MATH6380P Project-1: Feature Extraction and Transfer Learning

Kai Wang, Weizhen Ding

Department of Computer Science and Engineering

Introduction

Feature Extraction is the process of transforming raw data into features that better represent the underlying problem to the desired models, resulting in improved model accuracy on unseen data. A wide range of algorithms has been used for feature extraction in image processing. In this project, we employ the transfer learning method by utilizing the pre-trained VGG16, ResNet networks, as well as the scattering network to perform feature extraction. To compare the extracted representation from the different networks, we first visualize the features by mapping them into two dimensions, then conduct image classification based on the extracted features with various methods. The experiment demonstrates that the scattering network is better separated in visualization and has higher accuracy than VGG16 and ResNet.

Dataset

MNIST dataset is the collection of images of handwritten digits that is widely used in the field of machine learning. All the images are in greyscale with resolution of 28x28 pixel. In the original MNIST dataset, there are 60,000 training images and 10,000 test images.

Feature Extraction

Scattering Network: Scattering network computes the invariants to translation, rotations, scaling, and deformations, while preserving discriminative information. Unlike the standard convolution network, the filters are not learned but are fixed scaled and rotated wavelets[1]. With a scattering network, we can extract the essential properties for classification. In the project, we set the scale J = 2 and max_order = 1 and obtain 833 features for the 60,000 images.

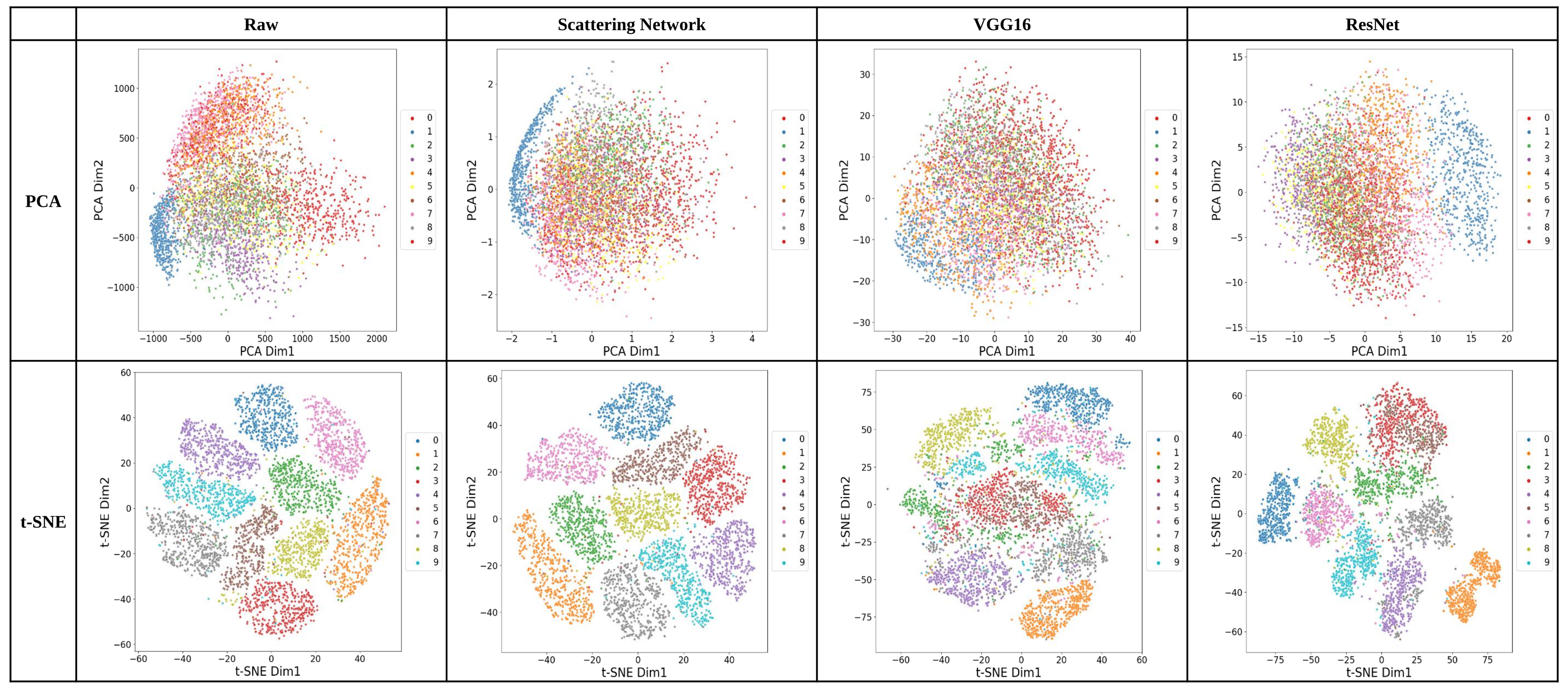
VGG16: VGG simplifies the previous CNN architectures by using a respective small field to replace large ones to reduce the number of parameters[2]. In this project, we select the VGG16, which performs robust but has a smaller number of parameters compared with VGG19. We stop at the FC2 layer and obtain 4096-dimension features for the randomly selected 5,000 samples.

ResNet: ResNet is the architecture designed to solve the vanishing gradient problem suffered by deeper neural networks. ResNet's introduces a so-called "identity shortcut connection" that skips one or more layers[3]. With ResNet architecture, we can now train neural networks with over a thousand layers to achieve decent performance. In the project, we stop at the flattern_2 layer and obtain 2048-dimension features for the randomly selected 5,000 samples.

Feature Visualization

Principal Component Analysis (PCA): PCA is a common exploratory data analysis tool. It constructs new features by combining existing columns linearly. In general, PCA maps the raw data into low-dimensional space, while the variance is maximized in the mapping space.

t-SNE: t-Distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear technique for dimensionality reduction that is particularly well suited for visualizing high-dimensional datasets. t-SNE converts relationships in original space into t-distributions with a small sample size and relatively unknown standard deviations. t-SNE is one of the most popular choices for high-dimensional visualization though the computational cost is very expensive.



Visualization of different feature extractors with PCA and t-SNE

Neural collapse in feature extraction

	Scattering Network	VGG16	ResNet
Std/Avg	1.87E+15	-1.05E+14	9.65E+14
Std/Avg	0.234318	0.200635	0.318246
Std(Cos)	2.04E-05	4.06E-16	1.65E-06
Avg(|Shifted Cos|)	1.11109	1.111111	1.11111

The features are extracted from the pretrained networks without fine-tuning. However, it is still noticeable that 1) the closeness to equal-norms of class-means, 2) equal-angularity, and 3) closeness to maximal-angle equiangularity has very small numbers, which demonstrate the neural collapse phenomenon in the feature extraction.

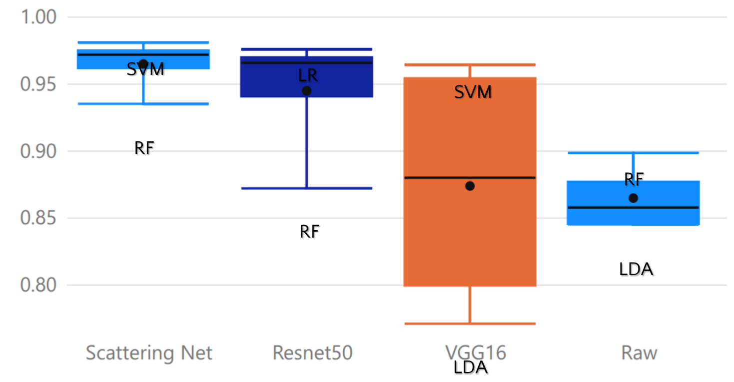
Image Classification

Logistic Regression: Logistic regression is a classification algorithm traditionally limited to only two-class classification problems. It can be extended for multi-class classification, but it is seldom used for this purpose in real-world problems. However, it can become unstable when the classes are well separated.

Linear Discriminant Analysis (LDA): LDA is commonly used for classification, dimension reduction, and data visualization. LDA model estimates the mean and variance from the data for each class and makes predictions by estimating the probability that a new set of inputs belongs to each class. Despite the simplicity, LDA often produces robust, decent, and interpretable classification results.

Support Vector Machine (SVM): A SVM is a discriminative classifier formally defined by a separating hyperplane. The hyperplane (line) is found through the maximum margin, i.e., the maximum distance between both classes' data points. It can solve non-linear and high-dimensional problems well though it is not robust to the outliers.

Random Forest: The random forest is a classification algorithm consisting of many decision trees. The random forest algorithm creates decision trees on data samples and then predicts each of them and finally selects the best solution through voting. It is an ensemble method that is better than a single decision tree because it reduces the over-fitting by averaging the result.



Performance of different feature extractors and classifiers

Computational cost of different feature extractors, visualization methods, and classifiers

	Extraction	Visualization		Classification			
		PCA	t-SNE	LDA	Logistic Regression	SVM	Random Forest
Scattering Network	11.33	1.10	3613.92	4.64	3.07	25.67	3.77
VGG16	888.72	0.56	185.43	47.71	6.17	23.11	1.68
ResNet50	790.85	0.43	97.87	16.91	4.30	17.65	1.68

Discussion & Conclusion

The feature extracted from the networks has very high dimensions. To compare the extracted features, we reduce features' dimensionality into two dimensions and visualize the samples. We can see the PCA provides minimal information for all the models. On the contrary, t-SNE separates the features very well in the scatterplots. In the t-SNE figures, the scattering network performs the best that separates the samples clearly and evenly. ResNet also performs well though it struggles to distinguish "3" and "5". VGG16 improves the raw data significantly, but it has noticeable overlapping with several digits.

We also conduct cross-validation by employing various image classification methods with the extracted features. Based on the experiment result, the scattering network has the highest accuracy, and it performs robustly with all the classifiers. The accuracy of ResNet is close to the scattering network despite it has a much lower minimum. VGG16 has the worst performance, and it is sensitive to classifiers. It is also worth noting that SVM provides the best accuracy for most models, while random forest and LDA have relatively low performance.

Moreover, in terms of efficiency, the scattering network, when setting scale as 2 and max order as 1, is much faster in feature extraction than the other two competitors. Besides, t-SNE runs much slower than PCA, although it performs much better in exhibiting the patterns. The difference among the classifiers is not significant in our project.

Overall, the scattering network performs better over the other two models. We can further investigate parameter settings' effect to feature extraction in image classification in the future study.

Contribution

Kai Wang conducted the Scattering network and VGG16 model and the relevant visualization and classification, and also the editing the poster of introduction, dataset, and feature extraction part.

Weizhen Ding conducted the ResNet feature extraction and other relevant visualization and classification, and also edit the poster of the introduction, image classification and conclusion part.

Acknowledgements

[1] J. Bruna and S. Mallat, "Invariant scattering convolution networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 8, pp. 1872–1886, 2013, doi: 10.1109/TPAMI.2012.230.

[2] K. Simonyan and A. Zisserman, "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION," 2015. Accessed: Oct. 02, 2020.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Dec. 2016, vol. 2016-December, pp. 770–778, doi: 10.1109/CVPR.2016.90.