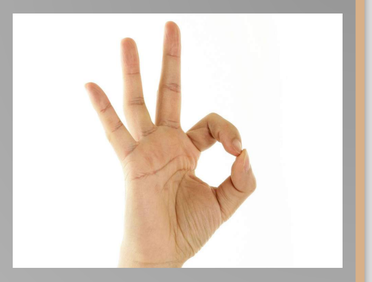
Comparing Supervised and Unsupervised learning in American Sign Language Recognition

Archana Anand, Shipra Behera, Yuchen Zhang
Professor Claire Monteleoni



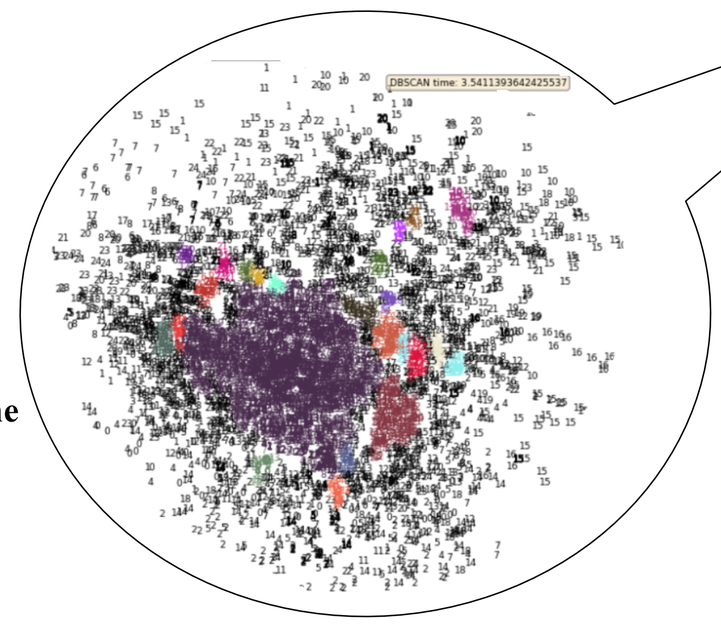
Objective

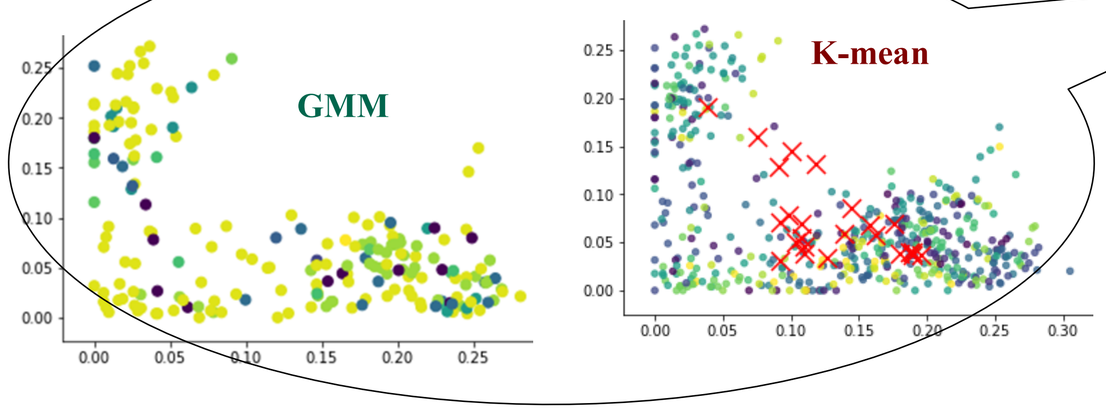
Comparison of supervised and unsupervised learning performance for American Sign language recognition. There has been a lot of ongoing work in this area, particularly involving neural networks. We will investigate different supervised and unsupervised techniques for the automatic recognition of signs and evaluate their accuracies over the 24-signs.

Clustering Algorithms

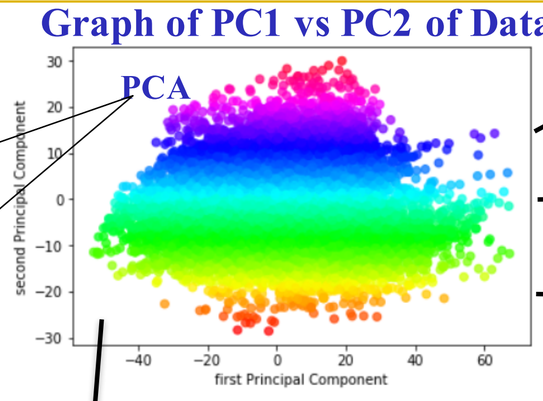
PCA+DBSCAN:
There are 43 clusters when we used PCA compared to 0 cluster when no dimension reduction methods were used

K-means and GMM with HOG:
24 clusters
We look at the labels of the clusters to understand what label each cluster corresponds to. Then we take an unlabeled data point, see into which cluster it fits best, and assign it a label.





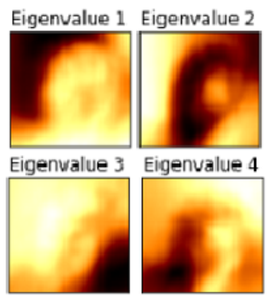
Feature Extractor/Descriptor



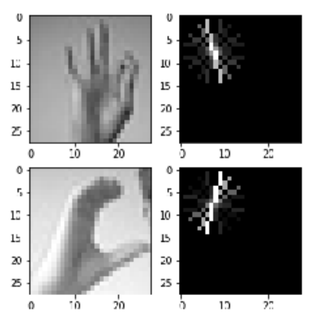
83.40% Random Forest 74.86%

82.9% AdaBoost(SAMME) 36.60%

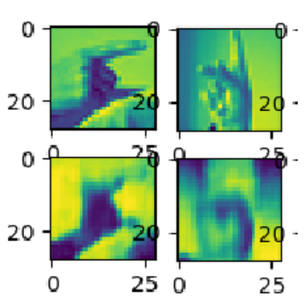
82.9% SVM 63.31%



HOG



Auto Encoder



HOG

We can see that HOG is a good descriptor for object detection, and good performance can be açhieved with Random Forest and linear SVM. One can expect even better performance with kernel SVM, if the computational complexity is not considered.

Feature Descriptor	Parameters	Classifier	Accuracy
HOG	Depth = 40, Estimator = 400	Random Forest	90.61 %
HOG	1 vs Rest	SVM	86.67 %
HOG		Naive Bayes	72.60 %
HOG	24 clusters	Kmean	50.55 %

Supervised Learning Algorithms

conv1	input	28*28*1
Kernel size 3*3*64	output	28*28*64
ReLU		

maxpool1	input	28*28*64
Kernel size 2*2	output	14*14*64

conv2	input	14*14*64
Kernel size 3*3*64	output	14*14*64
ReLU		

maxpool2	input	14*14*64
Kernel size 2*2	output	7*7*64

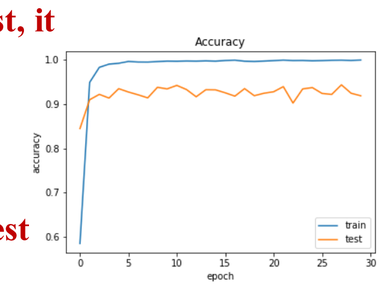
conv2	input	7*7*64
Kernel size 3*3*64	output	7*7*64
ReLU		

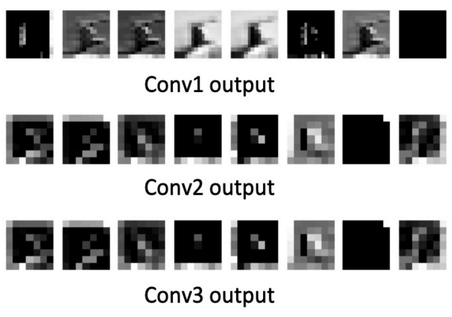
maxpool4	input	7*7*64
Kernel size 2*2	output	4*4*64

Full-connected1	input	4*4*64
Dropout 0.3	output	128*1
ReLU		

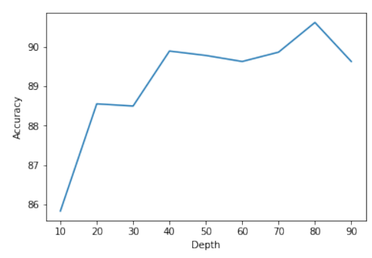
Full-connected2	input	128*1
	output	24
softmax		

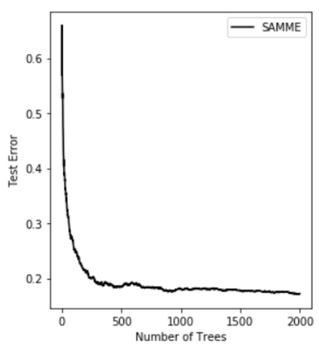
Neural Network performs best, it can classify image data without any dimension reduction. Test accuracy reaches 94.33%

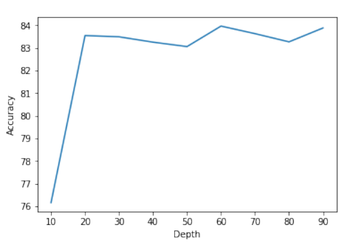


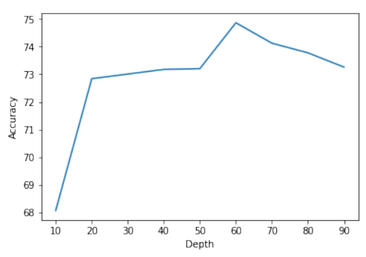


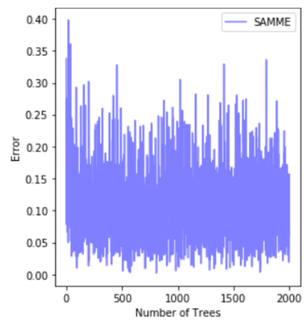
Random Forest test accuracies over PCA, HOG and Convolutional Autoencoder









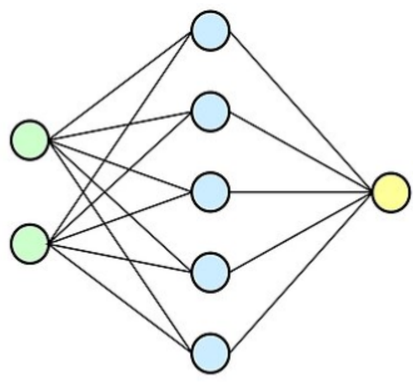


Variation of test error over number of trees for AdaBoost SAMME classifier

Classifier	Feature Descriptor	Accuracy
Neural Net	None	94.33%
Random Forest	HOG	90.61%
SVM	HOG	86.67%
AdaBoost	PCA	82.90%
Naive Bayes	HOG	72.60%

Conclusion

For different classification or Clustering algorithms, the best feature extraction method may vary. In general, supervised classification Algorithm performs better than unsupervised clustering in our project. Based on the experiment results, HOG performs better than other two feature extraction methods. Neural Network performs best, it can classify image data without any dimension reduction. The test accuracy reaches 94.33%.



References

[1] Garcia, B. & Viesca, S.A. (2016) Real-time American Sign Language Recognition with Convolutional Neural Networks.
[2] S. Ameen, S. Vadera, A convolutional neural network to classify american 530 sign language fingerspelling from depth and colour image
[3] Zheng L, Lian B, Jiang A. (2017) Recent Advances of Deep Learning for Sign Language Recognition.
[4] https://www.kaggle.com/datamunge/sign-language-mnist