MATH 6380O Mini-Project 1: Feature Extraction and Transfer Learning

Ng Yui Hong[1] yhngap@connect.ust.hk

[1]: Department of Mathematics, HKUST

1. Introduction

We have used scattering network, VGG16 and Restnet-18 to extract the features on MNIST dataset. Then Visualized the extracted features using PCA, t-SNE and various method. Image classification is done by the traditional supervised learning method like SVM, RF, etc..

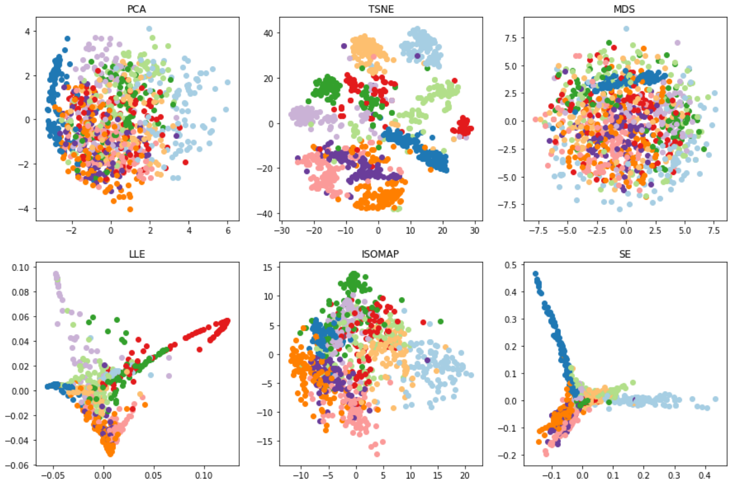
1.1 Dataset

We use MNIST as our dataset in this project.
MNIST is a simple yet effective dataset to explore the learnt feature of networks, given it is made up of numerous hand-written digits. It consists of 60000 gray-scale training images of size 28x28 *and 10 classes. In our experiment we resize all the images to 224*x224 and choose only a subset of the train/test images due to time concern.

2. Feature Extraction Parameter

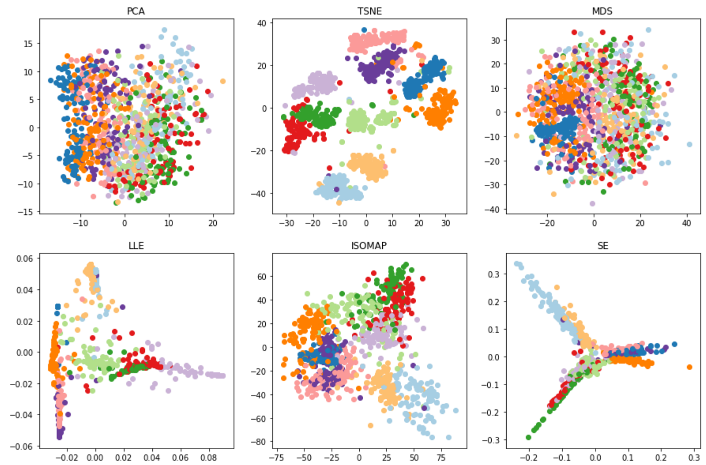
Feature dimension\Network	Scattering Net	Resnet	Vgg19
No. of dimension	1953	512	25088

3.1 Feature Visualization (Scattering net)



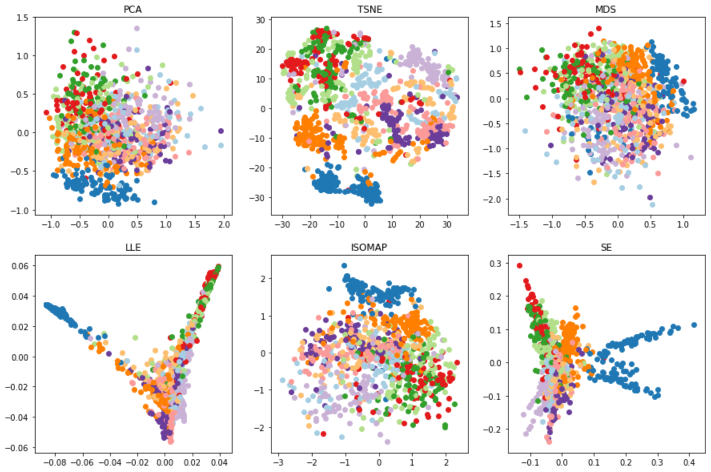
From the figure above, we could find that the scattering network seem to perform well in extracting and preserving those features in the MNIST dataset since there exist clusters in all these visualization figures, which means the key features of the data are well extracted and projected.

3.2 Feature Visualization (VGG 19)



VGG seems to perform better than Scattering Network from the visualization aspect. The clusters are clear cut and the cluster mixture seems to appear less than that of the Scattering Network. The t-SNE, SE and ISOMAP visualizations quite clearly indicate the ability in non-linear structure extraction of the VGG19 model

3.3 Feature Visualization (VGG 19)



Resnet seems to perform worse than other two models. Although the visualizations clearly project the extracted features, we could see that the Resnet seem to mix some features and cannot distinguish certain features with other features.

4. Visualization Summary

1. Deep Learning Networks Comparison through the Same Visualization

From the visualizations, we could see that Scattering Network and VGG19 seem to perform better than ResNet50. The features extracted by Scattering Network and VGG19 and dimensionally reduced by these visualization approaches are presented better than those of ResNet50. It implies that Scattering Network and VGG19 are better in extracting and preserving the features in the MNIST dataset.

2. Different Visualization Approaches through the Same Network

From the visualization figures above, non-linear structure in the data is very significant since PCA which tries to find the linear transformation among dimensions and MDS which measures the distance similarity between objects seem to mix clusters, which means some features may be very similar even though they are from different individuals in the dataset. Since t-SNE is particularly sensitive to local structure, while ISOMAP, LLE and variants are best suited to unfold a single continuous low dimensional manifold, t-SNE will focus on the local structure of the data and will tend to extract clusters local groups of samples. Therefore, we could clearly see a separate cluster for every digits and it shows that t-SNE is a superior way to visualize, reduce dimensionality for samples which exhibit manifold structure.

5. Classification Summary

	SVM	LDA	RandomForest	Logistic Regression
ScatNet (CE)	1.53 ± 0.13	0.42 ± 1.12	0.52 ± 0.13	0.22 ± 0.12
ScatNet (Acc)	0.18 ± 0.11	0.98 ± 0.06	0.94 ± 0.09	0.98 ± 0.06
VGG19 (CE)	0.81 ± 0.22	13.78 ± 8.53	0.58 ± 0.18	0.08 ± 0.11
VGG19 (Acc)	0.54 ± 0.22	0.24 ± 0.23	0.94 ± 0.09	0.98 ± 0.06
ResNet50 (CE)	1.62 ± 0.11	7.56 ± 6.95	0.76 ± 0.32	1.03 ± 0.22
ResNet50 (Acc)	0.18 ± 0.11	0.76 ± 0.22	0.86 ± 0.13	0.8 ± 0.13

In above table, CE stands for cross entropy and acc stands for accuracy.
As shown in the table, logistic Regression and Random Forest are the best. Especially when compared with the large gap between different features shown by the remaining two methods, we may conclude that Logistic Regression and Random Forest are less likely to be influenced by the different (quality of the) features.

We may further analyze the three types of features on a more general level. Features extracted by deep nets may be too complicated (and not necessary) for such simple tasks. This may be more obvious when we compare ResNet50 with VGG19, since the former one shows inferior results. Also, to feed the correct size into such deep networks, we need to resize the images to much larger sizes, which may degrade the image quality.

Reference:

Oyallon, E., Belilovsky, E., & Zagoruyko, S. (2017, October). Scaling the scattering transform: Deep hybrid networks. In International Conference on Computer Vision (ICCV).
Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).