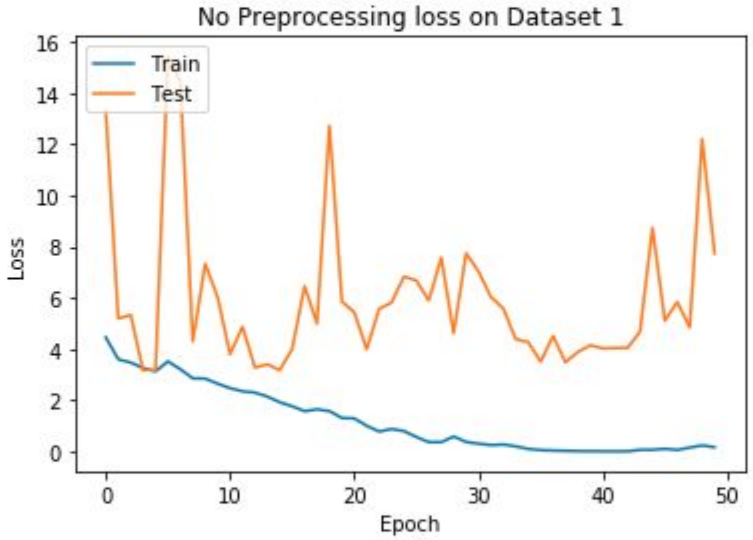
J or Z

- Accurate overall from experimentation
- Hard to quantitatively measure accuracy due to issues
    - Forearm matching
    - Environmental conditions

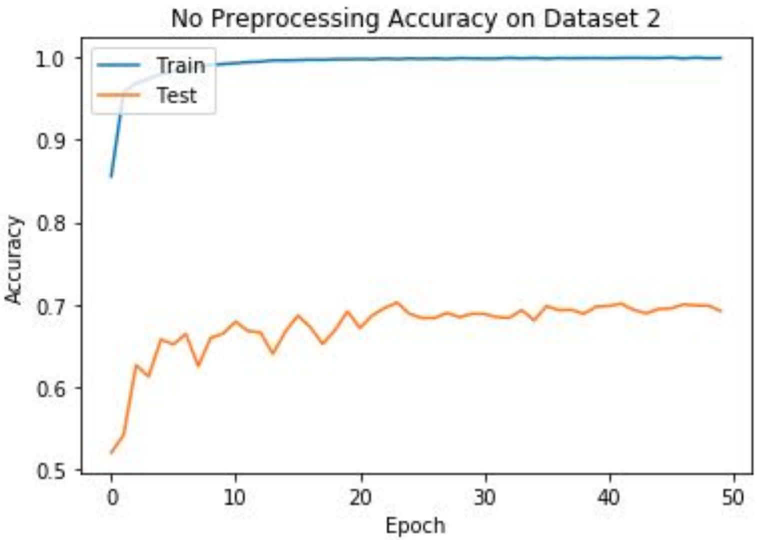


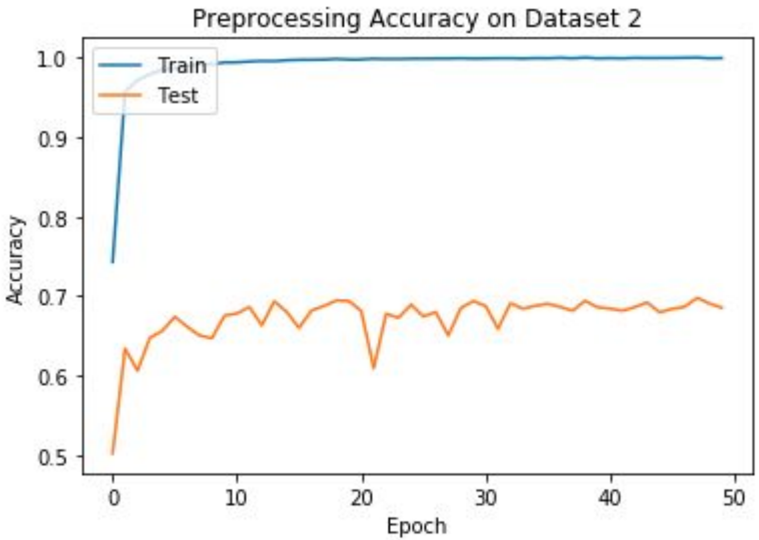
Dataset 1: Accuracy and Loss





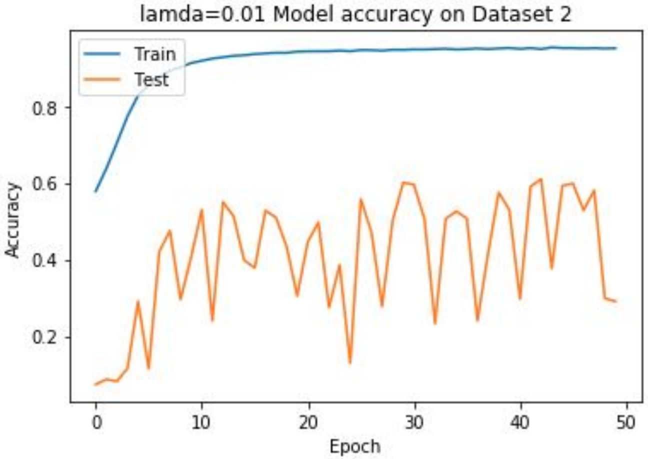
Dataset 2: Accuracy with/without Preprocessing

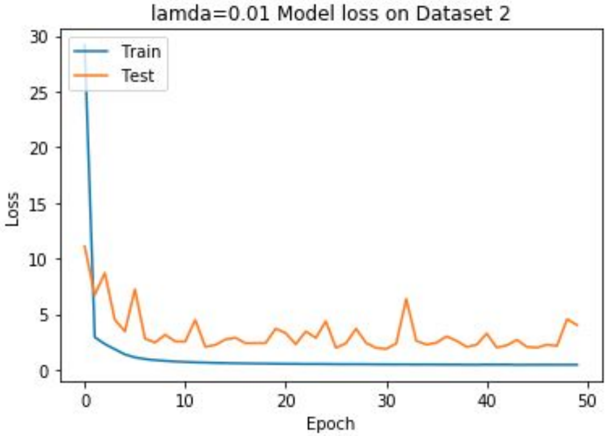




Dataset 2: Accuracy and Loss with Regularization

```
# Apply weight decay terms to l2 regularizer
decay = 0.001
for layer in self.model.layers:
    if isinstance(layer, keras.layers.Conv2D) or isinstance(layer, keras.layers.Dense):
        layer.add_loss(keras.regularizers.l2(decay)(layer.kernel))
    if hasattr(layer, 'bias_regularizer') and layer.use_bias:
        layer.add_loss(keras.regularizers.l2(decay)(layer.bias))
```





Demo

Conclusions

- Our methods did not provide an improvement over the other papers results
- Dataset plays a huge role
- Dynamic gestures work better with graphical models (HMMs, etc.)

Future Work

- Implement Real-time recognition using motion detection
- Improve CNN accuracy (Better dataset? Tuning? Coarticulation)
- Use motion tracking on top of pattern matching for J and Z
- Implement words and sentences

Questions?

Pictures

https://www.researchgate.net/figure/Winner-results-of-the-ImageNet-large-scale-visual-recognition-challenge-LSVRC-of-the_fig7_324476862

https://www.dlology.com/blog/quick-notes-on-how-to-choose-optimizer-in-keras/

https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/