MATH 6380P Project 1: Feature Extraction & Transfer Learning

Jimmy WU, Zhaoyu SU

Dept. of Civil & Environmental Engineering, HKUST

Introduction & Objectives

In this study, the wavelet based scattering net is compared with two popular deep convolutional neural network (DCNN) models regarding the ability of feature extraction.

Features extracted from the same batch of input data by both scattering net and DCNN models are subsequently compared both visually (t-SNE) and quantitatively (logistic regression), to check their feature extraction abilities.

Dataset

The dataset used in this study is the well-known MNIST, due to the following reasons:

- MNIST comprises only gray-scale images, leading to less computation requirement, especially for the final classification.
- MNIST has only 10 classes, which eases the t-SNE visualization.

Network Architectures

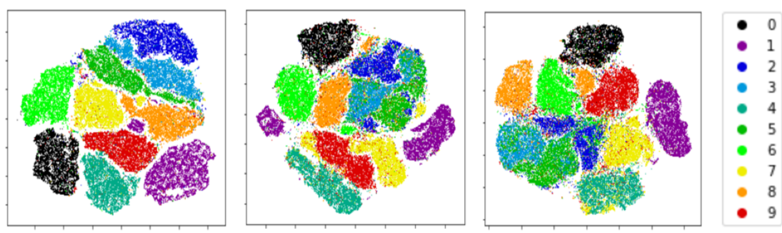
Wavelet Scattering Net

In this work, the wavelet scattering net is implemented following [1], and we adopted a 4-layer network architecture using pytorch and cuda.

DCNN Models

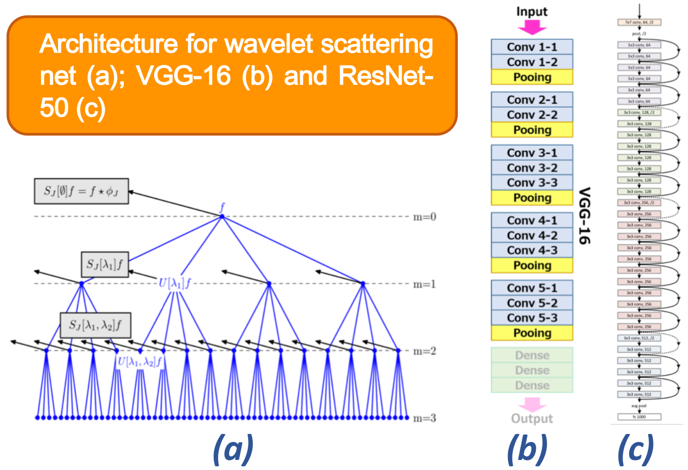
The DCNN models adopted here are VGG-16 [2] and ResNet-50 [3], and to get the features representations, both of these models are cut before the fully connected layer.

The features acquired from the different networks are firstly passed to the t-SNE for visualization. By comparison, the result from wavelet net seems to be the best among all the three networks, which has the largest degree of separation. But surprisingly, the ResNet-50 gets the worst performance.

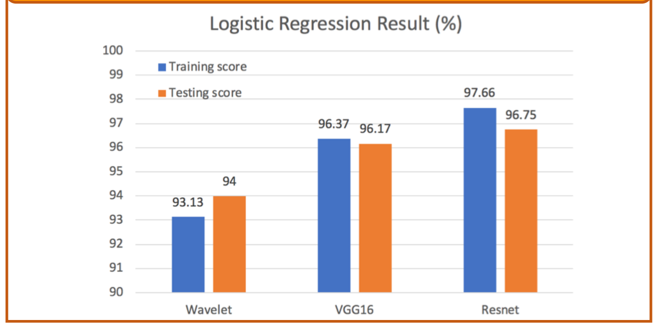


t-SNE visualization for scattering net, VGG-16 and ResNet-50 (from left to right)

Architecture for wavelet scattering net (a); VGG-16 (b) and ResNet-50 (c)



Supervised Logistic Regression for classification



Conclusion

Paradox results arise through visual and quantitative comparisons: t-SNE visualization indicating that wavelet scattering net seems has the largest feature separation degree, while the quantitative result just gives the opposite conclusion. After discussion, several potential answers are concluded as follows: 1) the number of feature representation from these three models are different, while scattering net has only 1,600 features, much less than the DCNN models (more than 2,000), which gives rise to the better visualization result; 2) t-SNE visualization cannot fully represent for the real case, as there is serious information loss during the dimension reduction.

Reference :

1. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
2. Oyallon, Edouard, Eugene Belilovsky, and Sergey Zagoruyko. "Scaling the scattering transform: Deep hybrid networks." In *International Conference on Computer Vision (ICCV)*. 2017.
3. He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770-778. 2016.