



Deep Gabor Scattering Network for Image Classification

Li-Na Wang, Benxiu Liu, Haizhen Wang, Guoqiang Zhong^(✉),
and Junyu Dong^(✉)

Department of Computer Science and Technology,
Ocean University of China, Qingdao, China
{gqzhong, dongjunyu}@ouc.edu.cn

Abstract. Deep learning models obtain exponential ascension in the field of image classification in recent years, and have become the most active research branch in AI research. The success of deep learning prompts us to make greater achievements in image classification. How to obtain effective feature representation becomes particularly important. In this paper, we combine the wavelet transformation and the idea of deep learning to build a new deep learning model, called Deep Gabor Scattering Network (DGSN). Concretely, in DGSN, we use the Gabor wavelet transformation to extract the invariant information of the images, partial least square regression (PLSR) for feature selection, and support vector machine (SVM) for classification. A key benefit of DGSN is that Gabor wavelet transformation can extract rich invariant features from the images. We show that DGSN is computationally simpler and delivers higher classification accuracy than related methods.

Keywords: Deep learning · Gabor filter · Invariant information
Deep Gabor scattering network (DGSN)

1 Introduction

In recent years, with the deepening of deep learning research, variety of deep learning models have been proposed and their application fields have been rapidly expanded. Nowadays, deep learning becomes to the most active branch of the artificial intelligence research. But the history of deep learning is not long. LeCun et al. [11] published the foreword exploration of the convolutional neural network in 1998, but it experienced years of deposition before deep learning really broke out. Since in 2006 Hinton and Salakhutdinov [8] proposed the concept of deep learning, researchers have gradually started the study of deep learning, constructed a variety of different deep learning models, and achieved a series of breakthrough research results. The rise of deep learning models has greatly improved the accuracy of object recognition and even surpassed the human level in some recognition tasks. However, in the actual image classification problems, deep learning models still face many challenges. How to construct an effective

deep network model and apply it to image classification is an important issue that needs to be solved urgently.

The challenge of image classification is mainly manifested in: lighting, shooting angle, deformation and other factors may cause image diversity. In image classification research, wavelet transform can solve this problem to a certain extent. It can recover images from unsuitable lighting and deformation situations and can better extract the invariant information in images.

In 1985, Meyer proved the existence of the wavelet function in the one-dimensional case and made a deep study in theory [14]. Based on the idea of multi-resolution analysis, Mallat proposed the Mallat algorithm, which plays an important role in the application of wavelet [19]. Its position in wavelet analysis is equivalent to CNN in deep learning. In image processing, the Gabor function is a linear filter for edge extraction. Its working principles are similar to the human visual system. Some existing study found that Gabor filter is very suitable for texture expression and separation. In the spatial domain, a two-dimensional Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. Mehrotra, Namuduri and Ranganathan [13] designed a computational model based on Gabor filters and the model was successfully used for edge detection, texture classification, etc. Chen, Cheng and Mallat [4] introduced a Haar scattering transform which computes invariant signal descriptors. Based on these, several researchers have demonstrated the utility of wavelet network for image analysis.

In this paper, we combine the wavelet transformation and the machine learning algorithms, in which we employ Gabor wavelet transformation to extract the invariant information of the images and use the partial least square regression (PLSR) [16,18] for feature selection and SVM for classification. We name the designed architecture as Deep Gabor Scattering Network (DGSN). Similar to LeNet-5 [11] and its variants [10], the structure of DGSN can be deeper by adding some convolutional and fully connected layers. However, to demonstrate the effectiveness of DGSN, we only use its prototype model here. The remaining part of this paper is composed of four sections. The deep learning methods based on wavelet are reviewed in Sect. 2. The architecture of our DGSN model is introduced in Sect. 3. Section 4 reports the experimental results with comparison to related work. Finally, Sect. 5 concludes this paper.

The main contributions of this paper are as follows:

- (1) We propose a new network structure called Deep Gabor Scattering Network (DGSN) for image classification.
- (2) A key benefit of DGSN is that based on the Gabor wavelet transformation, DGSN can extract rich invariant information of the images.
- (3) We show that DGSN is computationally simpler and delivers a better classification accuracy than compared approaches [4].

2 Related Work

In 2006, Hinton and Salakhutdinov [8] proposed the concept of deep learning which had gradually become more and more popular. Especially, in the field of

image recognition, the rise of convolutional neural networks (CNNs) had greatly improved the recognition accuracy and even in some tasks that had exceeded the recognition of human eyes, such as on object recognition [9, 12] and image classification [5]. With the development of the neural network, people tend to construct a deep network and believe that this deep network structure can learn more abstract features and have high learning ability.

As the application of deep network more and more extensive, many scholars have proposed a variety of deep network models to deal with different tasks. Among others, Krizhevsky et al. [10] put forward the AlexNet network structure to win the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC 2012). AlexNet contains eight layers, including five convolutional and three fully connected layers, and uses the “Dropout” technology. The model also performed well when migrating to other image classification tasks and became a commonly used model. In the ILSVRC 2014, researchers of Google proposed a deep network structure named GoogLeNet, which has 22 layers. This “deeper” network structure can learn more abstract features from images and achieved better classification results than AlexNet.

In practice, the problems of image diversity can be caused by lighting, shooting angle and deformation. To alleviate these problems, Bruna and Mallat [1] proposed the Invariant Scattering Convolution Networks (ISCN), which introduced the wavelet transformation into the deep learning area, and achieved fairly good classification results. Moreover, Guo and Mousavi [7] designed a deep CNN named Deep Wavelet Super-Resolution (DWSR) to predict the “missing details” of wavelet coefficients of the low-resolution images to obtain the Super-Resolution (SR) results. Said et al. [15] proposed a novel approach for image classification using wavelet networks and deep learning methods, which was called Deep Wavelet Network. The work of [4] introduced a Haar scattering transformation, which computes the invariant signal descriptors. It is implemented with a deep cascade of additions, subtractions and absolute values, iteratively computing the orthogonal Haar wavelet transformation. In this paper, we exploit the Gabor transformation for image feature learning.

In this paper, we aim to combining the Gabor wavelets with machine learning techniques for image classification. We propose a novel deep architecture called Deep Gabor Scattering Network (DGSN), which can effectively extract the invariant information of the images and categorize them.

3 Deep Gabor Scattering Network (DGSN)

In this section, we introduce the proposed new model, DGSN. We begin with a brief overview of the network architecture, followed by specific details. In order to classify the images, a deep CNN generally uses convolutional operations to extract the features of images. However, there are some difficulties to extract invariant information of the images with the convolutional operation, due to lighting, shooting angle and deformation, which may cause the image diversity problem. In order to overcome this problem, we use the Gabor wavelet transformation to replace the convolutional operation for the images’ feature extraction,

followed by PLSR for feature selection and SVM for classification. To the end, we obtain a deep Gabor scattering network, which can effectively learn the representations of the images and classify them.

3.1 The Structure of DGSN

DGSN is constructed by combining the Gabor wavelet transformation, PLSR and SVM classifier. It is constituted of five layers. The structure of DGSN is illustrated in Fig. 1, which follows the work of Deep Haar Scattering Network (HaarScat) [4]. Suppose the input images are of size $n \times n$. The Gabor layer consists of 32 Gabor filters corresponding to 4 scales and 8 orientations for image feature extraction, the PLSR layer is used for feature selection and dimensionality reduction, and the SVM layer for images classification. The main difference between DGSN and HaarScat is on the image feature learning. DGSN uses Gabor wavelet transformation, while the latter uses Haar wavelet transformation to extract the image features. Compared to HaarScat, DGSN can greatly reduce the training time and improve the classification performance under the condition of ensuring the classification accuracy.

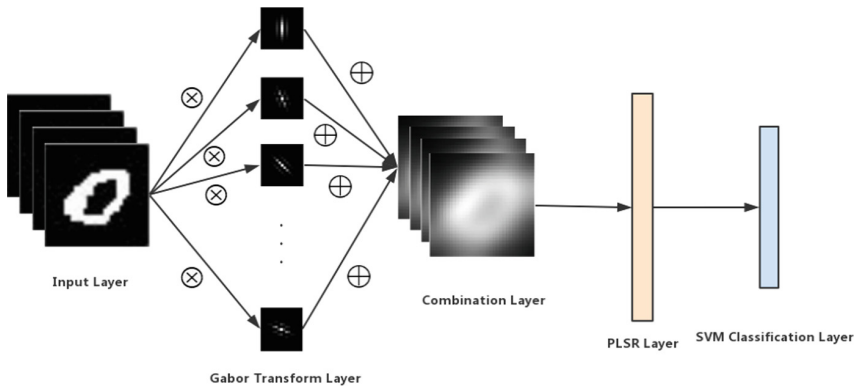


Fig. 1. The structure of DGSN.

3.2 Gabor Filters

In image processing, the Gabor wavelet transformation is an effective feature extraction algorithm. The working principles of the Gabor wavelet transformation are similar to the human visual system. Gabor wavelet transformation has good scale and direction selection characteristics. It is sensitive to the edge information of the image and able to adapt to the situation of light changes. Many pieces of work have found that Gabor wavelet transformation is very suitable for texture expression and separation. Compared with other methods in feature

extraction, Gabor wavelet transformation in general needs less training data and can meet the real-time requirements of some practical systems. Furthermore, it can tolerate to a certain degree of image rotation and deformation.

A Gabor kernel can obtain the response of the image in the frequency domain, and the result of the response can be regarded as a feature of the image. Then, if we use multiple Gabor kernels with different frequencies to obtain the response of the images, we can finally construct the representations of the images in the frequency domain.

In the space domain, a two-dimensional Gabor filter is a Gaussian kernel function with sine wave modulation. The filter consists of a real part and an imaginary part, which are orthogonal to each other. The mathematical expression of the Gabor function is as follows [6]:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right), \quad (1)$$

with the real part:

$$g_r(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi\frac{x'}{\lambda} + \psi\right), \quad (2)$$

and the imaginary part:

$$g_i(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \sin\left(2\pi\frac{x'}{\lambda} + \psi\right), \quad (3)$$

where

$$x' = x \cos \theta + y \sin \theta, \quad (4)$$

and

$$y' = -x \sin \theta + y \cos \theta. \quad (5)$$

Here, x and y indicate the position of the pixel on the x-axis and y-axis, λ is the wavelength, θ is the orientation, ψ is the phase offset, γ represents the aspect ratio, and σ is the standard deviation of the Gaussian factor of the Gabor function.

Gabor filters are self-similar. All Gabor filters can be generated from a mother wavelet after expansion and rotation. In many applications, when an image is given to a Gabor filter, the Gabor filter extracts the invariant feature of the image and produces different features with different scales and frequencies. Finally, all feature images are superimposed as a tensor, which is then normalized as the input of the next layer in DGSN.

3.3 PLSR and Classification

In DGSN, the Gabor transform layer is followed by PLSR for Gabor feature selection and the Gaussian kernel SVM for classification.

PLSR is a statistical method, and to some extent related to the principal components analysis (PCA). It can find a linear regression model through projecting the predicted and observed variables into a new space, rather than looking for the hyperplanes of maximum variance between independent variables and the response variables. PLSR can solve many problems which cannot be solved by ordinary multiple variable regression, and it can also realize the comprehensive application of various data analysis methods. Here, we use PLSR to select Gabor features and reduce the feature's dimensionality simultaneously.

Support vector machine (SVM) is one of the most commonly used classifiers and one of the most effective classifiers. It has excellent generalization ability and its own optimization goal is to achieve the least structural risk. In the classification layer, we use the SVM model in LibSVM [3], which is an easy, fast and effective SVM pattern recognition and regression package. The LibSVM Toolkit provides the default parameters. In most cases, researchers can use these default parameters to achieve a good classification effect and greatly reduce the time used to adjust the parameters. Even if the researcher wants to adjust the parameters, the toolkit provides a very convenient method of parameter selection. In the LibSVM toolkit, researchers can choose the type of SVM, kernel functions, and their parameters based on specific problems.

4 Experiments and Results

4.1 Data Sets

We train and evaluate our network on three standard data sets: MNIST, Yale [2] and Fashion-MNIST [17].

- (1) MNIST is a database which contains a training set of 60,000 handwritten digits images and a test set of 10,000 handwritten digits images with the resolution of $28 * 28$ pixels.
- (2) Yale is a face database which contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per individuals, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink. The resolution of each image is $32 * 32$ pixels.
- (3) Fashion-MNIST is a database of 70,000 positive images of 10 different products. The size, format, and the division of training set and test set of Fashion-MNIST are fully aligned with the original MNIST. It also contains a training set of 60,000 products images and a test set of 10,000 products images with the resolution of $28 * 28$ pixels. It contains 10 classes, some example images are shown in Fig. 2.

Over the past decades, classical MNIST dataset is often used as a benchmark for testing algorithms in the field of machine learning, machine vision, artificial intelligence and deep learning. MNIST is too simple and many algorithms have achieved 99% performance on the test set. Fashion-MNIST is an image data set

that replaces the handwritten digits set of MNIST. Therefore, in addition to train and test our DGSN on the MNIST data set, we also train and test DGSN on the Fashion-MNIST data set. For Fashion-MNIST, we don't need to modify any algorithm and can adopt the dataset directly. Furthermore, we have tested our DGSN on a data set, Yale, for face recognition.

4.2 Experimental Settings

During the training process, the size of the input images is fixed to $32 * 32$. In the Gabor transform layer, we use 32 Gabor filters with 4 scales and 8 orientations to extract the frequency information of the images. The parameters of the wave lengths, bandwidths, aspect ratios and angles are set to $\{2, 4, 6, 8\}$, 1, 0.5, $\{0, \pi/2\}$, and $\{0, \pi\}$, respectively. The orientation is set to $\{0, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8\}$. To show the effect of the wave length, we illustrate the classification accuracy against different values of λ in Fig. 3. We can see that the classification accuracy increases with the increment of the value of the wave length λ . In order to extract rich invariant feature from the images, we select 4 values for the Gabor filters.

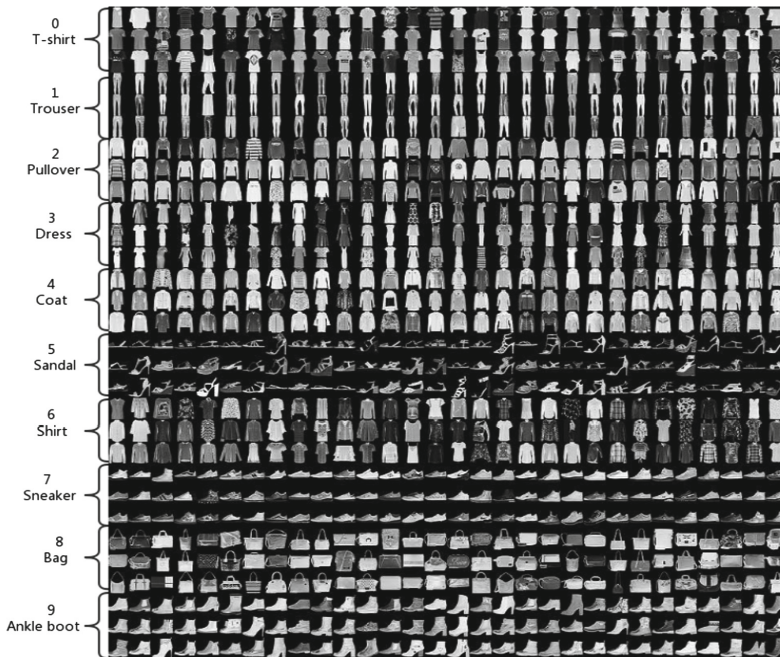


Fig. 2. Some example images in the Fashion-MNIST data set (each category takes up three lines).

In Gabor transform layer of DGSN, 32 Gabor filters are used as shown in Fig. 4. The image information extracted by each Gabor filter is shown in Fig. 5. Then, we combine them into one image as illustrated in Fig. 6.

4.3 Experimental Results

In this section, we mainly compare DGSN with the work of [4]. [4] used Haar wavelet transformation to extract invariant information of the images. Compared with this model, we use Gabor wavelet transformation to compute the invariant representations of the images. Classification results on the MNIST data set are shown in Table 1.

Table 1. Classification results on the MNIST data set.

Model	100	200	300	400	500	All
HaarScat [4]	91.20%	94.80%	95.47%	95.80%	96.18%	—
DGSN (ours)	94.70%	95.20%	95.93%	96.55%	96.86%	98.99%

Due to out of memory, we didn't obtain the results of [4] on the total MNIST data set (as well as on the Fashion-MNIST data set). Therefore, we select some images from each class in the MNIST data set to construct small training set. The 100, 200, . . . , 500 in Table 1 represent the numbers of images from each class. We can see that DGSN performs much better than HaarScat [4].

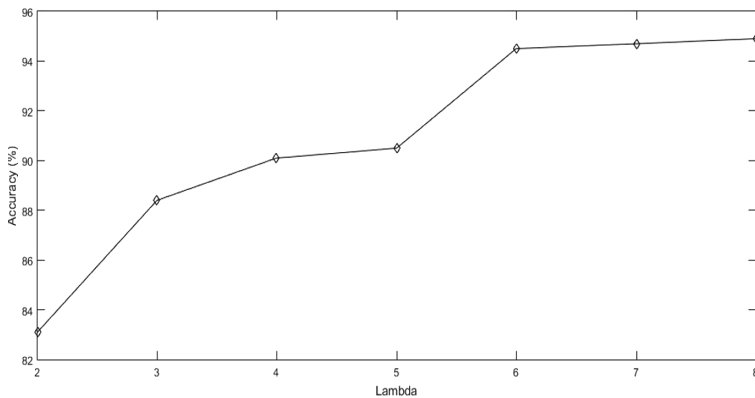


Fig. 3. The effect of the wave length λ with respect to the classification accuracy. The results are obtained on a subset of MNIST with 100 images in each class.

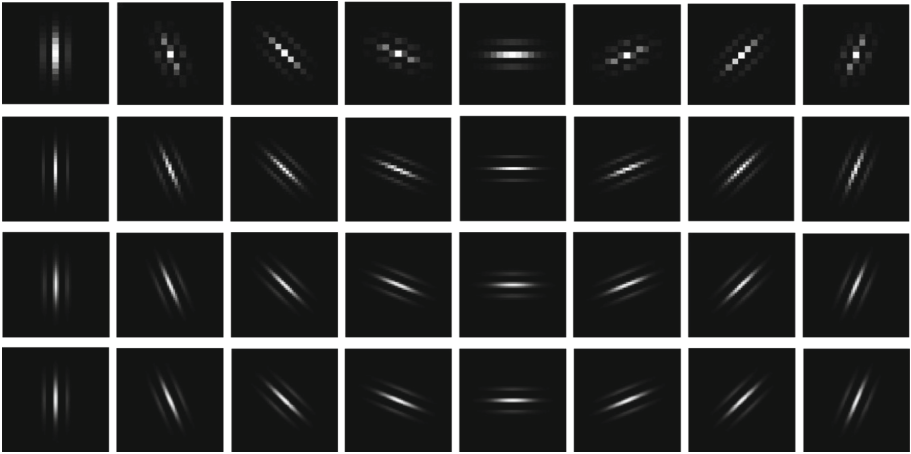


Fig. 4. The Gabor filters in the Gabor Layer with 8 orientation and $\lambda = 2, 4, 6, 8$ (corresponding to each row), respectively.

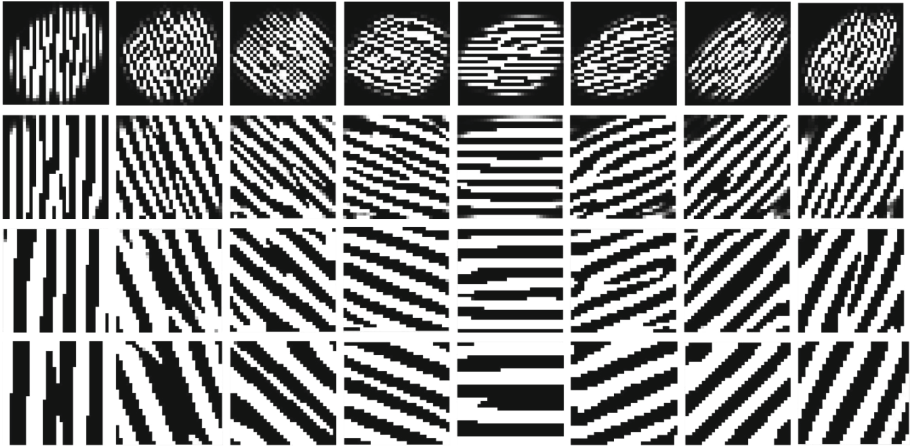


Fig. 5. The image information extracted by different Gabor filters in the Gabor Layer with 8 orientation and $\lambda = 2, 4, 6, 8$ (corresponding to each row), respectively.

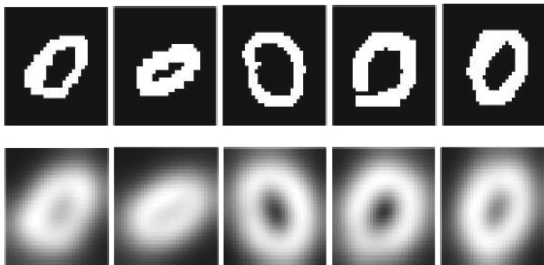


Fig. 6. The Gabor features. Top: the input images to the Gabor layer. Bottom: The Gabor features extracted by the Gabor filters.

Next, we apply our model to the Yale faces. The results obtained on the Yale faces is shown in Table 2. The “(4, 5, 6, 7, 8) Train” is a random subset of the images per individual, which is taken with labels to form the training set and the rest of the database was considered to be the test set. For each given setting, there are 50 randomly splits [4]. We can see that DGSN outperforms HaarNet [4] with a large margin.

Table 2. Results obtained on the Yale faces.

Model	4 Train	5 Train	6 Train	7 Train	8 Train
HaarScat [4]	36.19%	55.56%	60.00%	55.00%	62.22%
DGSN(ours)	49.52%	57.78%	65.33%	61.67%	68.89%

The training time of DGSN and HaarScat on the MNIST and Yale data sets is shown in Table 3. The training time on the MNIST data set is on the entire data set, and the training time on the Yale data set is the average training time on the 50 splits.

Table 3. Comparison of the training time on the MNIST and Yale data sets.

Model	MNIST	Yale
HaarScat [4]	—	773 s
DGSN(ours)	8172 s	18 s

According to the results shown in Tables 2 and 3, we can see that DGSN has a better performance than HaarScat on both the MNIST and the Yale data sets. More importantly, the training time of DGSN is much less than HaarScat, and DGSN can achieve high classification accuracy without large memory, which demonstrates the advantage of DGSN over HaarScat.

Finally, we apply DGSN to the Fashion-MNIST data set. The accuracy obtained by DGSN is 90.60%. We compare its classification results with that obtained by previous methods on this data set [17]. The results are shown in Fig. 7. From the histogram, we can see that DGSN delivers higher accuracy than the compared methods.

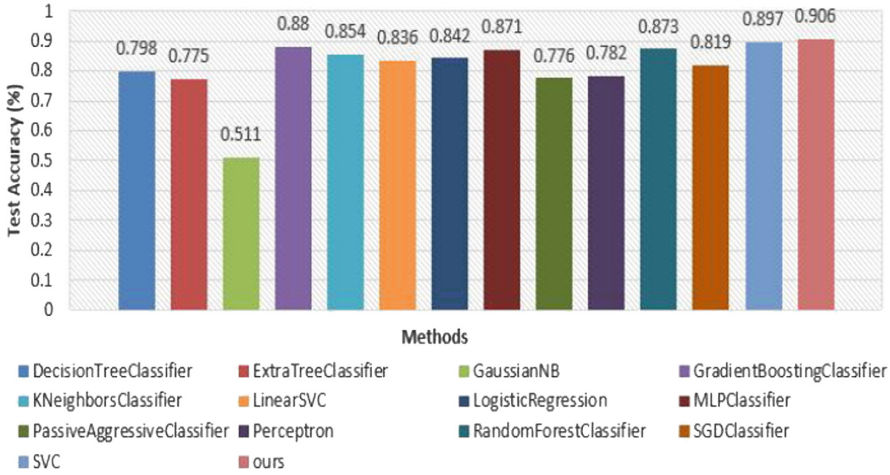


Fig. 7. Results obtained on the Fashion-MNIST data set.

5 Conclusion

We propose a new deep architecture called Deep Gabor Scattering Network (DGSN) for image classification. DGSN combines the wavelet transformation and the idea of deep learning. It uses Gabor wavelet transformation to extract the invariant information of the images, PLSR for feature selection, and SVM for classification. A key benefit of DGSN is that rich invariant features of the images can be extracted by the Gabor wavelet transformation. With extensive experiments, we show that DGSN is computationally simpler and delivers higher classification accuracy than compared methods. For future work, we plan to combine the Gabor wavelet transformation with deep convolutional neural networks (CNNs), so that we can replace the convolutional operation with Gabor transformation on account of the invariant feature extraction from images.

Acknowledgments. This work was supported by the Science and Technology Program of Qingdao under Grant No. 17-3-3-20-nsh, the CERNET Innovation Project under Grant No. NGII20170416 and the Fundamental Research Funds for the Central Universities of China.

References

1. Bruna, J., Mallat, S.: Invariant scattering convolution networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(8), 1872–1886 (2013)
2. Cai, D., He, X., Hu, Y., Han, J., Huang, T.S.: Learning a spatially smooth subspace for face recognition. In: *CVPR* (2007)
3. Chang, C., Lin, C.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2**(3), 1–27 (2011)

4. Chen, X., Cheng, X., Mallat, S.: Unsupervised deep haar scattering on graphs. In: NIPS, pp. 1709–1717 (2014)
5. Ciresan, D., Meier, U., Masci, J., Gambardella, L., Schmidhuber, J.: Flexible, high performance convolutional neural networks for image classification. In: IJCAI, pp. 1237–1242 (2011)
6. Grigorescu, C., Petkov, N., Westenberg, M.A.: Contour detection based on non-classical receptive field inhibition. *IEEE Trans. Image Process.* **12**(7), 729–739 (2003)
7. Guo, T., Mousavi, H., Vu, T., Monga, V.: Deep wavelet prediction for image super-resolution. In: CVPR Workshops, pp. 1100–1109. IEEE Computer Society (2017)
8. Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006)
9. Jarrett, K., Kavukcuoglu, K., Ranzato, M., LeCun, Y.: What is the best multi-stage architecture for object recognition? In: ICCV, pp. 2146–2153. IEEE Computer Society (2009)
10. Krizhevsky, A., Sutskever, I., Hinton, G.: ImageNet classification with deep convolutional neural networks. In: NIPS, pp. 1106–1114 (2012)
11. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998)
12. LeCun, Y., Huang, F.J., Bottou, L.: Learning methods for generic object recognition with invariance to pose and lighting. In: CVPR, pp. 97–104 (2004)
13. Mehrotra, R., Namuduri, K., Ranganathan, N.: Gabor filter-based edge detection. *Pattern Recognit.* **25**(12), 1479–1494 (1992)
14. Meyer, Y.: Principe D’Incertitude, Bases Hilbertiennes Et Algebres D’Operateurs. *Seminaire Bourbaki* **662**(145–146), 1985–1986 (1985)
15. Said, S., Jemai, O., Hassairi, S., Ejbali, R., Zaied, M., Amar, C.B.: Deep wavelet network for image classification. In: SMC, pp. 922–927 (2016)
16. Schwartz, W.R., Kembhavi, A., Harwood, D., Davis, L.S.: Human detection using partial least squares analysis. In: ICCV, pp. 24–31 (2009)
17. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-MNIST: A Novel Image Dataset for Benchmarking Machine Learning Algorithms. *CoRR* abs/1708.07747 (2017)
18. Zhang, H., Kiranyaz, S., Gabbouj, M.: Cardinal sparse partial least square feature selection and its application in face recognition. In: EUSIPCO, pp. 785–789 (2014)
19. Zhou, K., Wu, D.: Research of Mallat algorithm based on multi-resolution analysis. *Softw. Guide* **10**, 54–55 (2008)