MATH 6380P MINI-PROJECT 1
Feature Extraction and Transfer Learning on Fashion MNIST

Yipai DU (yduaz@connect.ust.hk) and Yongquan QU (yquai@connect.ust.hk)

Introduction

Fashion-MNIST is a dataset with the hope to better represent modern CV tasks. In this project, we hope to check the alignment of neural collapse in training and its classification accuracy. we apply feature extraction by scattering net, pre-trained VGG19 and ResNet18. We then visualize these features by PCA and t-SNE. For classification, we use LDA, Logistic Regression and Random Forests based on both raw image and extracted features. We train the last layer and fine tune the pre-trained neural networks on Fashion MNIST, compare the classification accuracies of all the methods mentioned. We check the consistency of accuracy in training data and neural collapse phenomenon.

Feature Extraction

Wavelet Scattering Network is in fact not a trainable network, but a transform with fixed parameters based on wavelet transform. The transform itself is a translation-invariant signal representation implemented as convolutional networks of wavelet filters. Our implementation is based on Kymatio library. The aim is that through the translation-invariant transform the extracted features can be easily classified.

Pre-trained Neural Networks: VGG19 and **ResNet18** are two famous neural networks. VGG shows that classification accuracy can be improved by using deeper convolutional networks with smaller kernels, ResNet solves the gradient explosion and vanishing problem. Here, we use VGG19 and ResNet18 pre-trained on ImageNet dataset, as feature extractors, the shapes of features are (4096) and (512) respectively.

Transfer Learning

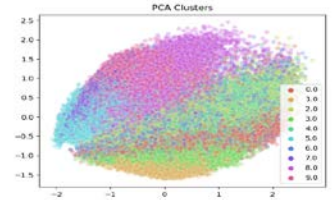
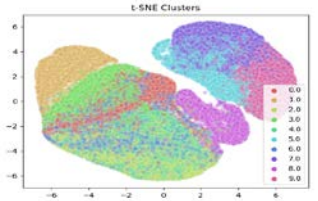
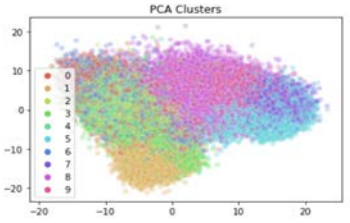
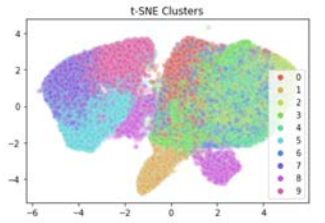
Fine Tuning is a common technique in transfer learning. Although the images in ImageNet can be different from that in Fashion MNIST, we can still utilize features learned by pre-trained neural networks. Here we again use VGG19 and ResNet18 with appropriately modified full connected layers, train the last layer as well as fine-tune them on Fashion MNIST dataset.

Contribution

Yongquan QU: Feature extraction by pre-trained networks; Calculation, Visualization and Classification of corresponding features; Transfer Learning
Yipai DU: Feature extraction by Wavelet Scattering Network; Calculation, Visualization and Classification of corresponding features and original Fashion MNIST dataset

Visualization

PCA and **t-SNE** are used to visualize the features extracted by the networks. Specifically, PCA is used to project the extracted features to 2D space. t-SNE is carried out by applying firstly PCA to reduce the dimensionality to 50 and then t-SNE to cast them to 2D space. The point colors represent different classes. We can see from the plots that the extracted features looks promising since they are grouped well even in 2D space, indicating simple classifiers may work well.

	PCA Clusters	t-SNE Clusters
Scat Net		
VGG19		

Neural Collapse Statistics

	Scattering Net	Pre-trained VGG19	Pre-trained ResNet18	Fine-tuned ResNet18	Raw Image
NC1	2.225E8	-3.082E8	-1.456E9	4.98E6	1.051E9
Equal norms	0.2364	0.2172	0.1729	0.2493	0.2144
Equal angle	0.5465	0.4615	0.3781	0.4149	0.4900
Maximal angle equiangularity	0.4696	0.4042	0.3127	0.3703	0.4161

Classification Results

We train simple classifiers listed below on extracted features and raw images. In each entry, the first number indicates training accuracy while the second is testing accuracy. Overall we can conclude Fashion-MNIST is a simple dataset that can be classified even with simple classifiers directly with more than 80% accuracy. But to some extent feature extraction gives a bit improvements.

	Scat Net	VGG19	ResNet18	Raw Image
LDA	93.15% 90.19%	91.15% 88.47%	87.80% 86.90%	83.26% 81.51%
Logistic Regression	85.73% 84.96%	91.70% 87.05%	88.51% 87.09%	85.29% 82.94%
Random Forests	100.0%, 88.79%	100.0% 84.54%	100.0% 85.98%	100.0% 87.52%
Train the last layer	N.A	N.A, 87.43%	N.A, 87.51%	N.A
Fine-Tune	N.A	N.A	N.A, 94.01%	N.A

Conclusion

In this project we applied two feature extraction techniques: scattering network and pre-trained networks and checked the consistency of training accuracy and neural collapse. From neural collapse statistics, we found feature extraction is effective for all methods compared to raw image (the closer the statistics to zero, the better the feature extraction). This conforms with the training accuracy of these networks. However, in terms of testing accuracy the conclusion is more complicated to be drawn because even though feature extraction is working well it may be overfitting on training data therefore will not work well on the testing set. That's why we see inconsistency in neural collapse and testing accuracy. Deep neural networks, although fit the training data perfectly, work worse than scattering net at test phase. Neural collapse statistics closer to zero indicating overfitting is indeed happening in these deep models. Scattering network, on the other hand, extract features without need of training, and works best among all the models.