MATH6380P Project-1: Feature Extraction and Transfer Learning

LIANG Zhicong, HUANG Zhichao, WEN Ruixue
Department of Mathematics, Hong Kong University of Science and Technology

Introduction

This is the report of MATH6380P project-1, 2018 Fall, HKUST. It is aiming at extracting various features of the same datasets through different models, visualizing these features and classifying images based on these features.

We use three models for feature extraction, which are ResneXt[1], VGG[2], Scatter Net[3]. To compare the features extracted by these models, we use unsupervised method to visualize these features. They are PCA, MDS, t-SNE. Besides, we also train multiples classifier to estimate the performance of the three models above, including SVM, Random Forest, Logistic Regression and LDA.

We use CIFAR-10 as the dataset to conduct the experiment. It contains 10 categories of images with size 32*32. It is much more complex than MNIST, but it is also a smaller dataset that feature can be extracted by CPU in reasonable time.

Classifiers will be trained on features extracted from 50000 training samples and accuracy will be evaluated on 10000 testing samples.

Feature Extraction

VGG: By stacking convolution layers, VGG has successfully expended the receptive filed of its 3*3 filters, increased nonlinearity while reduced the number of parameters. Feature is extracted from the last pooling layer, (bottleneck) with dimension of 512.

ResneXt: Resnet use add residual path to build a more powerful neural network. It could help solving the problem of gradient vanishing. We extract feature from the final pooling layer like what we do in VGG. The dimension of the feature is 1024.

Scatter Net: ScatNet extracts features by applying a scattering propagator, which is based on wavelets. One of its advantages is keeping the invariance of translation and rotation. By changing the parameter of filters, we can get different sizes of features and choose a best one eventually according to the classification accuracy.

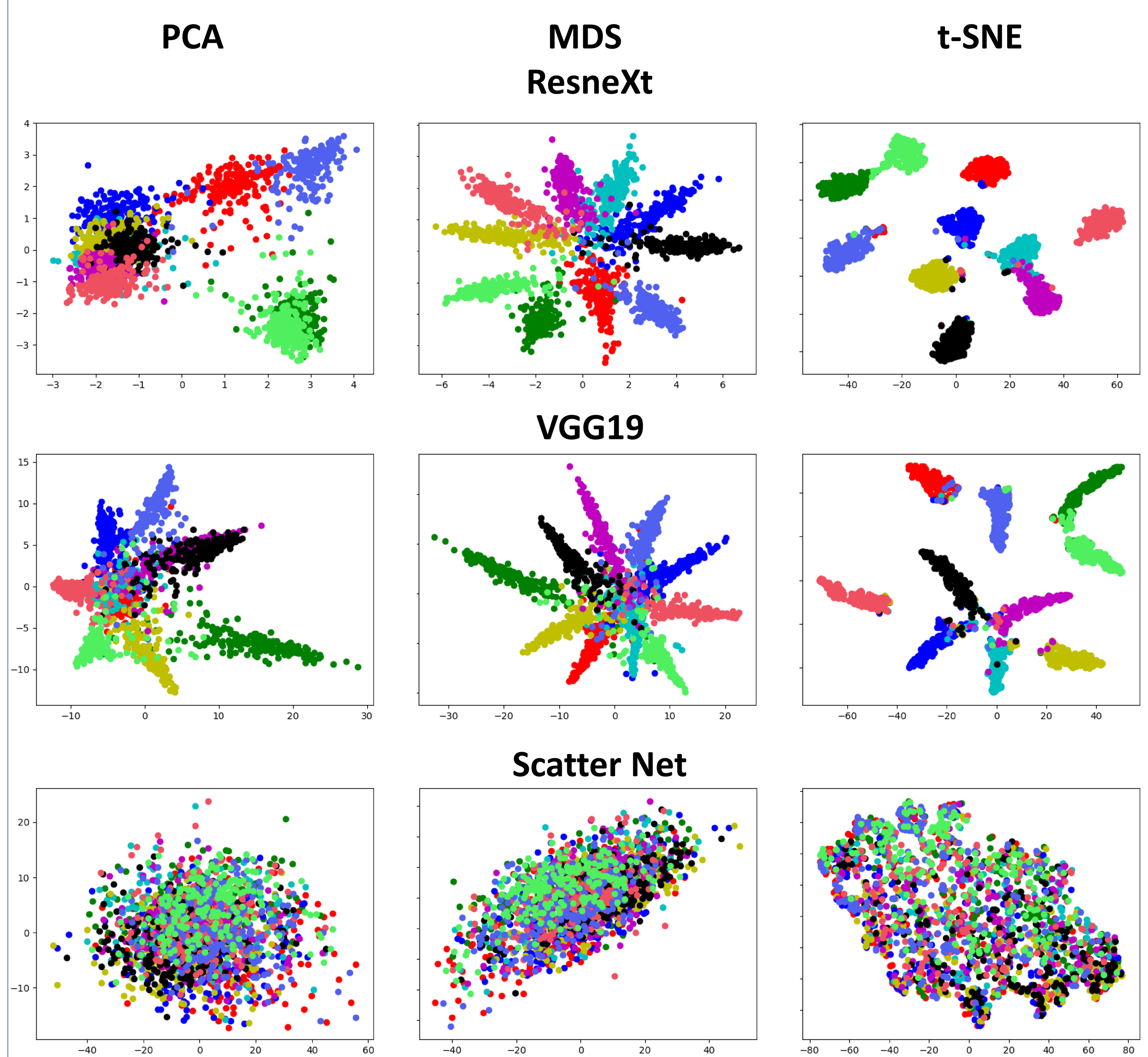
Considering computational expense, we will used pre-trained VGG and ResneXt to extract features instead of training ones from scratch.

Visualization

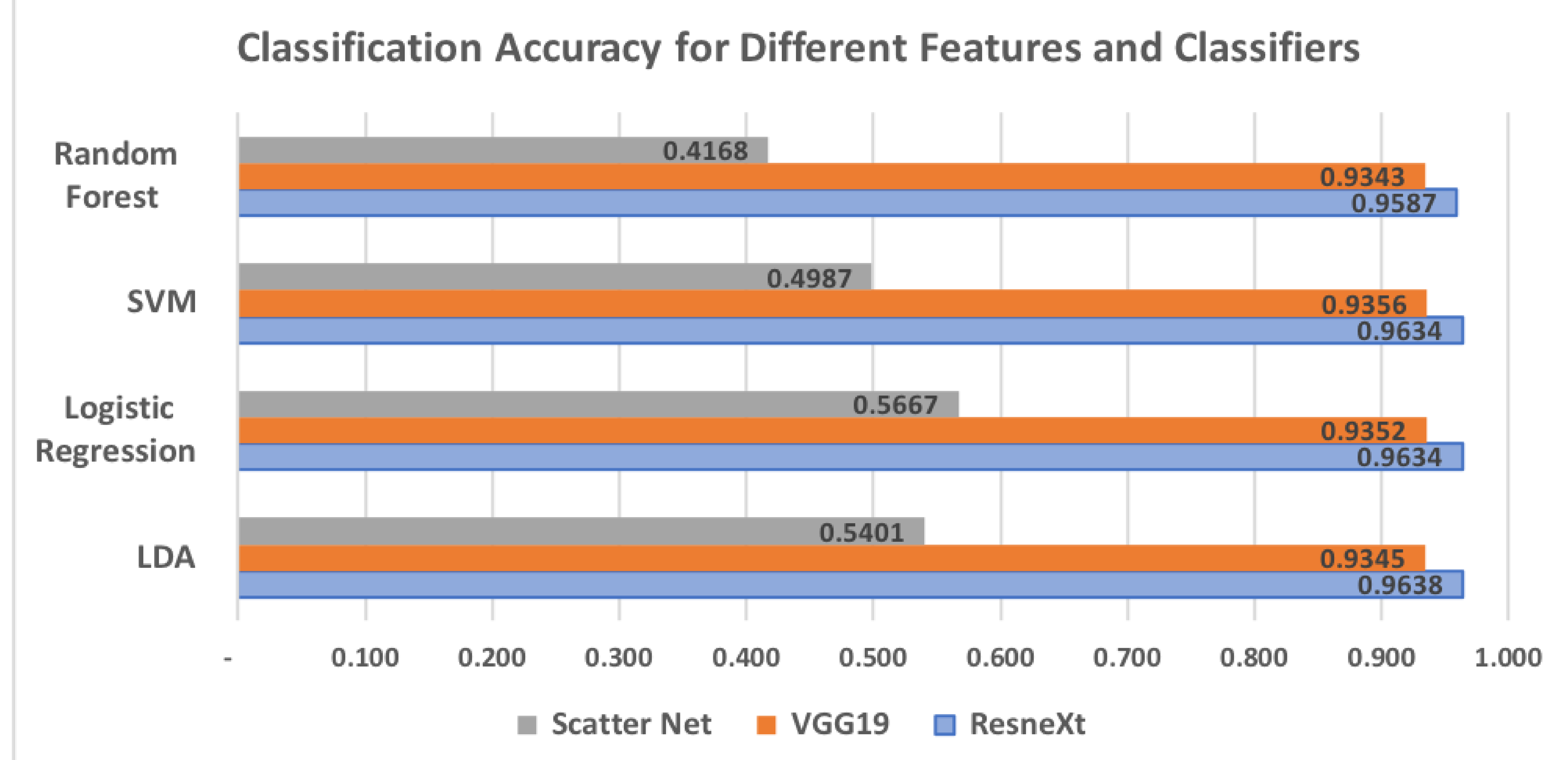
We use three techniques to visualize the feature: PCA, MDS and t-SNE. Each color represents a category.

As shown below, PCA is not good to reduce the dimension and visualize the feature because it is only a linear operation. Nearly all classes are not separated. However, the classification result shows these features are indeed very good for class separation. On the other hand, MDS and t-SNE are good for visualization.

Features extracted by pretrained model like ResneXt and VGG express the distinction of different classes very well. In fact, they all achieve over 90% accuracy in the origin models and our classifiers. But features extracted by Scatter Net does not have clear boundary in 2 dimensional embedding space. Visualization figures showed below are plotted from the testing samples.



Classification



Conclusions and Discussion

The classification accuracy of VGG and ResNet are up to 93.56% and 96.38% respectively. However, there is an obvious drop in classification accuracy in Scatter Net. We proposed several explanation for this phenomenon.

- One reason is that images in CIAFR-10 are of too small size to represent complicate items with clearer texture, while the highlight of Scatter Net is its ability in capturing texture features.
- Secondly, the parameters in Scatter Net are customized and no learning process is involved, which means it cannot adapt to the dataset by itself, lacking flexibility compared with the other two method
- In addition, both VGG and ResneXt are trained aiming at classification, so they tend to extract features which are prone to classification (even cheating), while Scatter Net just extracts features by pre-established parameters without goal. In this case, parameter tuning and feature selection become essential to Scatter Net.

Future

To verify our hypothesis, more experiments need to be conducted on two kinds of image datasets, one of which contains more detailed textures while the another contains more deformations like rotation and scaling. Feature selection should be taken into consideration, too.

Links

- Codes: https://github.com/ZhicongLiang/deeplearning-project-1
- Scatter Net: https://www.di.ens.fr/data/software/scatnet/
- CIFAR-10: https://www.cs.toronto.edu/~kriz/cifar.html
- MATH6380P: https://deeplearning-math.github.io/

References

1.Saining Xie et.al. (2016). Aggregated Residual Transformations for Deep Neural Networks. *arXiv preprint arXiv:1611.05431.*

2, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

3. Mallat, S. (2012). Group invariant scattering. *Communications on Pure and Applied Mathematics*, *65*(10), 1331-1398.