Math 6380O Project 1
Feature Extraction and Transfer Learning

Chunyan BAI, Haoye CAI, Yuan CHEN, Wenshuo GUO

Department of Computer Science and Engineering

Hong Kong University of Science and Technology

Introduction

- A warm-up aims to explore feature extractions using existing networks

- The existing networks used in the project include: pre-trained deep neural networks and scattering nets

- Image classifications with traditional machine learning methods

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Introduction

- **2** datasets to perform classification task - Raphael's dataset and MNIST's dataset
- **2** pre-trained deep neural networks and scattering nets
- **6** visualization approaches: PCA, MDS and t-SNE etc. to visualize the extracted features
- **4** traditional supervised learning methods implemented based on the extracted features, including LDA, logistic regression, SVM, random forests to classify images

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Related Work: Scattering Network

- We hereby adopt the improved version of scattering network as proposed by Oyallon et al [1].

- This invariant scattering network utilized two wavelet transformation to build a non-trivial invariant and retain both high-frequency and low-frequency domain.

Related Work: Scattering Network

- To deal with images, we apply scattering transformation on each RGB channel of the image, resulting in a final channel of $3\times(1+JL+12J(J-1)L2)3\times(1+JL+12J(J-1)L2)$, and the image is downsampled by a factor of 2^J, where J is the spatial scale of scattering transform and L is an integer parametrizing a discretization of $[0,2\pi]$.

- We use scattering network to explore its efficiency compared with modern data-driven neural networks.

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Related Work: VGG19 and ResNet50

- Both VGG [3] and ResNet [4] are convolutional neural networks designed to perform image recognition. Their success on ImageNet [5] dataset has proven their efficacy in extracting meaningful feature from general images.

- Therefore we choose VGG19 and ResNet50 pre-trained on ImageNet as the base neural network for feature extraction in this project.

Related Work: VGG19 and ResNet50

- These two networks also make good comparisons against each other

- VGG19 is considered shallow but may preserve more details in its feature map, while ResNet50 is much deeper and the final feature consists of high-level visual cues.

Dataset 1: Raphael's paintings identification

- A small dataset of Raphael's/forgeries paintings. There are 28 images in total.

- The idea of whether extracted feature works on such a small dataset would further illustrates the difference/connection of various methodologies. There has been previous study on authenticating Van Gogh paintings by Liu et al [2].

- We resize all the images to 224*224, and use transfer learning to train classifiers to identify forgeries. We use leave-one-out validation as the dataset is of limited size.

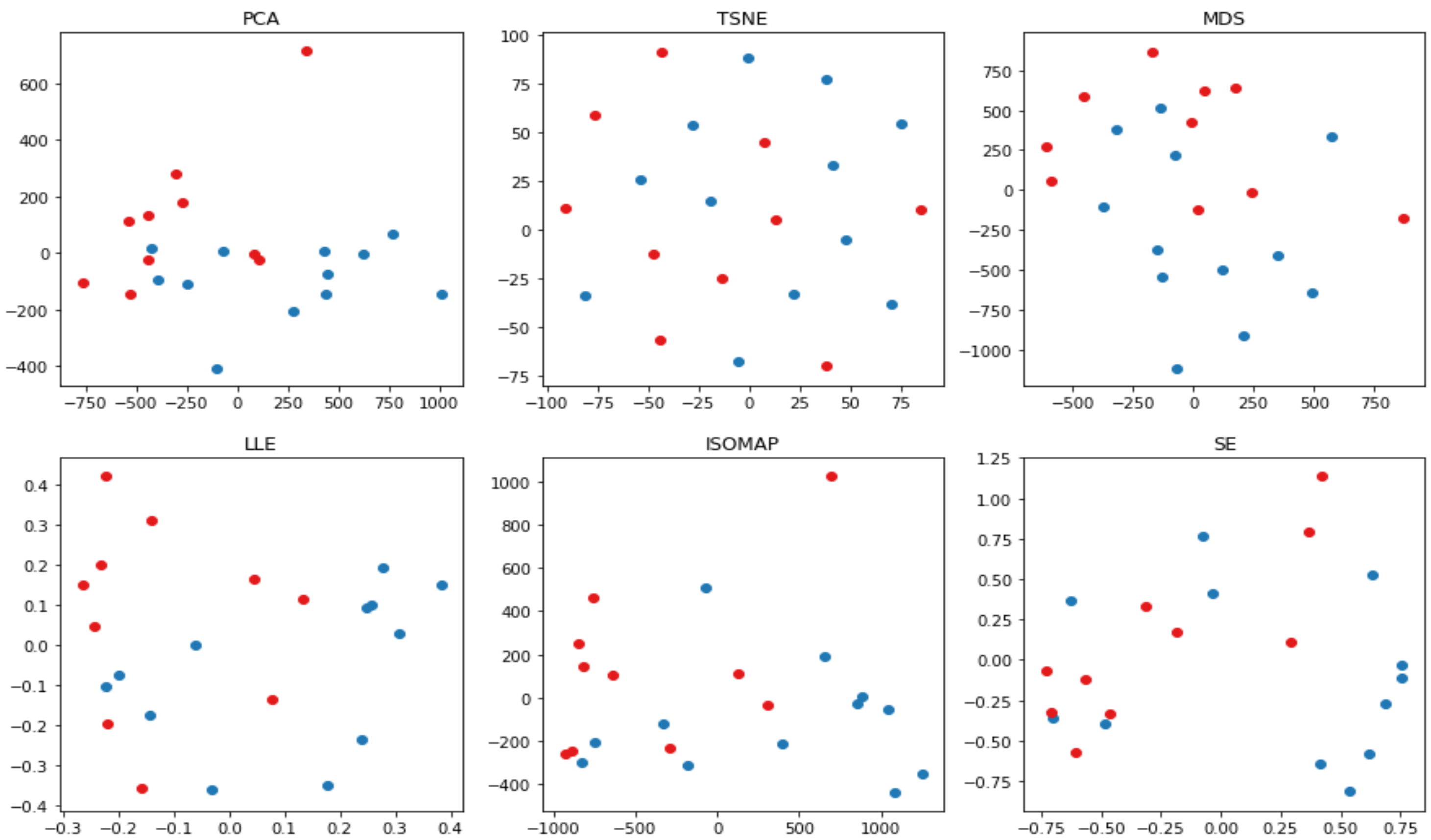
香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Dataset 2: MNIST

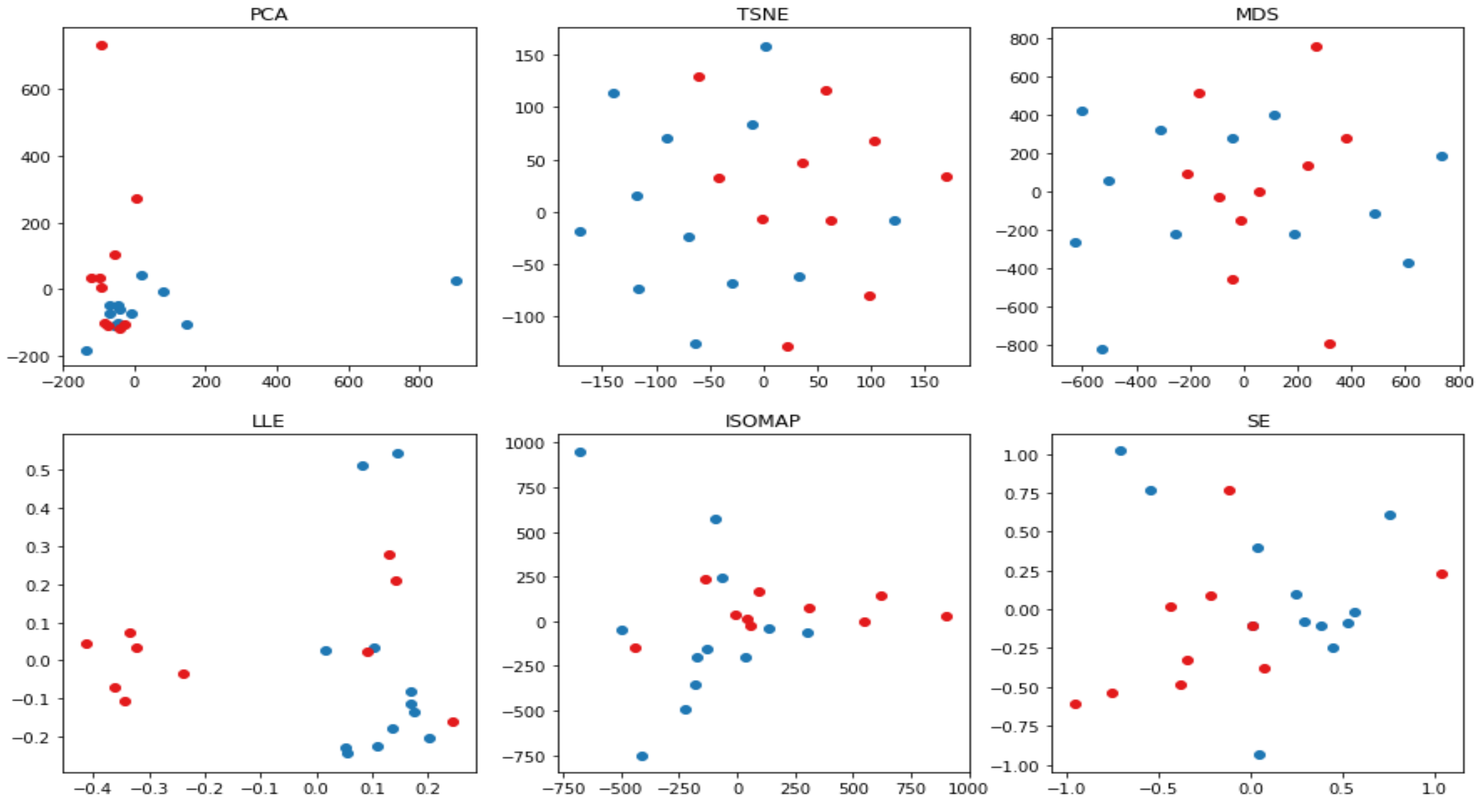
- MNIST is a simple yet effective dataset to explore the learnt feature of networks, given it is made up of numerous handwritten digits.
- It consists of 60000 gray-scale training images of size 2828 and 10 classes.
- In our experiment we resize all the images to 224224, and choose only a subset of the train/test images due to time concern.

Result: Dataset 1 Raphael's Painting

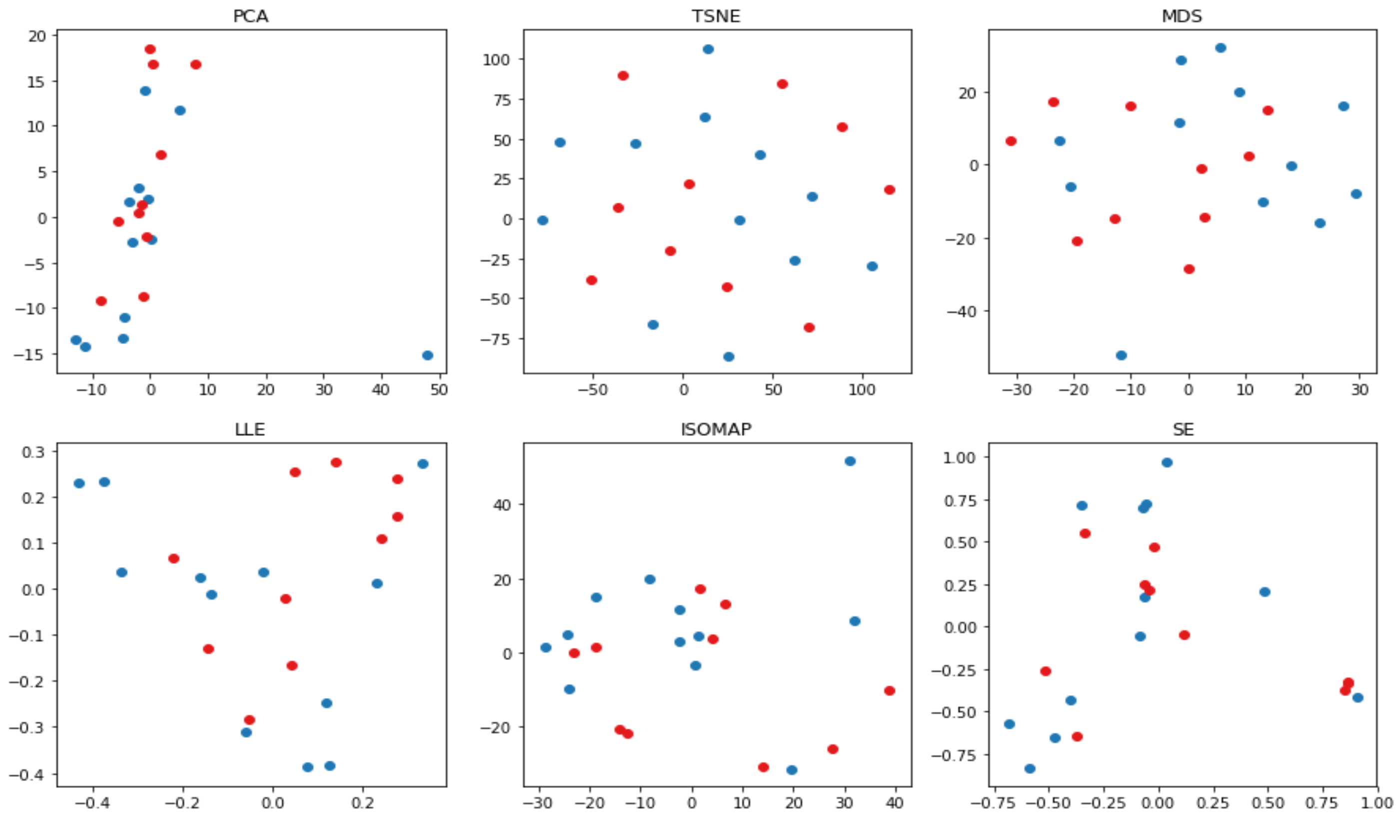
Results: Visualization of Scattering Network



Results: Visualization of VGG19 Network



Results: Visualization of ResNet50 Network



Visualization Summary: Raphael Dataset Result

Deep Learning Networks Comparison through the Same Visualisation

- The scattering network tends to separate the features into two parts.
- VGG19 tends to project most features in a relative small space
- ResNet50 tends to preserve more linear structures
- Three networks perform differently in feature extraction, which may due to their different architectures.

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Visualization Summary: Raphael Dataset Result

Different Visualisation Approaches through the Same Network

- PCA is a linear transformation of the dimensions, and often miss important non-linear structure in the data.
- t-SNE seems to have the best performance among all six methods. The clusters are obvious than those in other figures and the points are projected almost uniformly in the space.

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Results: Scattering Net - Classification on the extracted features

Classification evaluation function: Outputs are cross entropy (negative log likelihood) and accuracy. The number before and after the '|' is the mean and the standard deviation respectively.

```
1  classification_raphael(scat_feature)
```

```
SVC: bce: 0.606957081328 | 0.329827523921, acc: 0.590909090909 | 0.491666083018
LinearDiscriminantAnalysis: bce: 6.67414304328 | 13.1406707877, acc: 0.590909090909 | 0.491666083018
RandomForestClassifier: bce: 0.647441553212 | 0.322877862044, acc: 0.681818181818 | 0.465770489362
LogisticRegression: bce: 0.333787774529 | 0.752503497757, acc: 0.863636363636 | 0.343174292512
```

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Results: Scattering Net - Classification on the extracted features

For the features extracted by the Scattering Network, Logistic Regression has the lowest cross entropy loss and the highest accuracy, which indicates the best classification performance.

When taking standard deviation into consideration, however, since all the standard deviations are relatively large, especially compared to the means, we find that all methods are actually similar (at least similarly unstable). This is understandable because as we can see in the visualizations, the extracted features cannot be easily linearly separated, and positive and negative features tend to mix together.

Results: VGG19 - Classification on the extracted features

Classification evaluation function: Outputs are cross entropy (negative log likelihood) and accuracy. The number before and after the '|' is the mean and the standard deviation respectively.

```
1 classification_raphael(vgg19_feature)
```

```
SVC: bce: 0.732629443491 | 0.0830013836569, acc: 0.545454545455 | 0.497929597732
LinearDiscriminantAnalysis: bce: 0.920833405594 | 0.433443803941, acc: 0.363636363636 | 0.481045692921
RandomForestClassifier: bce: 0.706377798049 | 0.224337240833, acc: 0.5 | 0.5
LogisticRegression: bce: 1.06694353663 | 1.59343385384, acc: 0.681818181818 | 0.465770489362
```

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Results: VGG19 - Classification on the extracted features

For the features extracted by the VGG19 Network, Logistic Regression still has the highest accuracy but Random Forest and SVM show the lowest cross entropy loss.

Therefore we conclude for these features, the three classifiers mentioned indicate reasonably good performance. The bad performance shown by LDA may indicate this feature distribution violates the LDA gaussian assumptions.

Results: ResNet50 - Classification on the extracted features

Classification evaluation function: Outputs are cross entropy (negative log likelihood) and accuracy. The number before and after the '|' is the mean and the standard deviation respectively.

```
1  classification_raphael(ResNet50_feature)
```

```
SVC: bce: 0.848902438299 | 0.199775666422, acc: 0.545454545455 | 0.497929597732
LinearDiscriminantAnalysis: bce: 1.75302978976 | 4.2157984159, acc: 0.5 | 0.5
RandomForestClassifier: bce: 0.664865473447 | 0.235994520728, acc: 0.590909090909 | 0.491666083018
LogisticRegression: bce: 0.723769099988 | 0.81084204489, acc: 0.636363636364 | 0.481045692921
```

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Results: ResNet50 - Classification on the extracted features

From the statistics, for the features extracted by the ResNet50 Network, Logistic Regression still has the highest accuracy but the difference seems to further shrink, i.e. four methods have similar accuracy and cross entropy (except for LDA which has larger cross entropy).

The bad/unstable output distribution shown by LDA may indicate this feature distribution violates the LDA gaussian assumptions. But more importantly, as shown in the visualization, the features from different classes seem to further mix together, making classifiers more difficult to distinguish. This may explain the not-so-good results shown by all methods.

Classification Summary: Raphael Dataset Result

	SVM	LDA	RandomForest	Logistic Regression
ScatNet (CE)	0.61 ± 0.33	6.67 ± 13.14	0.65 ± 0.32	0.33 ± 0.75
ScatNet (Acc)	0.59 ± 0.49	0.59 ± 0.49	0.68 ± 0.47	0.86 ± 0.34
VGG19 (CE)	0.73 ± 0.08	0.92 ± 0.43	0.71 ± 0.22	1.07 ± 1.59
VGG19 (Acc)	0.55 ± 0.50	0.36 ± 0.48	0.50 ± 0.50	0.68 ± 0.47
ResNet50 (CE)	0.85 ± 0.20	1.75 ± 4.20	0.66 ± 0.24	0.72 ± 0.81
ResNet50 (Acc)	0.55 ± 0.50	0.50 ± 0.50	0.59 ± 0.49	0.63 ± 0.48

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Classification Summary: Raphael Dataset Result

- Logistic Regression achieve best accuracy/log loss for all types of features; LDA almost always perform the worst

- The first implication: this Raphael Drawing dataset has a data distribution that is very different from the LDA gaussian assumption. This may serve as the main reason for the bad performance.

- The next important observation: the simplest method seems to be more suitable for this dataset. More complicated methods like SVM have high potential to overfit, and simpler methods have better chance to make good decisions in the test set.

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Which features are more robust or more suitable for this dataset?

- As shown by the numbers, no matter for the best or the average performance by the classifiers, Scattering Network seems to provide more reasonable features.

- This may be explained by the ability inherent in the wavelet operations to process/find fixed patterns, which may enforce its analysis on the textures of drawings, thus leading to easier thresholds between classes.

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Which features are more robust or more suitable for this dataset?

- For the two deep nets, since they are trained on large scale datasets, they are more likely to find universal patterns which may not be helpful for this specific task.

- In addition, since ResNet50 has much more layers than VGG19 (increased depth), it may provide much more complicated and high-level features in deeper layers.

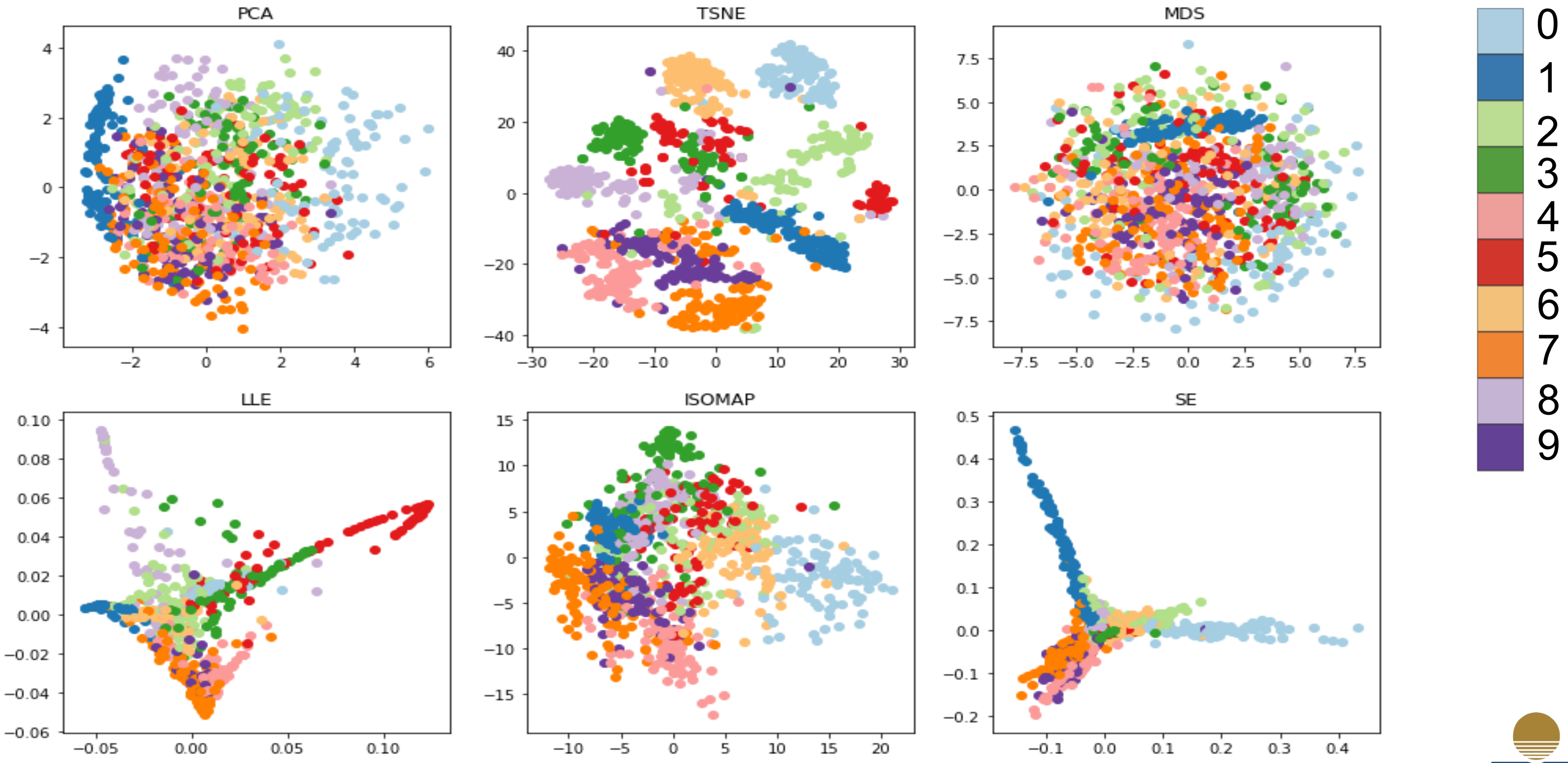
- This may be the reason why ResNet features seem to be more mixed.

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Result: Dataset 2 MNIST

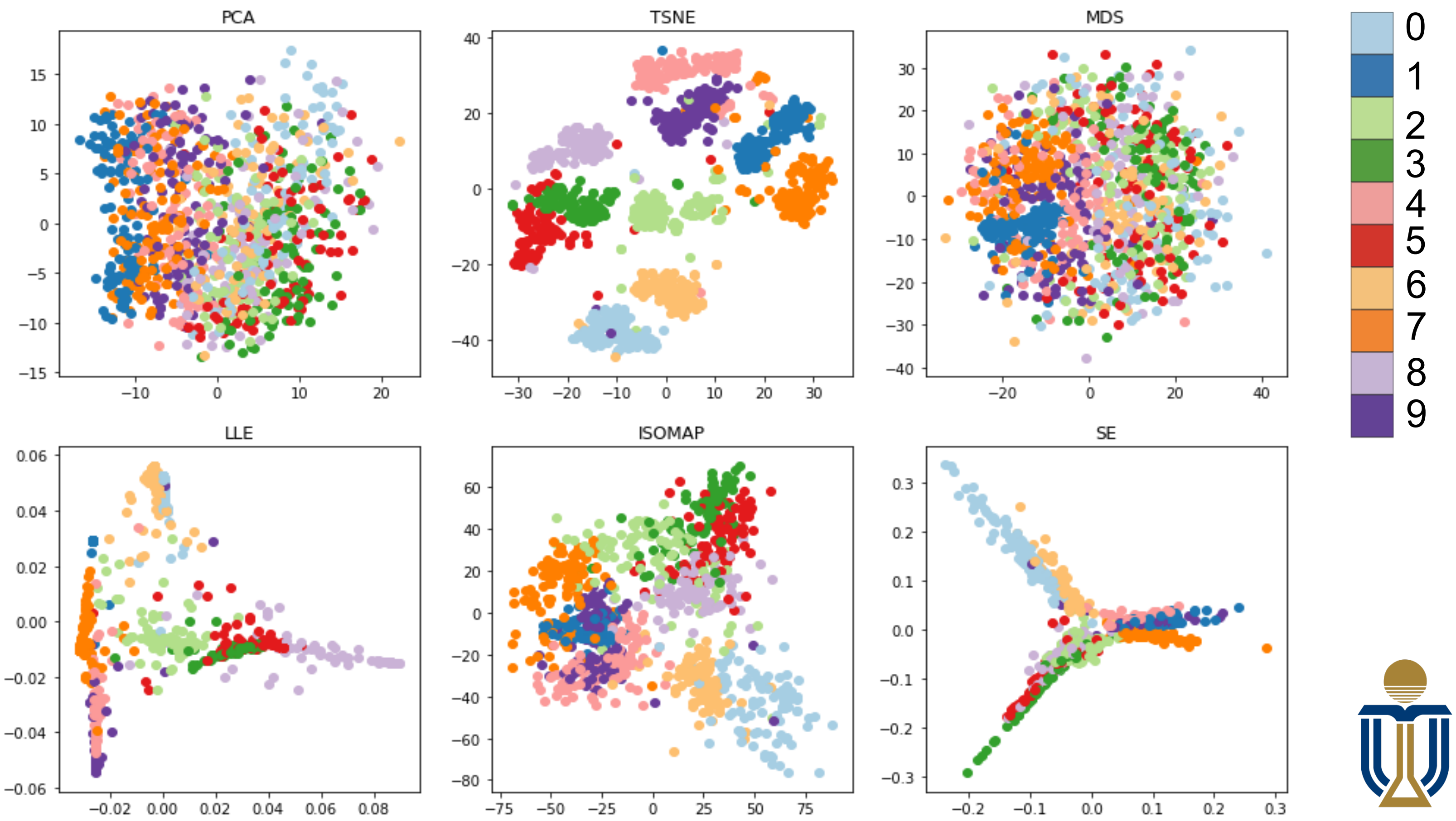
香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Results: Visualization of Scattering Network

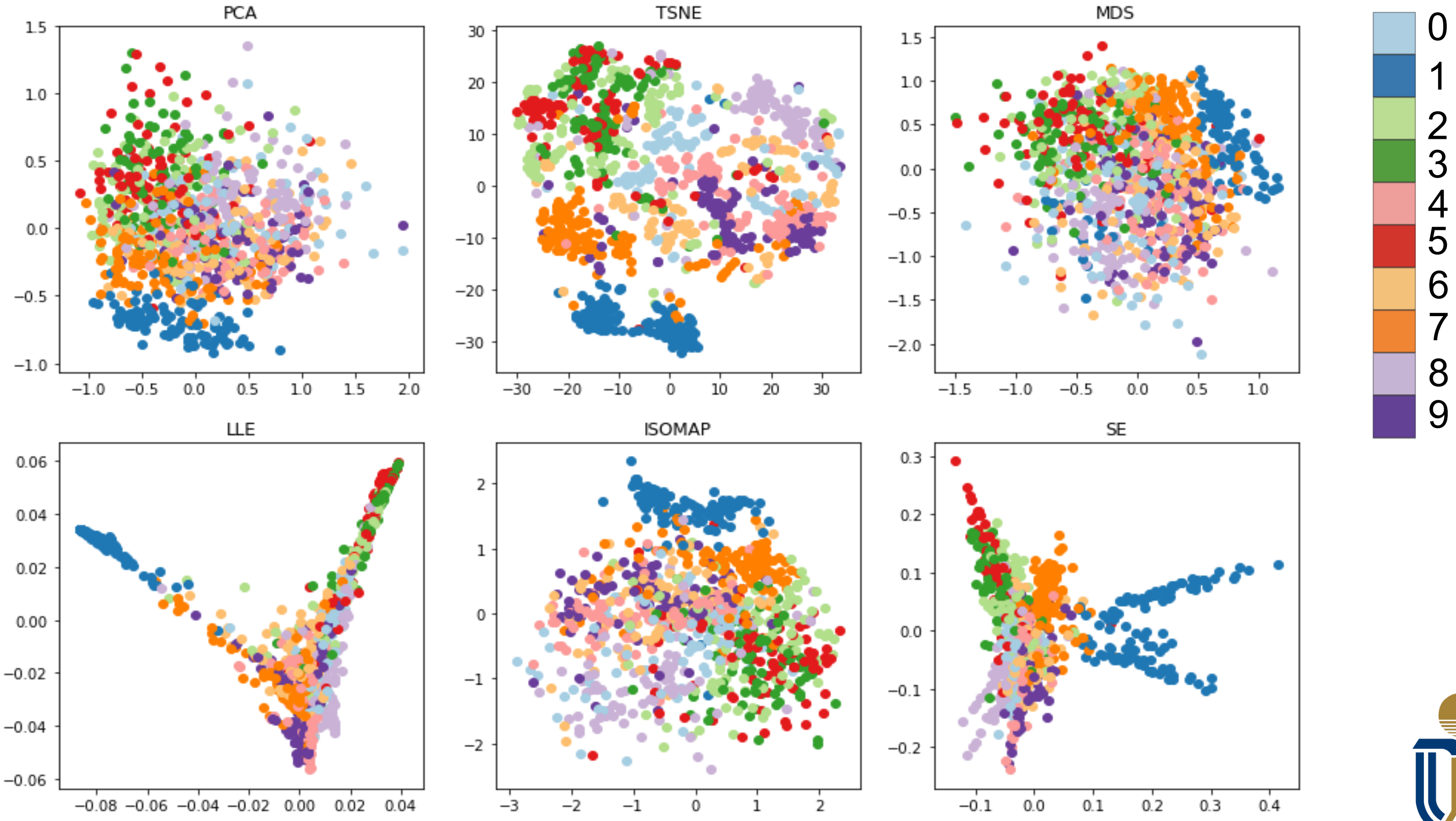


香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Results: Visualization of VGG19 Network



Results: Visualization of ResNet50 Network



香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Visualization Summary: MNIST

Deep Learning Networks Comparison through the Same Visualisation

- The features extracted by Scattering Network and VGG19 and dimensionally reduced by these visualisation approaches are presented better than those of ResNet50.

Visualization Summary: MNIST

Different Visualisation Approaches through the Same Network

- Non-linear structure in the data is very significant since PCA which tries to find the linear transformation among dimensions and MDS which measures the distance similarity between objects seem to mix clusters, which means some features may look very similar even though they are from different groups in the dataset.

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Results: Scattering Net - Classification on the extracted features

Classification evaluation function: Outputs are cross entropy (negative log likelihood) and accuracy. The number before and after the '|' is the mean and the standard deviation respectively.

```
1 classification_mnist(scat_feature1, y_train, scat_feature2, y_test)
```

```
SVC: ce: 1.52501802048 | 0.134726171697, acc: 0.18 | 0.107703296143
LinearDiscriminantAnalysis: ce: 0.417484625077 | 1.11500235588, acc: 0.98 | 0.06
RandomForestClassifier: ce: 0.519676359259 | 0.133950668986, acc: 0.94 | 0.0916515138991
LogisticRegression: ce: 0.223123390989 | 0.120358834566, acc: 0.98 | 0.06
```

Results: Scattering Net - Classification on the extracted features

For the features extracted from MNIST by the Scattering Network, Logistic Regression still has the lowest cross entropy loss and the highest accuracy, which indicates the best classification performance.

Linear Discriminant Analysis also has similar level of performance while non-linear classifier like SVM and Random Forest performed suboptimally, we can see that a linear decision boundary is already sufficient for MNIST features. When taking standard deviation into consideration, 3 out of 4 has low std, indicating its highly linearity, which is consistent with the visualization.

Results: VGG19 - Classification on the extracted features

Classification evaluation function: Outputs are cross entropy (negative log likelihood) and accuracy. The number before and after the '|' is the mean and the standard deviation respectively.

```
classification_mnist(vgg19_feature_train, y_train, vgg19_feature_test, y_test)
```

```
SVC: ce: 0.805665570568 | 0.218437973267, acc: 0.54 | 0.22
LinearDiscriminantAnalysis: ce: 13.7808244169 | 8.53223598554, acc: 0.244444444444 | 0.226623089493
RandomForestClassifier: ce: 0.584524227972 | 0.175466817887, acc: 0.94 | 0.0916515138991
LogisticRegression: ce: 0.075207969605 | 0.110081354029, acc: 0.98 | 0.06
```

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Results: VGG19 - Classification on the extracted features

For the features extracted from MNIST by VGG19, Logistic Regression still has the lowest cross entropy loss and the highest accuracy, which indicates the best classification performance.

Random Forest also has similar level of performance while other classifier like SVM and LDA performed suboptimally. We can see that the feature from VGG-19 is of high semantic meaning. When taking standard deviation into consideration, 2 out of 4 has low std, indicating its certain non-linearity, which is consistent with the visualization.

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Results: ResNet50 - Classification on the extracted features

Classification evaluation function: Outputs are cross entropy (negative log likelihood) and accuracy. The number before and after the '|' is the mean and the standard deviation respectively.

```
classification_mnist(resnet_feature_train, y_train, resnet_feature_test, y_test)
```

```
SVC: ce: 1.62142374547 | 0.107362974968, acc: 0.18 | 0.107703296143
LinearDiscriminantAnalysis: ce: 7.56229853181 | 6.9493266934, acc: 0.76 | 0.215406592285
RandomForestClassifier: ce: 0.760802879416 | 0.315333514733, acc: 0.86 | 0.128062484749
LogisticRegression: ce: 1.03198858952 | 0.224656799192, acc: 0.8 | 0.126491106407
```

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Results: ResNet50 - Classification on the extracted features

For the features extracted from MNIST by ResNet50, Random Forest has the lowest cross entropy loss and the highest accuracy, which indicates the best classification performance.

Logistic Regression and LDA also has similar level of performance while other classifier like SVM performed suboptimally. We can see that the feature from ResNet50 is of relatively low semantic meaning, resulting in lower accuracy overall. It is possibly due to the deep network structure where most information are thrown away. When taking standard deviation into consideration, most of them are quite high, indicating its not quite linear, which is consistent with the visualization and analysis.

Classification Summary: MNIST

	SVM	LDA	RandomForest	Logistic Regression
ScatNet (CE)	1.53 ± 0.13	0.42 ± 1.12	0.52 ± 0.13	0.22 ± 0.12
ScatNet (Acc)	0.18 ± 0.11	0.98 ± 0.06	0.94 ± 0.09	0.98 ± 0.06
VGG19 (CE)	0.81 ± 0.22	13.78 ± 8.53	0.58 ± 0.18	0.08 ± 0.11
VGG19 (Acc)	0.54 ± 0.22	0.24 ± 0.23	0.94 ± 0.09	0.98 ± 0.06
ResNet50 (CE)	1.62 ± 0.11	7.56 ± 6.95	0.76 ± 0.32	1.03 ± 0.22
ResNet50 (Acc)	0.18 ± 0.11	0.76 ± 0.22	0.86 ± 0.13	0.8 ± 0.13

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Classification Summary: MNIST

- Logistic Regression and Random Forest are the best among all results.

- Smaller standard deviation (i.e. variance) compared with the Raphael's Drawing dataset

- By analyzing performance gaps of certain algorithms, we may learn some characteristics about our learned feature distributions. For example, the low accuracy of LDA on VGG19 features may indicate the feature of VGG19 is not combination of gaussian distribution with equal variance.

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Distribution of Work

- Chunyan BAI: Feature extraction by scattering net with known invariants;
- Wenshuo GUO: Feature extraction by pre-trained deep neural networks, e.g. VGG19, and resnet50;
- Yuan CHEN: Visualize these features using classical unsupervised learning methods, e.g. PCA, t-SNE, MDS, LLE, isomap, Spectral Embedding (SE);
- Haoye CAI: Image classifications using traditional supervised learning methods based on the features extracted, e.g. LDA, logistic regression, SVM, random forests;
- All members equally: Report writing. Compare the results obtained and give analysis on explaining the phenomena.

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY

Reference

1. Oyallon, E., Belilovsky, E., & Zagoruyko, S. (2017, October). Scaling the scattering transform: Deep hybrid networks. In International Conference on Computer Vision (ICCV).
2. Liu, H., Chan, R. H., & Yao, Y. (2016). Geometric tight frame based stylometry for art authentication of van Gogh paintings. Applied And Computational Harmonic Analysis, 41(2), 590-602.
3. Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
4. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
5. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on (pp. 248-255). IEEE.

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Thank You!

Q&A

香港科技大學
THE HONG KONG
UNIVERSITY OF SCIENCE
AND TECHNOLOGY