Mini-Project 1: Feature Extraction and Transfer Learning

Rongrong GAO (20619663), Junming CHEN (20750649), Zifan SHI (20619455)

1 Introduction

Feature extraction plays a decisive role on the performance of deep learning models. In this project, we study the influence of different feature extraction methods on classification task as well as the transferability of VGG16 and Resnet18. To analyze the quality of the extracted features, features are visualized through t-SNE.

2 Fashion-MNIST Dataset

Fashion-MNIST is a dataset collected from Zalando's article images, which consists of 60,000 training images and 10,000 test images. All the images are in grayscale style and have the resolution of 28*28. There are 10 classes in total. The samples are shown in Fig. 1.

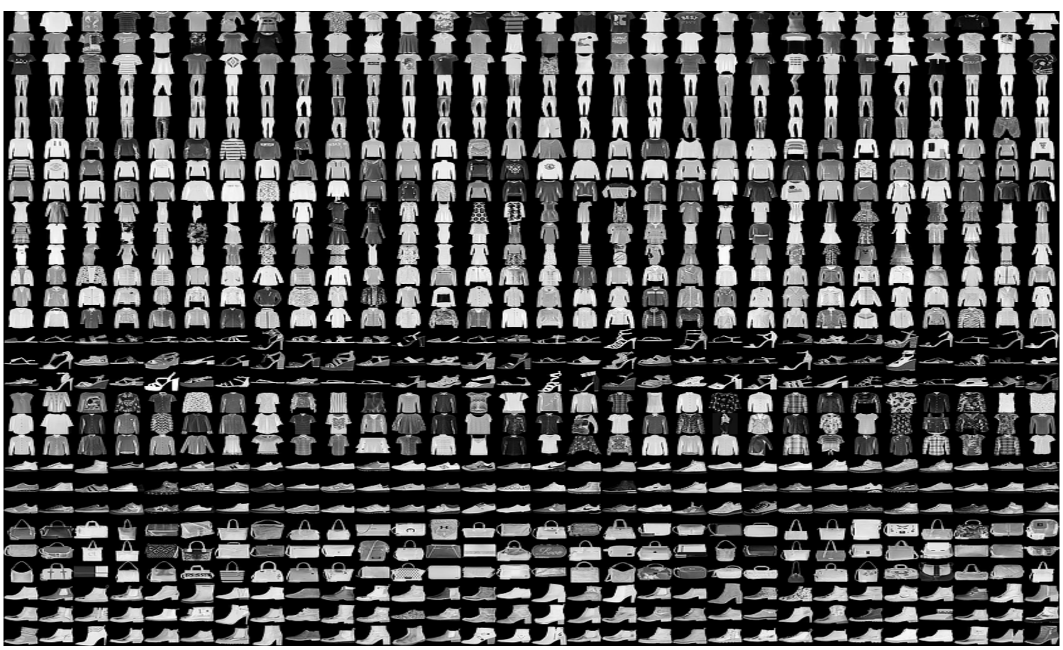

Figure 1: The fashion-mnist datasets.

3 Method

3.1 Random Forest

Random forest is a well-known ensemble learning method for classification, where a set of decision trees are constructed to fit on the various sub-samples of the dataset and averaging is adopted among all the decision tress to give the final prediction, reducing the problem of over-fitting. The input of random forests can be various and influence the performance of random forest. In this study, we explores the original images as well as transformed features as inputs.

3.2 Wavelet Scattering Net

A scattering transform is a non-linear signal representation that builds invariance to geometric transformations while preserving a high degree of discriminability. These transforms can be made invariant to translations, rotations (for 2D or 3D signals), frequency shifting (for 1D signals), or changes of scale. These transformations are often irrelevant to many classification and regression tasks, so representing signals using their scattering transform reduces unnecessary variability while capturing structure needed for a given task. This reduced variability simplifies the building of models, especially given small training sets.

The scattering transform is defined as a complex-valued convolutional neural network whose filters are fixed to be wavelets and the non-linearity is a complex modulus. Each layer is a wavelet transform, which separates the scales of the incoming signal. The wavelet transform is contractive, and so is the complex modulus, so the whole network is contractive. The result is a reduction of variance and a stability to additive noise. The separation of scales by wavelets also enables stability to deformation of the original signal. These properties make the scattering transform well-suited for representing structured signals such as natural images, textures, audio recordings, biomedical signals, or molecular density functions.

3.3 Transfer learning

Transfer learning is a process to apply an existing model to new and relevant tasks on the premise that certain additional data is obtained. As we "migrate" these learned features, we don't need to train a neural network from scratch. Based on the data set size and the similarity of the data between the source and target data sets, there are different ways to use the pre-trained model. (1)Feature extraction: We can use the pre-trained model as a feature extraction device. The specific method is to remove the output layer, and then use the remaining entire network as a fixed feature extraction machine to apply to the new data set.(2)Structure with pre-trained model: We can also use the structure of the pre-training model, but first randomize all the weights, and then train based on our own data set. (3)Train specific layers, freeze other layers: Another way to use a pre-trained model is to partially train it. The specific method is to keep the weights of some initial layers of the model unchanged, and retrain the subsequent layers to obtain new weights. In this process, we can try many times to find the best match between frozen layers and retrain layers based on the results. We also use two of the famous CNN architectures VGG-16 and ResNet-18 to do the finetune process.

4 Experiments

4.1 Feature Extraction

We use four different methods for feature extraction on Fashion-MNIST dataset: raw image, wavelet scattering net, VGG16 pretrained on ImageNet and Resnet18 pretrained on ImageNet. For wavelet scattering net, the length of it is 2. The number of dilations and rotations are set to 6 and 3 respectively. For VGG16 and Resnet18, we extract the feature maps just before the fully-connected layer. The extracted features are then fed into random forests to predict the class labels of the images.

The classification accuracy for different methods on Fashion-MNIST dataset are shown in Tab. 1. Results of pre-trained models are slightly better than simply using raw images and features from scattering network, because VGG16 and Resnet are two more well-designed networks than wavelet scattering network. Besides, there are larger number of parameters in VGG16 and Resnet18 compared with those transformations in scattering network, therefore, VGG16 and Resnet18 can model more complex and adaptive feature extraction functions that describe the images better.

Table 1: Classification accuracy on Fashion-MNIST test dataset. t is the number of decision trees used in the random forest.

	Raw image	Scattering	VGG16	Resnet18
t=50	0.8719	0.8714	0.8834	0.8808
t=100	0.8749	0.8740	0.8896	0.8849
t=200	0.8762	0.8792	0.8924	0.8872

4.2 Feature Visualization

4.2.1 TSNE for Scattering Net Features

Firstly we utilize 2-component t-SNE for the visualization of extracted feature maps of all of 60000 images, which takes about an hour to render. For simplicity, we then visualize the first 2000 images which has the similar data distribution shape although it is much sparse. The hyper parameters for t-SNE are perplexity=30, random state=42.

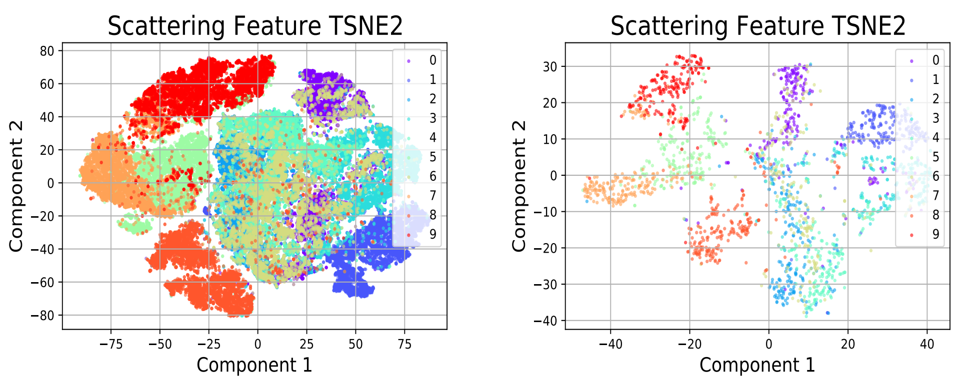

Figure 2: TSNE 2 component visualization for Scattering Net Features. The left one is visualization of all the 60000 feature vectors and the right one is that of first 2000 feature vectors.

4.2.2 TSNE of fine-tuned feature map

Next we visualize extracted feature maps of the same sub dataset for fine-tuned VGG-16 and ResNet-18, the visualized feature of VGG-16 show similar results with Scattering Net, while ResNet-18 separates each class the results well in the feature space.

4.3 Transfer Learning

We test the transferability of VGG16 and Resnet18. VGG16 and Resnet19 are first pre-trained on Imagenet. Then the parameters of the whole network are fixed and a trainable fully-connected layer is added so that the number of output units match the number of categories in Fashion-MNIST dataset. The classification accuracy of pre-trained VGG16 and Resnet18 are shown in Tab. 2. Finally, we fine-tune the whole networks on Fashion-MNIST dataset. The classification results of fine-tuned models reported in Tab. 2 are better than those of pre-trained models. The reason is that there exists a domain gap between ImageNet dataset and Fashion-MNIST dataset. The feature extraction function learnt on ImageNet is not suitable to directly apply to Fashion-MNIST dataset. After fine-tuning, the model is able to learn the dataset-specific features that result in the best classification performance.

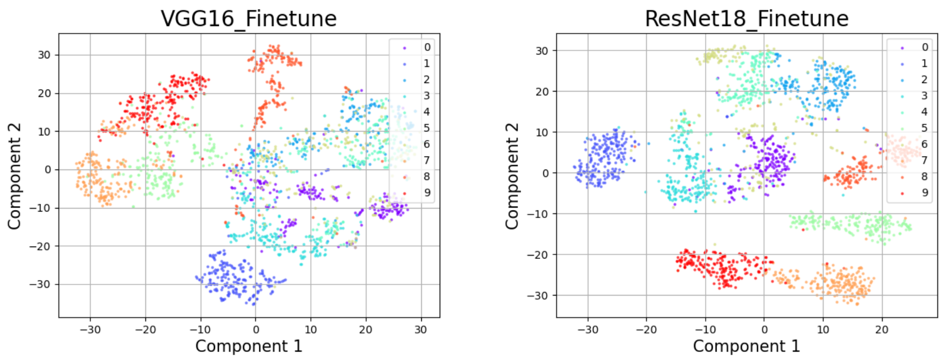

Figure 3: TSNE 2 component visualization for fine-tuned model

Table 2: Classification accuracy on Fashion-MNIST test dataset. P denotes the pre-trained model and FT denotes the fine-tuned model.

	VGG16(P)	VGG16(FT)	Resnet18(P)	Resnet18(FT)
Acc	0.9203	0.9338	0.9448	0.9454

5 Conclusion

In this project, we compare the classification performance on fashion-mnist dataset of three popular classification methods: the traditional Random Forest, Scattering Net, and pre-trained neural neworks of VGG-16 and ResNet-18. In order to verify the results of such algorithms, we use t-SNE to visualize the learned feature in subspace. Besides, the quantitative evaluation and statics analysis is conducted. From the results, we can see that all these are all effective feature extraction methods for this dataset. Meanwhile, the fine-tuned neural networks obtains the best performance which prove the strong learning ability of deep learning methods.

6 Acknowledgement

Chen Junming: Scattering Net implementation, scattering feature extraction, feature visualization and feature statistics analysis.
Shi Zifan: Exploring and implementing feature extraction using scattering network and pre-trained models, and the afterward Random Forest classification.
Gao Rongrong: Implementing fine-tune based on pre-trained deep neural networksVGG-16 and ResNet-18for feature extraction, transfer learning and feature visualization.

7 Reference

[1] J. Bruna and S. Mallat, ”Invariant scattering convolution networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, No. 8, pp. 1872-1886, 2013.
[2] M. Shaha and M. Pawar, ”Transfer Learning for Image Classification,” 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, 2018, pp. 656-660.