Multi-Layered Dictionary Learning for Face Spoofing Detection

Atif Ahmed (aa3931)

Srinidhi Srinivas (ss5145)

Thejaswi Muniyappa (tm2848)

28th April, 2017

Introduction

- Ease of availability of silicone masks has increased the crime related to face spoofing.
- Spoofing is difficult to be identified by single-layered pre-defined dictionaries.
- Adaptive multi-layered dictionaries for sparse approximation provide a superior performance for such applications.
- Motivation: Improve the face spoofing detection using multi-layered dictionaries.

Problem Statement

"Face Spoofing Detection in unconstrained environment"

- Given an image, compute the sparse approximation and classify if the image is spoofed.

- How to compute the sparse approximation?
  Learn the dictionary and sparse approximation of the input dataset of images.

- How to classify?
  Train the classifier using the sparse representation of the input data.

Literature Survey

- Presentation Attack Detection (PAD) algorithms rely on domain knowledge.
- Properties like Motion, Texture, Reflectance and Image Quality are considered by PAD.
- Image Spoofing is difficult to incorporate texture in Boltzmann machines.
- Regularized Sparse Coding is also used for Image Representation

Multi-layered Dictionary Learning

- Let D_1 be the Dictionary at the first level, X be the input data and Z be the sparse representation.

$$\boldsymbol{X} = \boldsymbol{D_1 Z}$$

This is similar to the shallow one-level dictionary learning problem

- Extending the shallow set-up to the second layer, we get

$$\boldsymbol{X} = \boldsymbol{D_1 \varphi(D_2 Z)}$$

Here, φ is the activation function

- The full optimization for the N layers can be written as:

$$\boldsymbol{X} = \boldsymbol{D_1 \varphi(D_2 \varphi(\ldots(D_N Z)))}$$

- Hence, the minimization problem becomes

$$\min_{\mathbf{D_1, D_2, \ldots D_N}} ||\boldsymbol{X} - \boldsymbol{D_1 \varphi(D_2 \varphi(\ldots(D_N Z)))}||_F^2$$

Multi-layered Dictionary Learning via Greedy Algorithm – 1st Layer

- The entire problem can be divided into a smaller bunch of greedy problems.
- For the first layer,

$$Z_1 = \varphi(D_2\varphi(\ldots(D_N Z)))$$

$$\min_{D_1, Z_1} ||X - D_1 Z_1||_F^2$$

- This can be solved by Method of Alternating Directions

$$Z_k \leftarrow \min_{Z} ||X - D_{k-1} Z||_F^2$$

$$D_k \leftarrow \min_{D} ||X - D Z_k||_F^2$$

Multi-layered Dictionary Learning via Greedy Algorithm – Continuing it with other layers

- After formulating the coefficients for the first layer, we formulate that of second layer as:

$$\boldsymbol{\varphi^{-1}(Z_1) = D_2 Z_2}$$

$$\boldsymbol{Z_1 = \varphi(D_3\varphi(\ldots(D_N Z)))}$$

- Method of alternating directions is invoked again to find to solve the optimization at every level

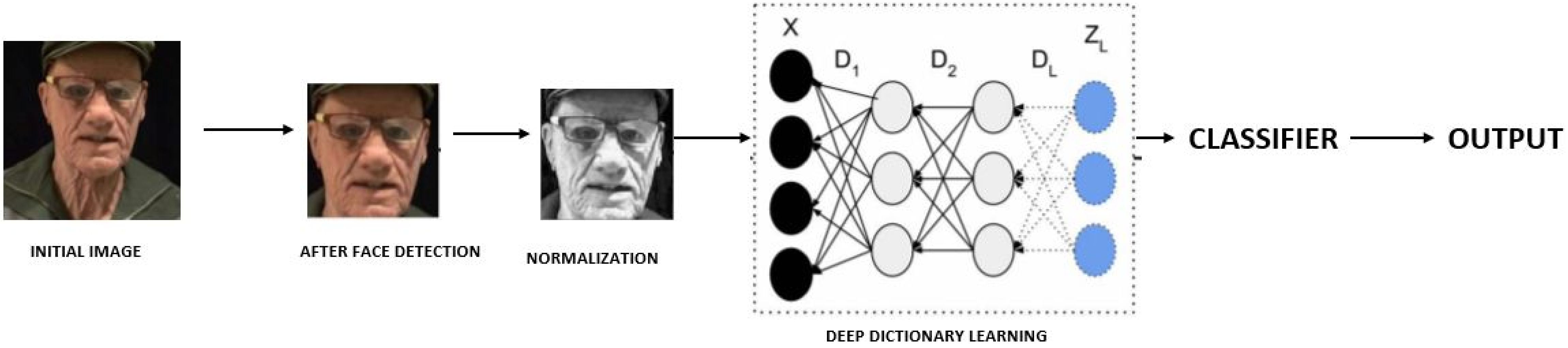
- Sparsity can be introduced to this optimization problem by re-writing it as:

$$\min_{\mathrm{D_1,D_2,\ldots D_N}} \left|\left|X - \boldsymbol{D_1\varphi(D_2\varphi(\ldots(D_N Z)))}\right|\right|_F^2 + \lambda\left|\left|\boldsymbol{Z}\right|\right|_1$$

- This can be imposed at the final layer as:

$$\min_{\mathrm{D_N},Z} \left|\left|\varphi^{-1}(\boldsymbol{Z_{N-1}}) - \boldsymbol{D_N Z}\right|\right|_F^2 + \lambda\left|\left|\boldsymbol{Z}\right|\right|_1$$

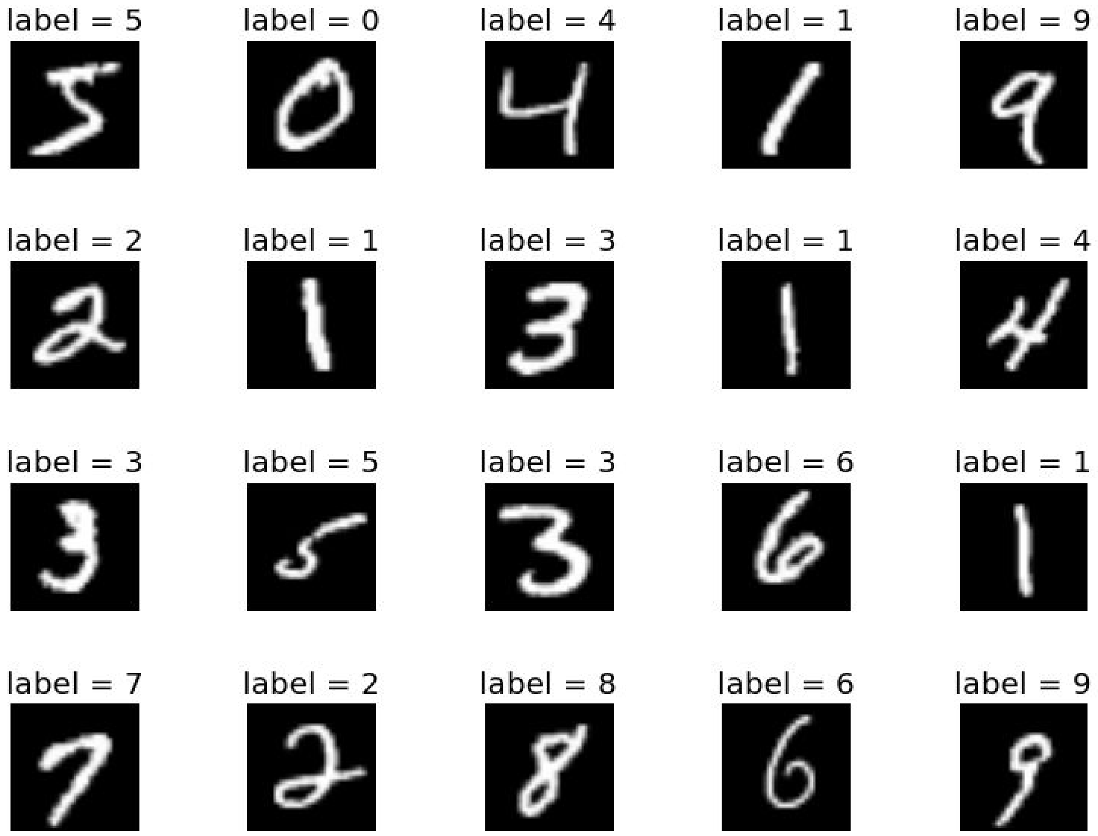
Architecture



Dataset - Evaluation on MNIST

- Dataset of handwritten digits provided by Yann Lecun*.
- 28 x 28 images for 10 classes

Training	Test
40,000	8,000



* http://yann.lecun.com/exdb/mnist/

Dataset - Training on MNIST

- First we tried with **all features.**

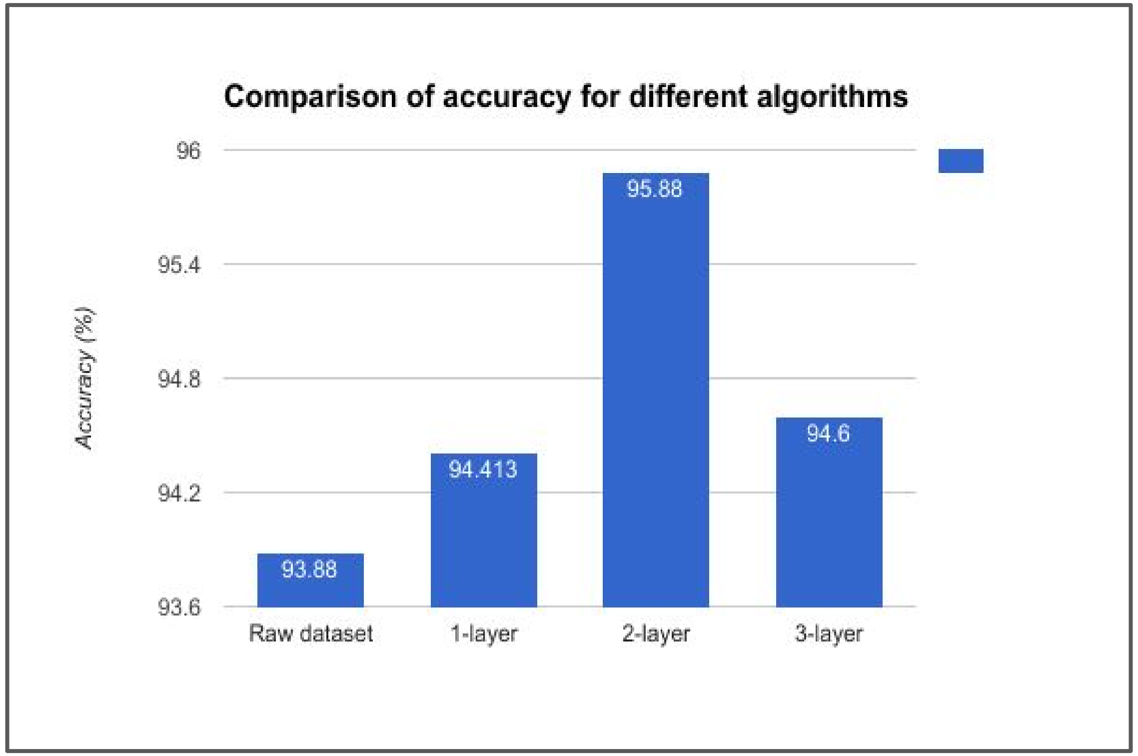
Dataset	Feature Size	Training Size	Test Size
MNIST	784	40,000	8,000

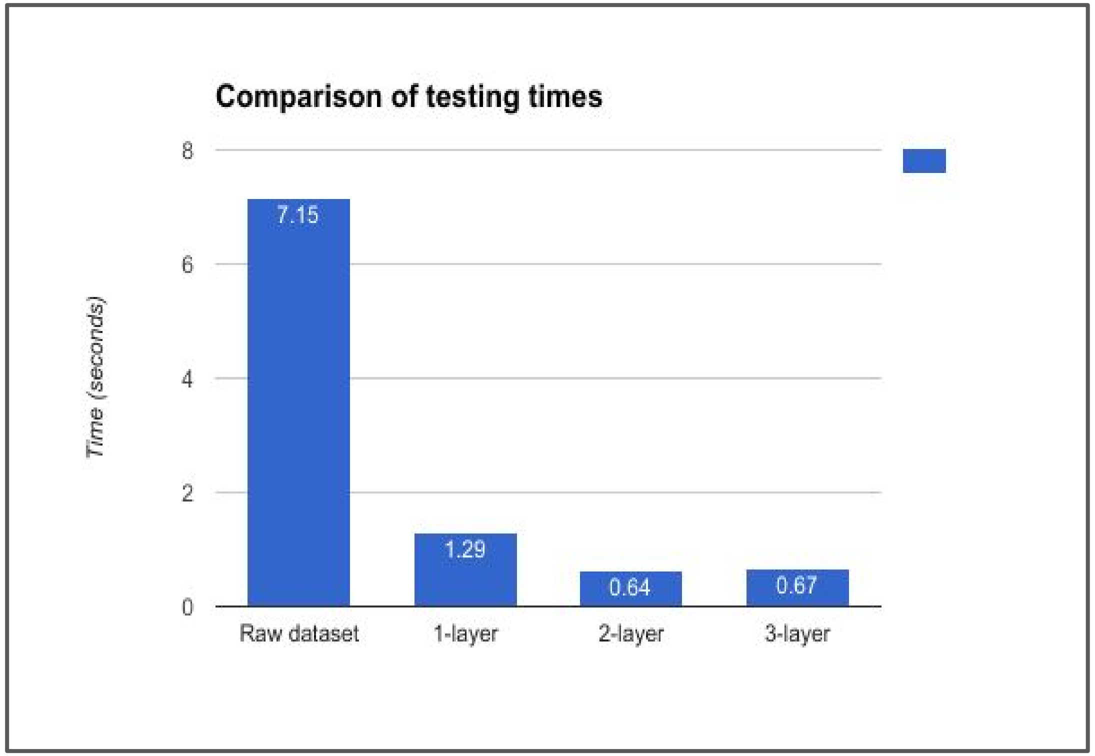
- Dimensionality Reduction with **Dictionary learning.**

Dataset	Feature Size (1 Layer)	Feature Size (2 Layer)	Feature Size (3 Layer)	Training Size	Test Size
MNIST	100	200 - 100	400 - 200 - 100	40,000	8,000

Testing and Results on MNIST

MNIST Dataset	Raw Dataset	Shallow (1-layer)	2-layer	3-layer
Accuracy (%)	93.58	94.413	95.88	94.6
Testing Time (sec.)	7.15	1.29	0.64	0.67

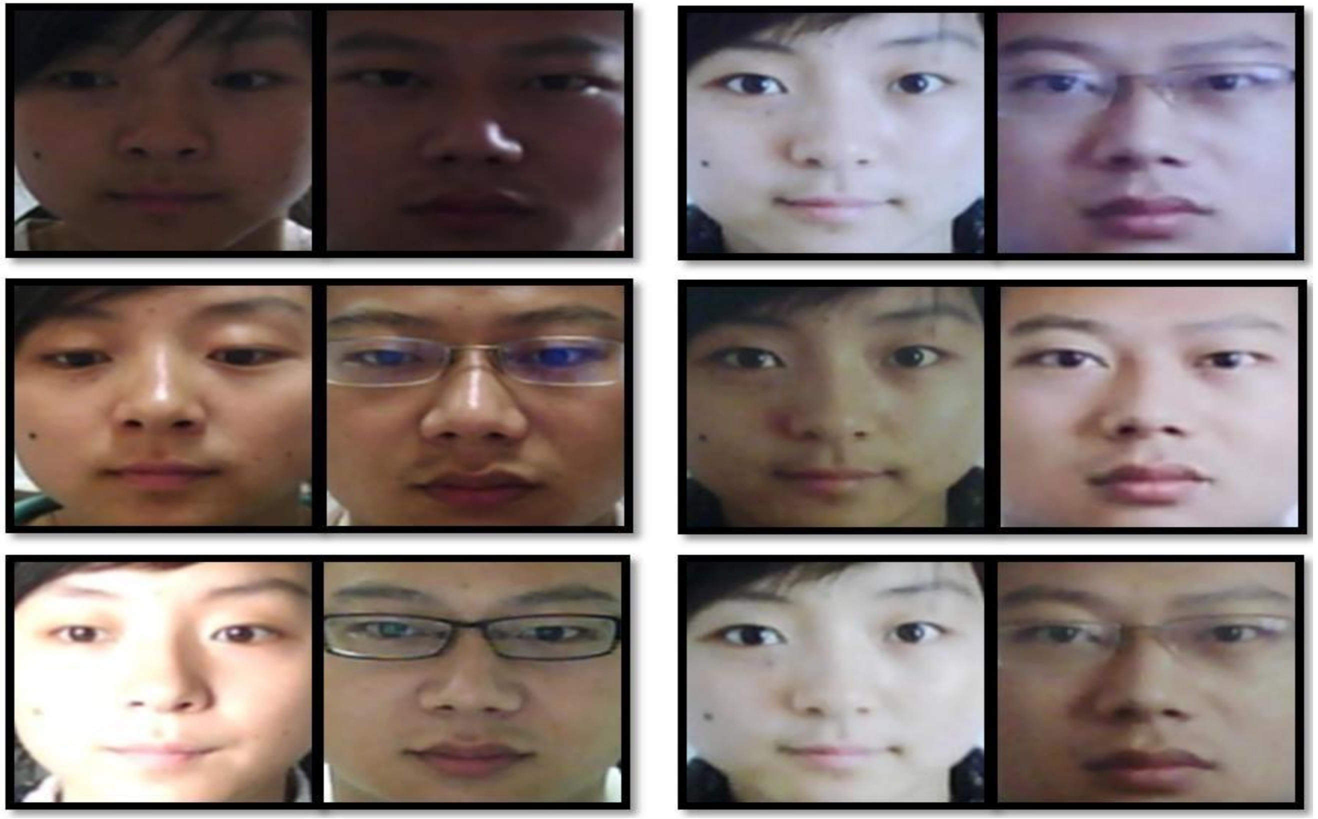




Dataset - NUAA

- Database of photograph imposters collected with cheap webcams.
- 64 x 64 images of 2 classes (client or impostor).

Training	Test
3491	9123



- We extract the face using Viola Jones face detection algorithm.

Original Image

- After detecting the face, the image is converted to grayscale and is normalized

- All the cropped images are 64 x 64. Hence, every image is represented as a vector of length 4096

Training on NUAA

- First we tried with **all features.**

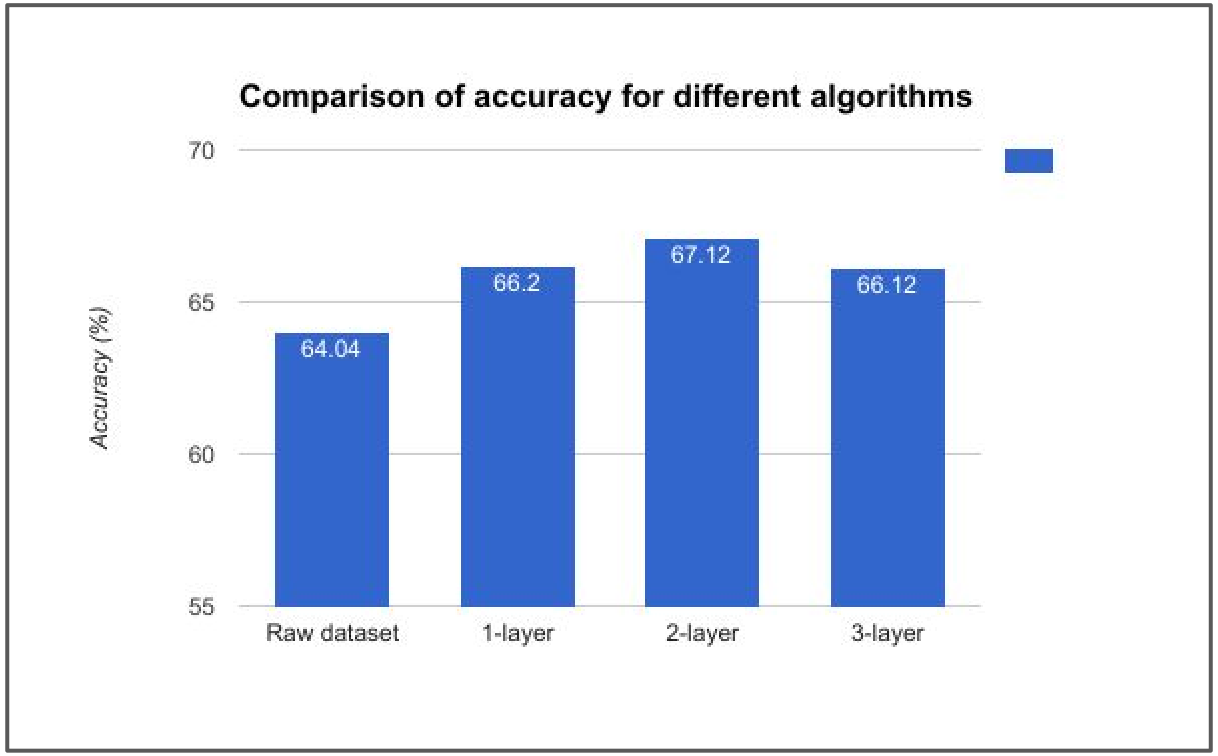
Dataset	Feature Size	Training Size	Test Size
MNIST	784	40,000	8,000
NUAA	4096	3,491	9123

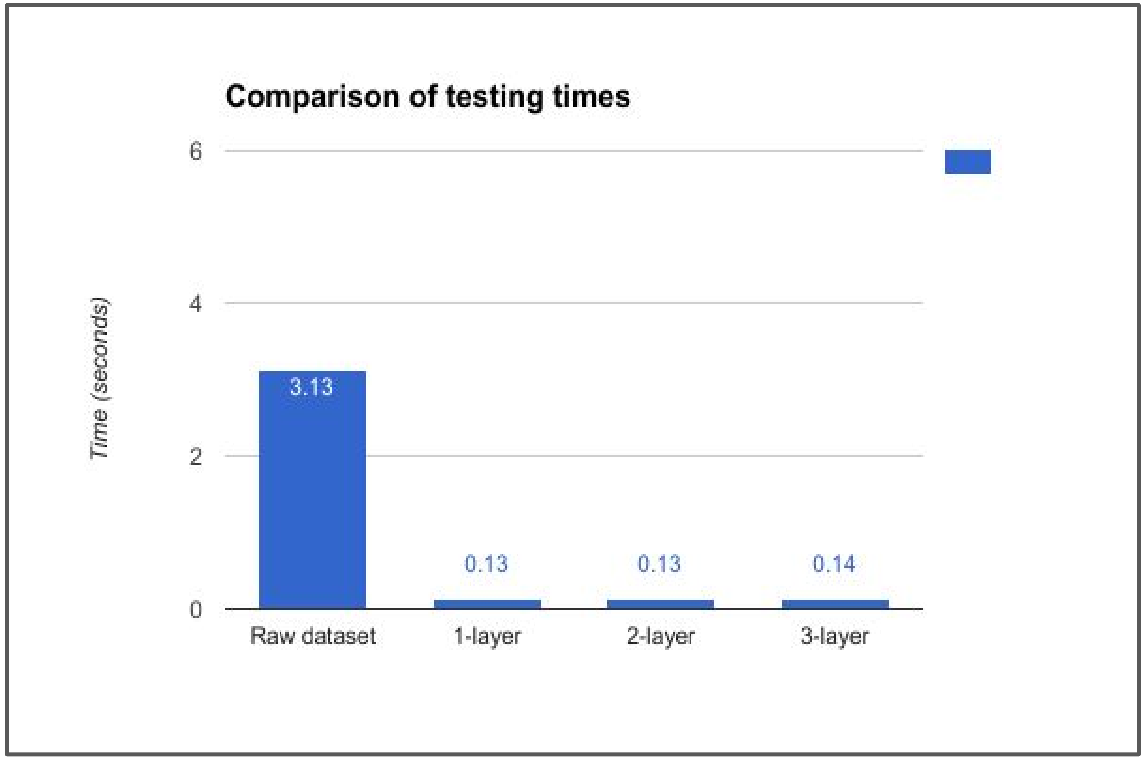
- Dimensionality Reduction with **Dictionary learning.**

Dataset	Feature Size (1 Layer)	Feature Size (2 Layer)	Feature Size (3 Layer)	Training Size	Test Size
MNIST	100	200 - 100	400 - 200 - 100	40,000	8,000
NUAA	512	1024 - 512	1024 - 512 - 256	3,491	9123

Testing and Results on NUAA

MNIST Dataset	Raw Dataset	Shallow (1-layer)	2-layer	3-layer
Accuracy (%)	64.04	66.2	67.12	66.12
Testing Time (sec.)	3.13	0.13	0.13	0.14





Effect of various activation functions

- The reference paper [2] uses only linear activation functions in all the layers.
- We tried with other activation functions also like tanh and sigmoid.
- In the algorithm, inverse of the activation functions have to be used.
- At each layer we project the values to a specified interval (e.g. [-1, +1] for arctanh).

Layers	200, 100	400, 200	50, 25
Accuracy	63.86	51.31	60.61

Challenges for multi-layered dictionary learning

- Studies have proven the convergence guarantees for single level dictionary learning. These proofs are very hard to replicate for multiple layers.

- The number of parameters required to be solved increases when multiple layers of dictionaries are learned simultaneously. With limited training data, this could lead to overfitting.

Conclusion

- Multi-layered dictionary gives better results and testing times than shallow dictionary learning.
- Learning different levels of dictionaries along with the coefficients is not the same as learning a single (collapsed) dictionary and its corresponding features (even with linear activation functions).
- The single-level problem (single level) is a bi-linear problem and 2-level problem is a tri-linear problem
- Greedy approach for layer-wise learning ensures the convergence of each layer.
- Regularization at final layer can be used to control sparsity and prevent overfitting.

Future Work

- Using images in 2D format instead of a vector. The data will then become multi-dimensional which calls for multi-dimensional dictionary learning.

- Using other kind of sparsity measures and comparing performance.

- Improving performance for nonlinear activation functions.

References

[1] Ishan Manjani, Snigdha Tariyal et. al "Detecting Silicone Mask Based Presentation Attack via Deep Dictionary Learning" *IEEE January 2017*

[2] Snigdha Tariyal, Angshul Majumdar "Deep Dictionary Learning" *IEEE May 2011*

[3] Teresa Bufford, Yuxin Chen "When Dictionary Learning Meets Classification" *UCLA April 2013*

[4] Di Wen, Hu Han "Face Spoof Detection with Image Distortion Analysis" *IEEE Transactions on Information Forensics and Security 2015*

[5] Miao Zheng, Jiajun Bu "Graph Regularized Sparse Coding for Image Representation" *IEEE May 2011*

[6] Yang, Jimei, and Ming-Hsuan Yang. "Top-down visual saliency via joint crf and dictionary learning." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2016).

[7] X.Tan, Y.Li, J.Liu and L.Jiang "Face Liveness Detection from a single image with sparse low rank bilinear discriminative model " *Proceedings of 11th European Conference on Computer Vision (ECCV 10), Crete, Greece. September 2010*

Thank You!