香港浸會大學
HONG KONG BAPTIST UNIVERSITY

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

Feature Extraction and Transfer Learning

-- MATH 6380p Course Project 1

Dong Qian and **Kejing Yin**
{dongqian, cskjyin}@comp.hkbu.edu.hk
Department of Computer Science, Hong Kong Baptist University

Introduction

- **Project Task**
  - Extract Features using Scattering networks and pre-trained DNNs, AlexNet and ResNet-18.
  - Visualize the extracted features using PCA and t-SNE.
  - Perform classification task using logistic regression.

Datasets

- **MNIST**
  - The MNIST dataset contains 60,000 training samples and 10,000 test samples. There are 10 digits (0-9) to predict. Each image is represented by 28 * 28 pixels.
- **CIFAR10**
  - The CIFAR-10 dataset contains 60,000 images divided into 10 classes, with 6000 images per class. There are 50,000 training samples and 10000 test samples. Each image is represented by 32*32 pixels in color with red, green and blue.

Feature (Representation) Extraction

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

In the pixel space, a simple operation can yield a great variation of the value of each pixel.

Transformations such as translations, scaling and rotations are generally uninformative for classification.

A good representation should be invariant to this kind of deformations.

Scattering Network

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

The scattering network learns translation-invariant representations of the images and preserves high frequency.

It cascades wavelet transforms and modulus pooling operators.

Given a band-pass wavelet $\psi(u)$, it is possible to build a bank of multiscale directional wavelets based on scales and orientations.

$$\psi_{2^j r} = 2^{2j}\psi(2^j r^{-1} u)$$

Given a family wavelets $\{\psi_\lambda(u)\}_\lambda$, the convolution operation on the data x with each wavelet is defined as $\{x \star \psi_\lambda\}_\lambda$.

[1] Joan Bruna and Stéphane Mallat. Invariant Scattering Convolution Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1872--1886, 2012.

Scattering Network

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

To guarantee the invariance, the modulus operation is introduced.

Adding a spatial window for local stability

The scattering network architecture is

$$S_J[p]x = ||||x \star \psi_{\lambda_1}| \star \psi_{\lambda_2}|\ldots| \star \psi_{\lambda_m}| \star \phi_{2^J}$$

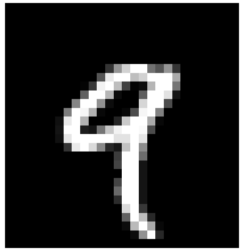
[1] Joan Bruna and Stéphane Mallat. Invariant Scattering Convolution Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 8, pp. 1872--1886, 2012.

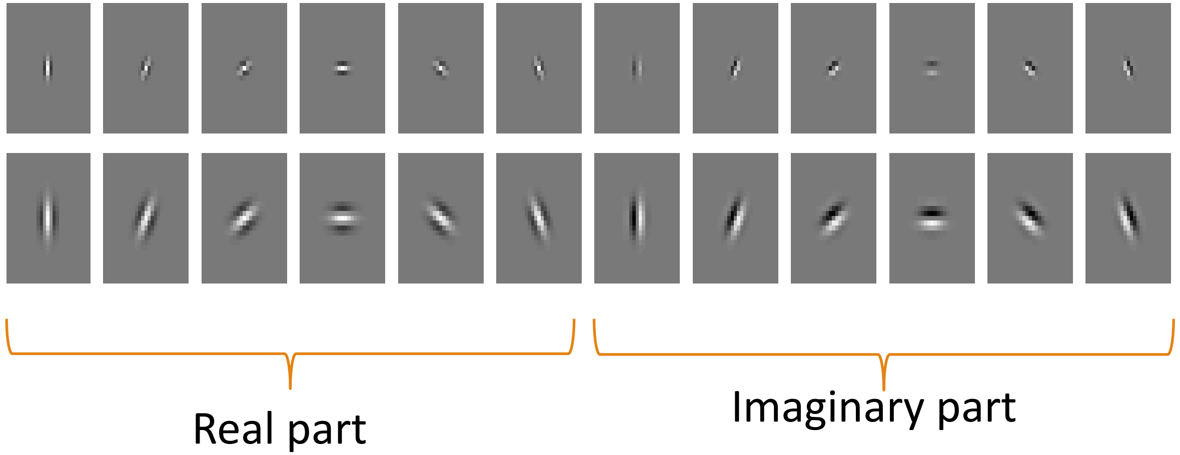
Scattering Network

Parameters in the Scattering network:

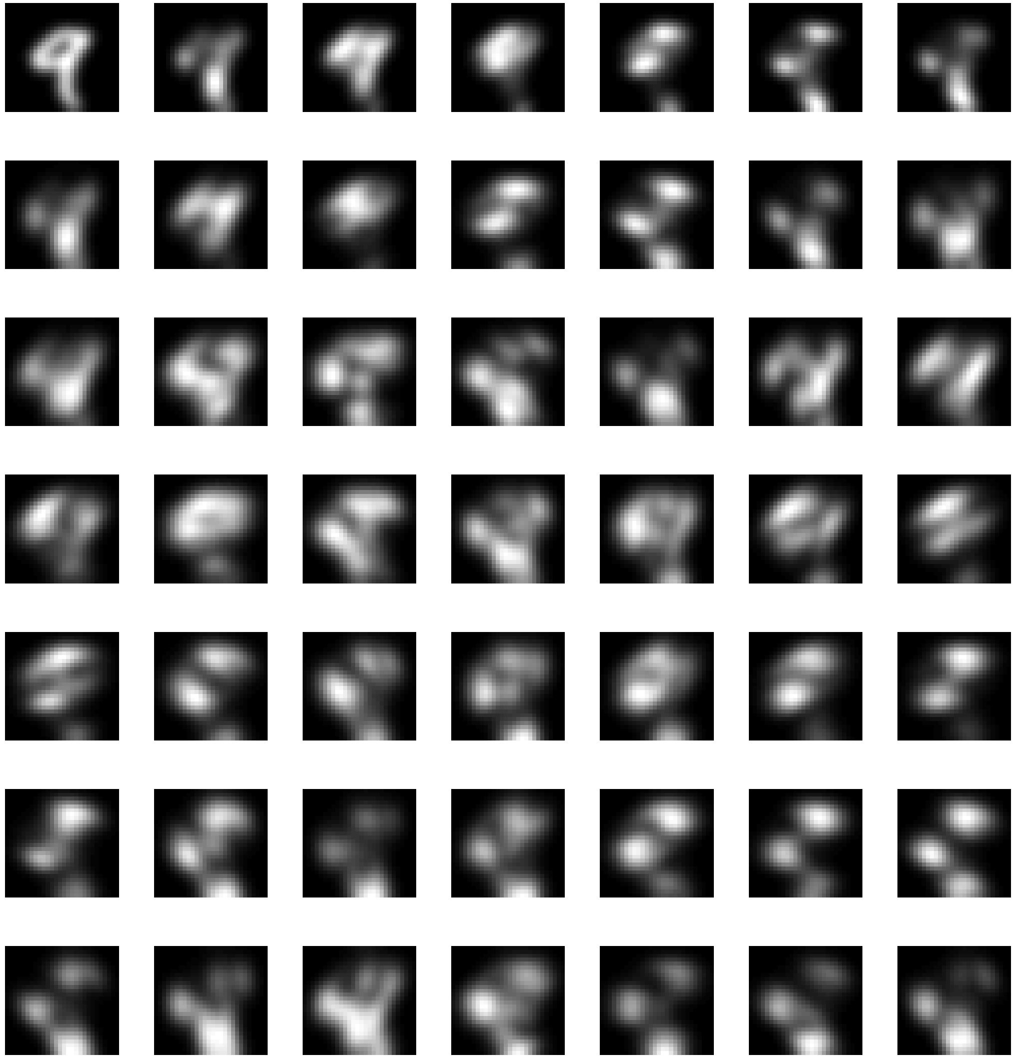
- The number of scale J,
- The number of orientations L and
- The maximum scattering order M

Scattering Network



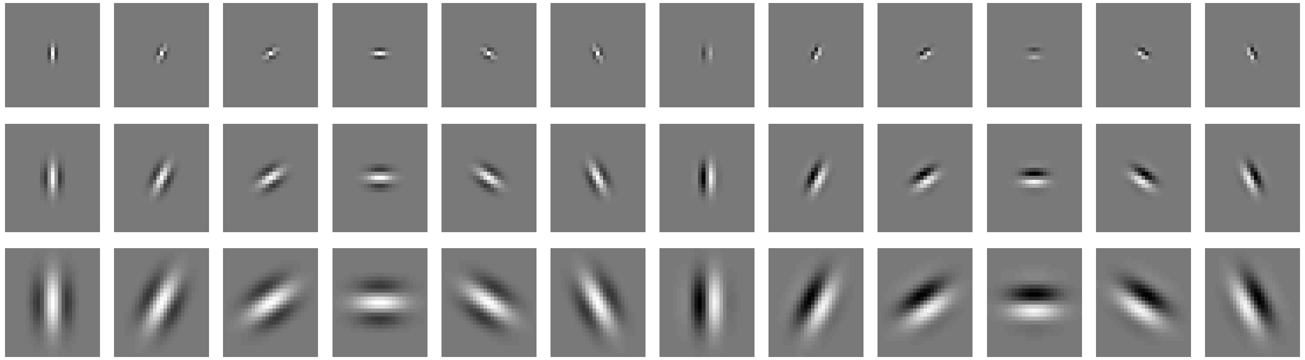


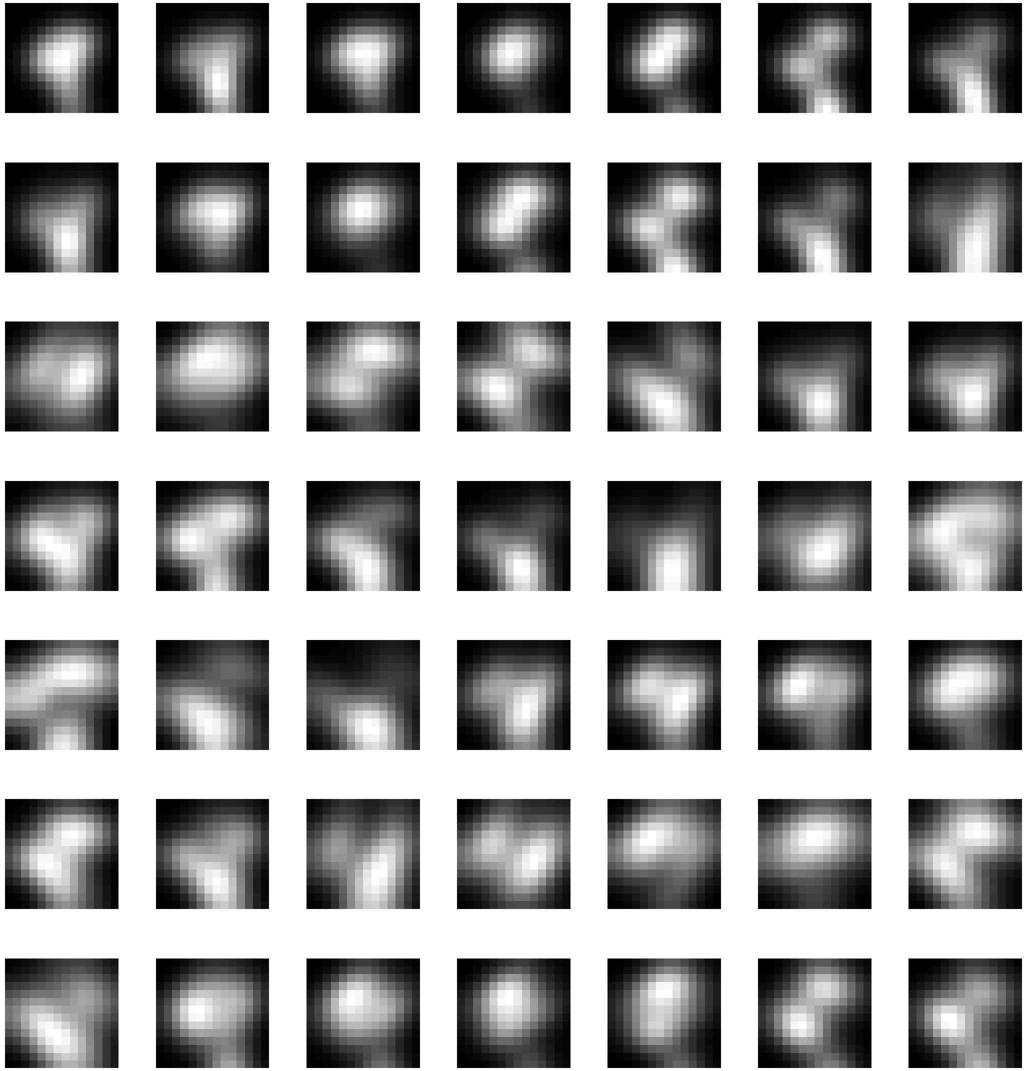
A filter bank of complex Morlet wavelets with 2 scales (J) and 6 orientations (L).



Scattering Network

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系



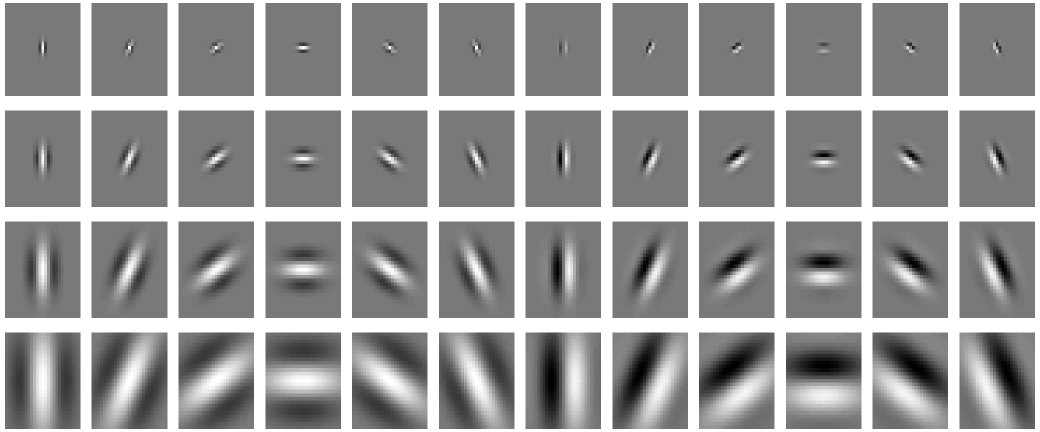


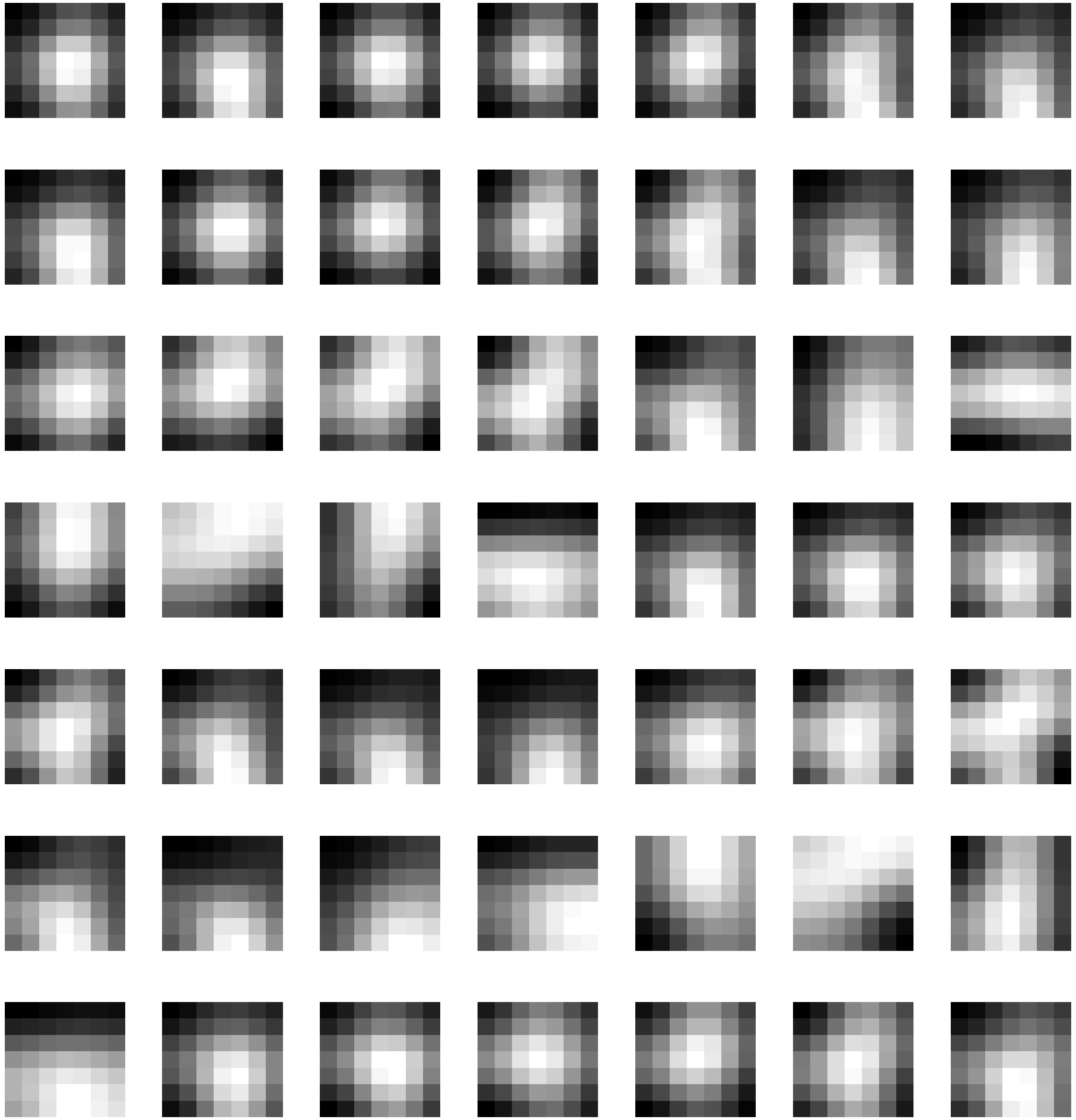
Parameters:

- J = 3
- L = 6
- M = 2

Scattering Network

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

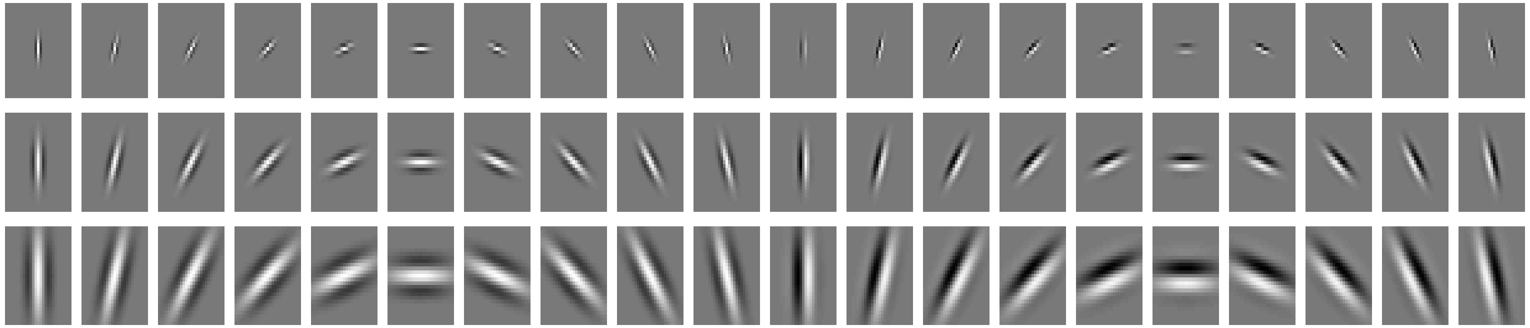


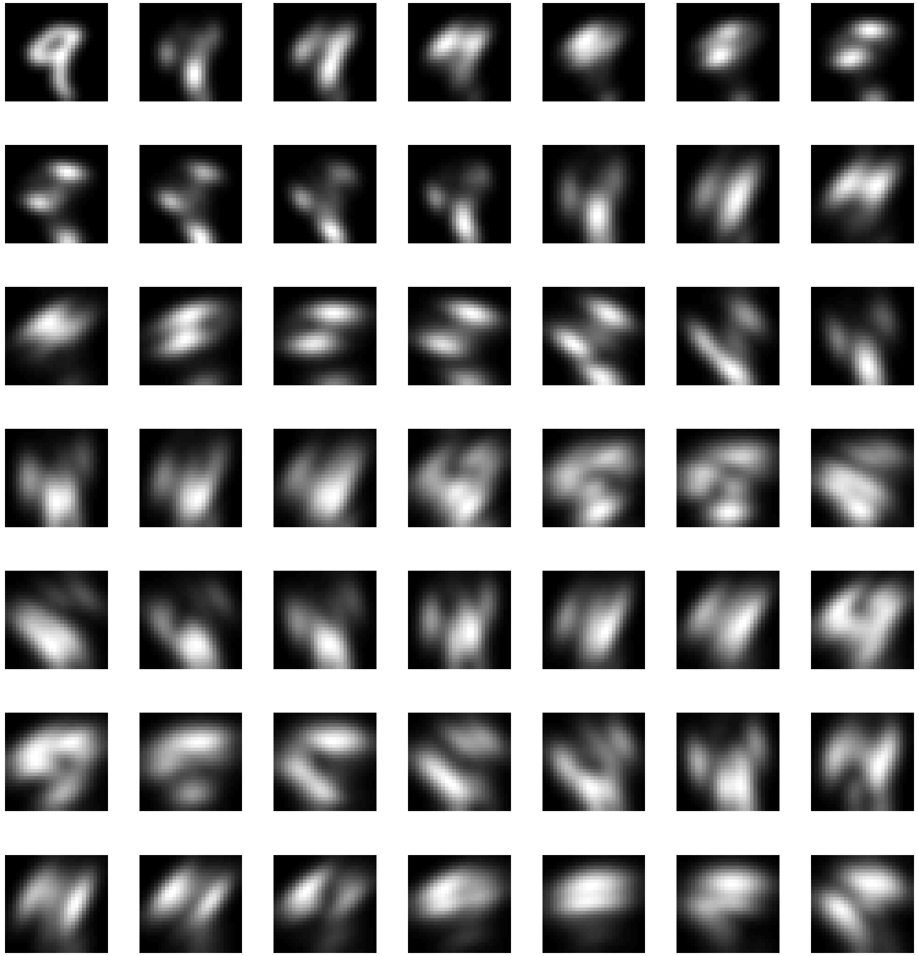


Parameters:

- J = 4
- L = 6
- M = 2

Scattering Network

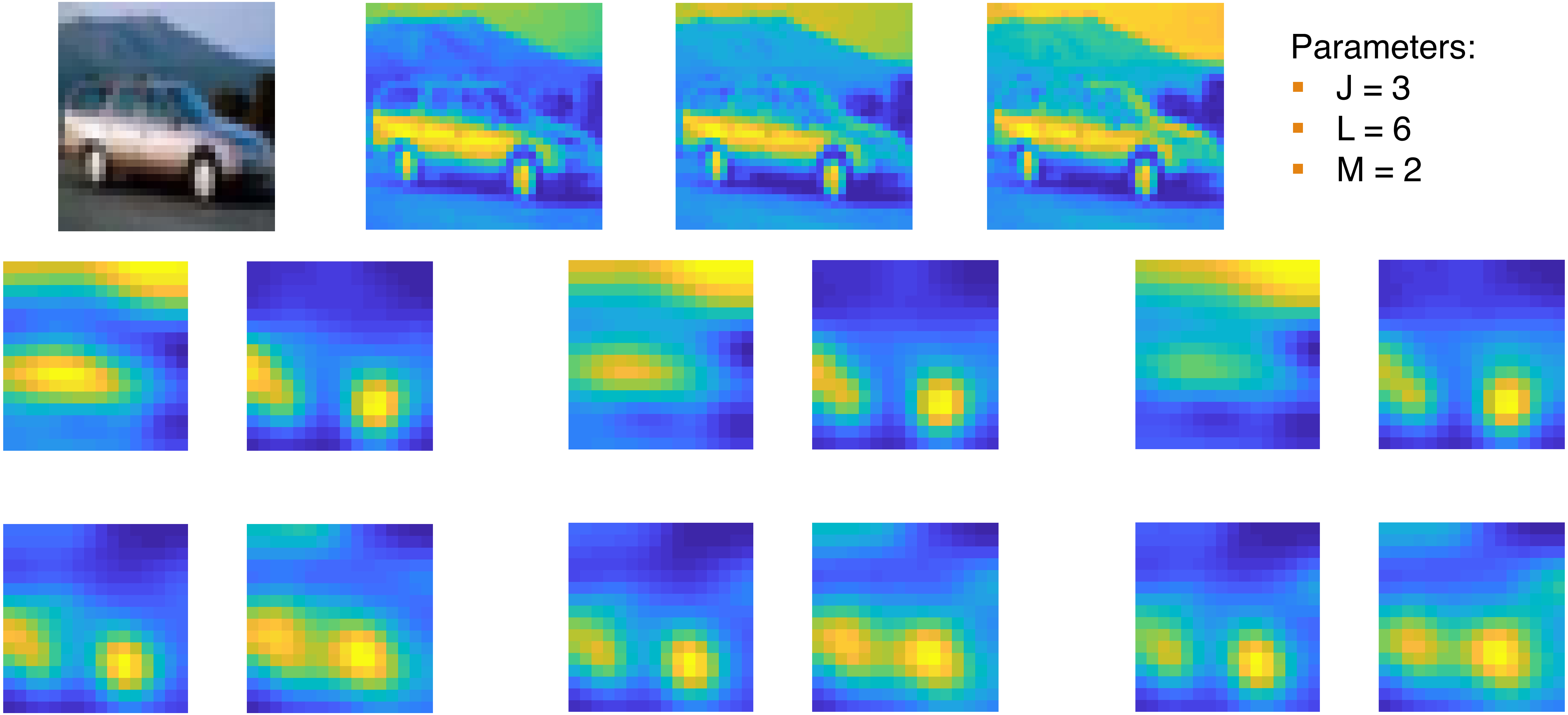




Parameters:

- $J = 3$
- $L = 10$
- $M = 2$

Scattering Network

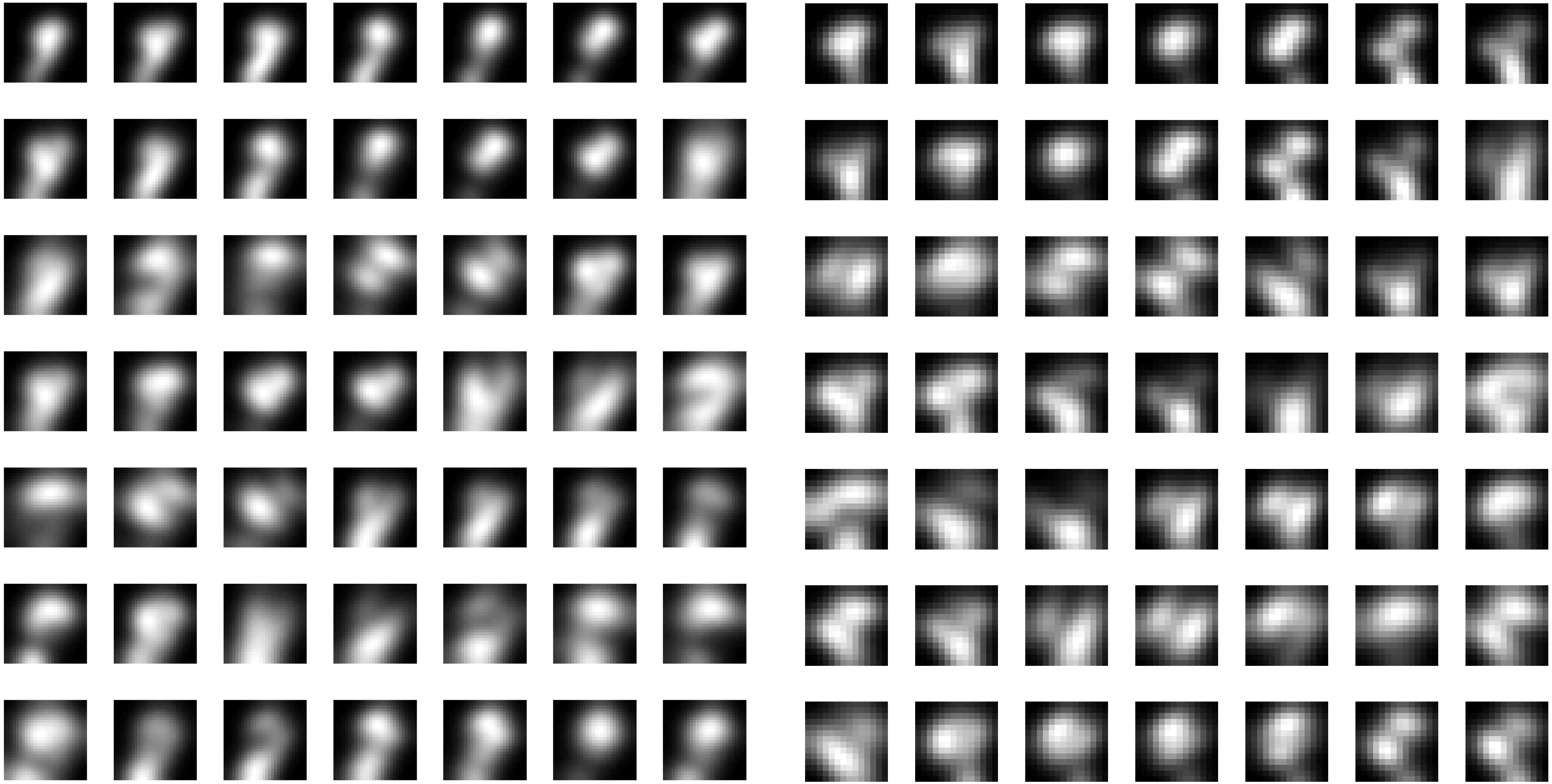


Parameters:

- J = 3
- L = 6
- M = 2

Scattering Network

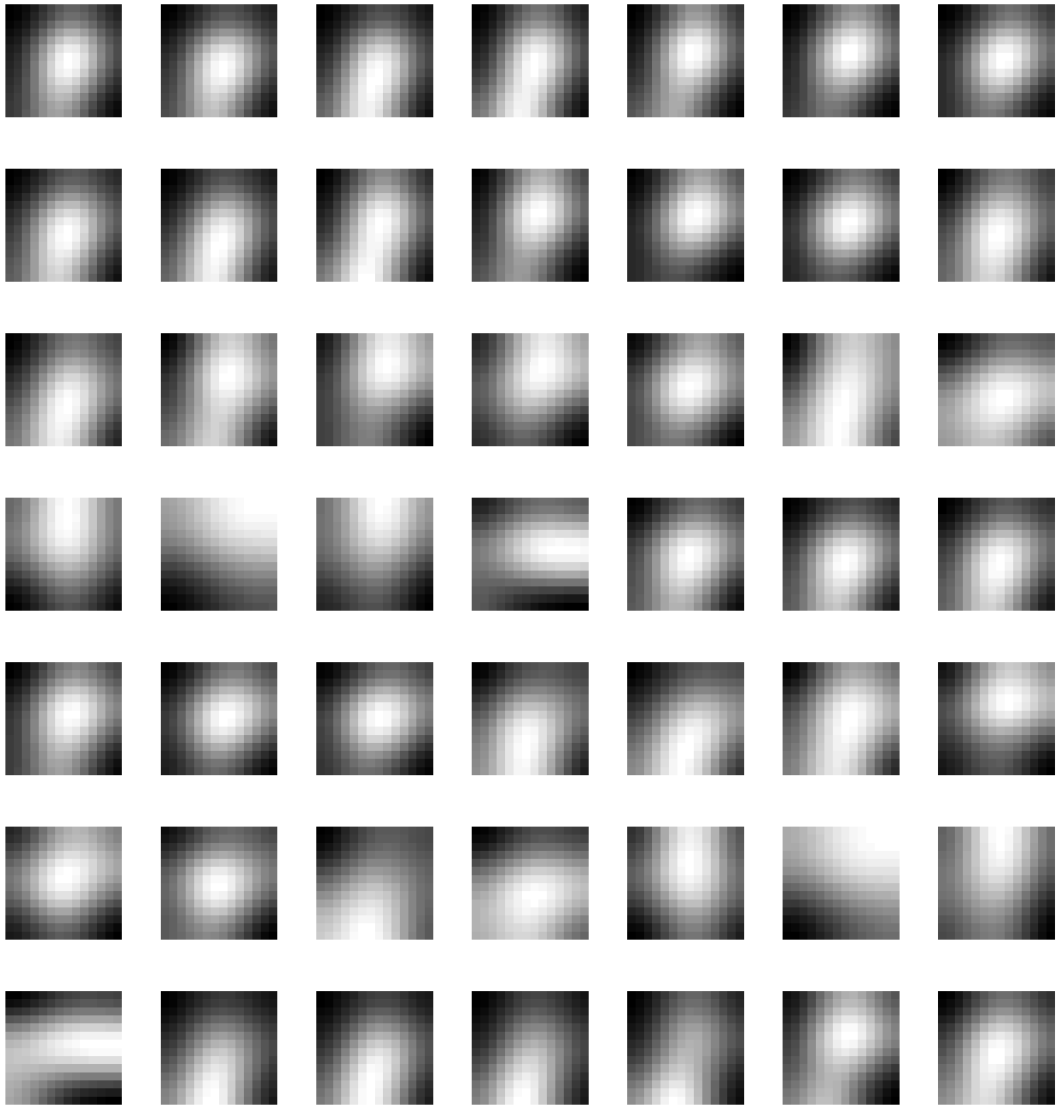
DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

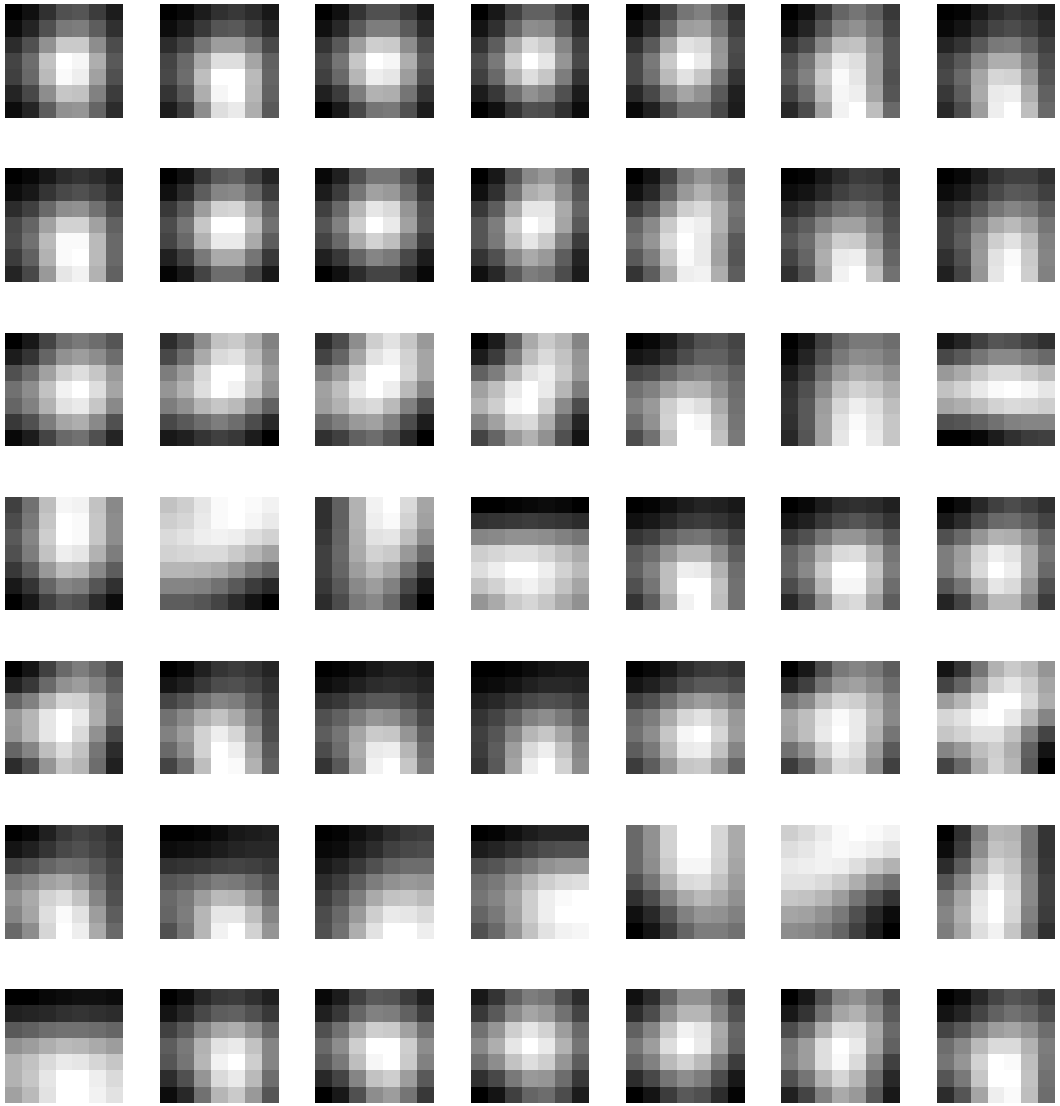


Feature maps of two Images of digit ‘9’

Scattering Network

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

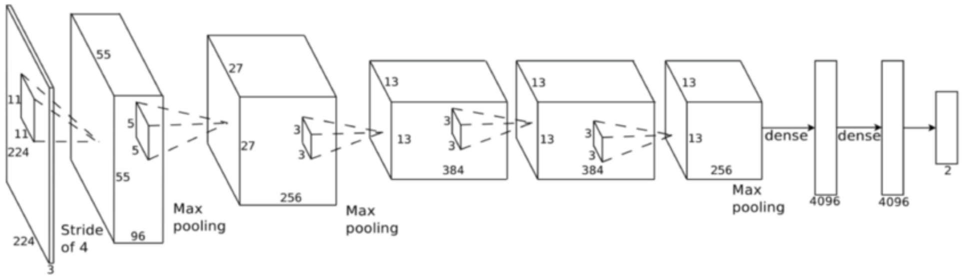




Pre-trained DNNs

AlexNet and ResNet

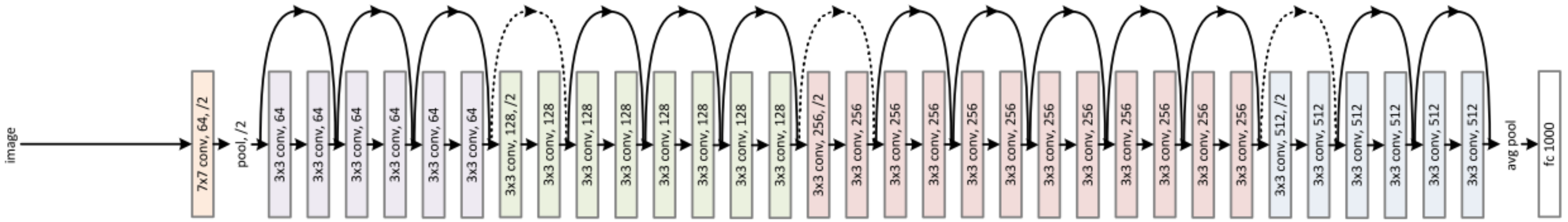
AlexNet
(2012)



Winner of 2012 ImageNet with top-5 error of 15.3%
- ReLU activation function.
- Local response normalization.
- Dropout.

$$\mathcal{F}(\mathrm{x}) = \mathcal{H}(\mathbf{x}) - \mathrm{x}$$

ResNet
(2016)



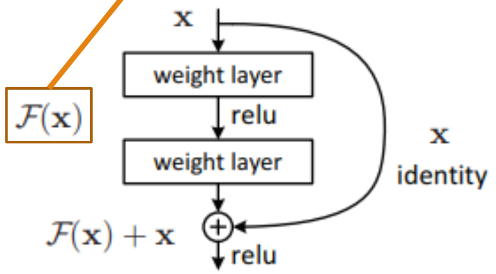


Figure 2. Residual learning: a building block.

Enabling hundreds of layers being trained
- Identity mapping by shortcuts.
- 3.57% top-5 error on ImageNet.
- 152 layers, yet lower complexity than VGG nets.

[1] Alex Krizhevsky, Ilya Sutskever and Geoffrey E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, pp. 1097--1105, 2012.
[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren and Jian Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770--778, 2016.

Pre-trained DNNs

Feature Extraction using Pre-trained AlexNet and ResNet-18

- Pre-trained models provided in torchvision package
- Models are pre-trained on ImageNet.

Data Preparation

```
cifar10_transform = transforms.Compose([
    transforms.Resize((256, 256)),
    transforms.ToTensor(),
    transforms.Normalize(mean=[0.485, 0.456, 0.406],
                         std=[0.229, 0.224, 0.225])
    ])

mnist_transform = transforms.Compose([
    transforms.Resize((256, 256)),
    transforms.ToTensor(),
    transforms.Lambda(lambda x: x.repeat(3, 1, 1)),
    transforms.Normalize(mean=[0.485, 0.456, 0.406],
                         std=[0.229, 0.224, 0.225])
    ])
```

- Resize the images to a larger size
- Normalize the input images
- For MNIST data, map input images to 3-channel

[1] torchvision package: https://pytorch.org/docs/stable/torchvision/index.html

Visualization

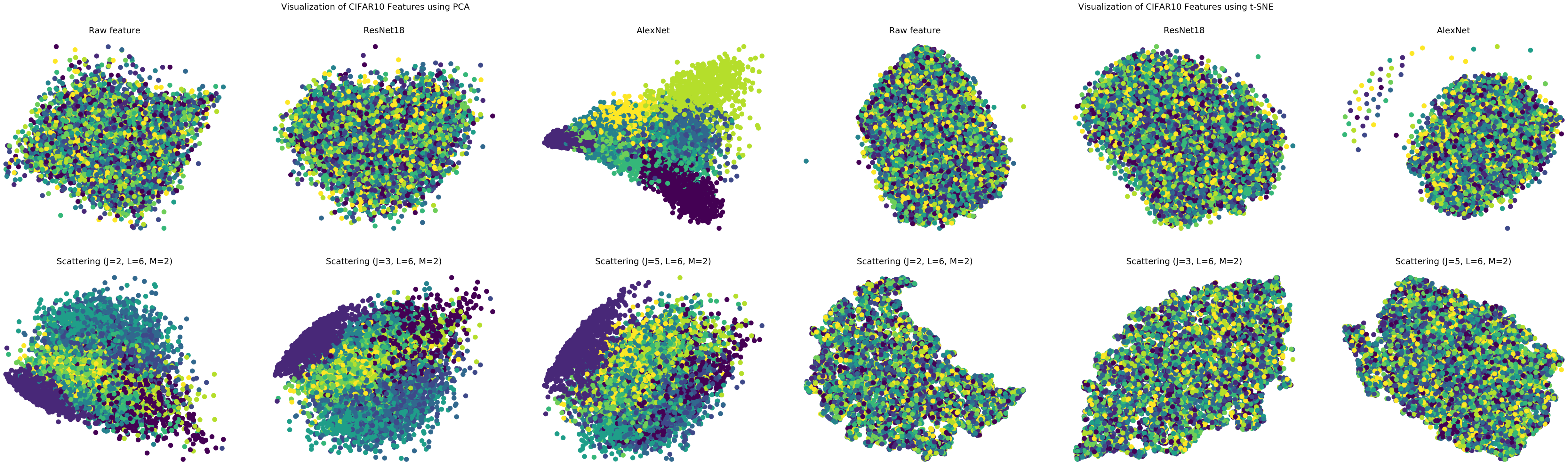
DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

- **Use classic dimensionality reduction methods: PCA and t-SNE**

- **MNIST:**
  - Raw feature is already very discriminative (dataset is not very complex).
  - With the parameter J increasing, discriminative power of Scattering net increase.
  - Features extracted by Scattering net features are far less discriminative than that by DNNs. (pre-defined vs. learnt from data)

Visualization

Use classic dimensionality reduction methods: PCA and t-SNE



CIFAR10:

- Dataset is much complicated than MNIST.

Classification

DEPARTMENT OF
COMPUTER SCIENCE
HONG KONG BAPTIST UNIVERSITY
香港浸會大學計算機科學系

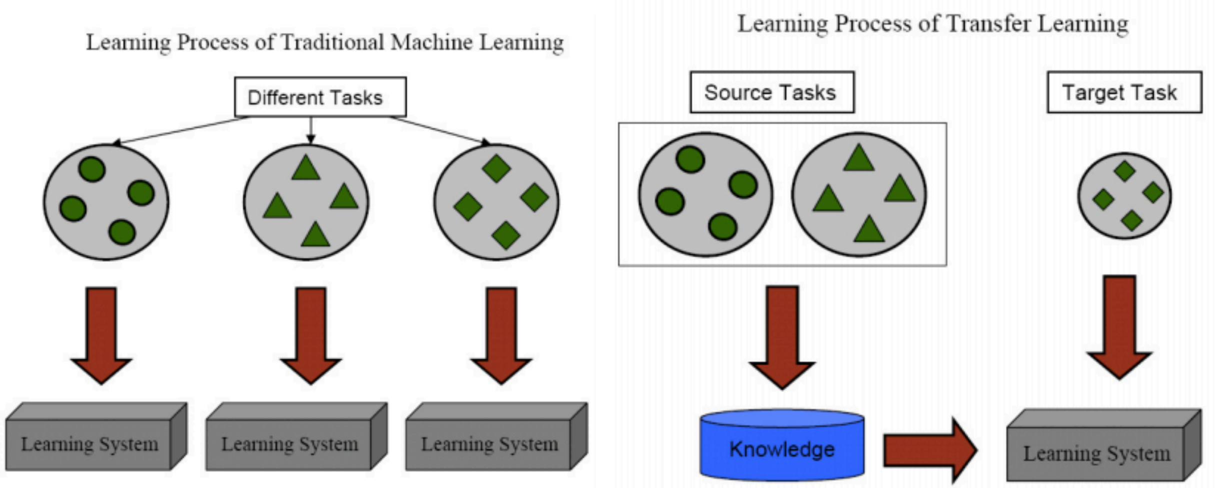
Logistic Regression

	MNIST	CIFAR10
Raw feature	0.92	0.41
ResNet-18	**0.96**	**0.79**
Scattering (J=2, L=6, M=2)	0.66	0.37
Scattering (J=3, L=6, M=2)	0.85	0.40
Scattering (J=5, L=6, M=2)	0.93	0.44

- Raw features of MNIST dataset are very discriminative, while that of CIFAR10 are not.
- ResNet achieved best performance in both datasets.
- Classification performance of the features extracted by Scattering net is much worse.
- With J increasing, the performance of Scattering net increase. When J=5, the performance of Scattering net exceeds the raw features. (Increase J will increase the range of translation invariance.)

Open Discussion

- **Transfer learning and domain adaptation, in general, is not a trivial problem.**



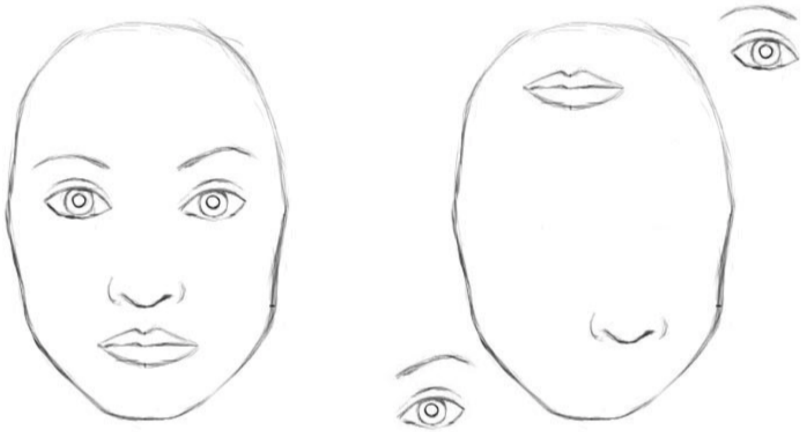
(a) Traditional Machine Learning (b) Transfer Learning

- Data distributions of ImageNet and CIFAR10 differ from each other.
- ResNet-18 trained on CIFAR10 achieve accuracy of 0.93.
- Large gap between the results obtained by transfer learning and that by direct training on the target data.
- Fine-tuning could potentially improve the performance futher.

[1] Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2010): 1345-1359.
[2] CNN performance on CIFAR10: https://github.com/kuangliu/pytorch-cifar

Open Discussion

- **Limitations of existing CNNs**
  - Detect the existence of features, but not the exact location (“part-to-whole” relationship not learned).



“Failure” of CNN:
Recognize both images as faces.
(two eyes + one mouth + one nose) = a face!

  - Hinton recently proposed the Capsule network to address this problem.
  - Recall the advantages of Scattering net: more control on the filters.
  - **Possible to preserve the complex spatial information by imposing more controls on the filters?**

[1] image from: https://medium.com/ai%C2%B3-theory-practice-business/understanding-hintons-capsule-networks-part-i-intuition-b4b559d1159b
[2] Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." *Advances in Neural Information Processing Systems*. 2017.
[3] Hinton, Geoffrey E., Sara Sabour, and Nicholas Frosst. "Matrix capsules with EM routing." in *ICLR*. 2018.

Thank you!

All questions and comments are greatly appreciated!

Dong Qian and **Kejing Yin**
{dongqian, cskjyin}@comp.hkbu.edu.hk