Journal of Southeast University (English Edition) Vol. 31, No. 4, pp. 469 - 473 Dec. 2015 ISSN 1003—7985

Kernel principal component analysis network for image classification

Wu Dan[1,4] Wu Jiasong[1,2,3,4] Zeng Rui[1,4] Jiang Longyu[1,4]
Lotfi Senhadji[2,3,4] Shu Huazhong[1,4]

([1]Key Laboratory of Computer Network and Information Integration of Ministry of Education, Southeast University, Nanjing 210096, China)
([2]Institut National de la Santé et de la Recherche Médicale U 1099, Rennes 35000, France)
([3]Laboratoire Traitement du Signal et de l'Image, Université de Rennes 1, Rennes 35000, France)
([4]Centre de Recherche en Information Biomédicale Sino-Français, Nanjing 210096, China)

Abstract: In order to classify nonlinear features with a linear classifier and improve the classification accuracy, a deep learning network named kernel principal component analysis network (KPCANet) is proposed. First, the data is mapped into a higher-dimensional space with kernel principal component analysis to make the data linearly separable. Then a two-layer KPCANet is built to obtain the principal components of the image. Finally, the principal components are classified with a linear classifier. Experimental results show that the proposed KPCANet is effective in face recognition, object recognition and handwritten digit recognition. It also outperforms principal component analysis network (PCANet) generally. Besides, KPCANet is invariant to illumination and stable to occlusion and slight deformation.

Key words: deep learning; kernel principal component analysis net (KPCANet); principal component analysis net (PCANet); face recognition; object recognition; handwritten digit recognition

doi: 10.3969/j.issn.1003 - 7985.2015.04.007

A major difficulty of image classification is the considerable intra-class variability, arising from different illuminations, rigid deformations, non-rigid deformations and occlusions, which are useless for classification and should be eliminated. Deep learning structures like deep convolutional networks have the ability to learn invariant features[1]. Bruna et al. [2] built a scattering network (ScatNet) which is invariant to both rigid and non-rigid deformations. Chan et al. [3] constructed a principal component analysis network (PCANet), which cascaded principal component analysis (PCA), binary hashing, and block-wise histogram. PCANet achieves the state-of-the-art accuracy in many datasets of classification tasks, such as extended Yale B dataset, AR dataset, and FERET dataset. Kernel PCA (KPCA) [4-5] is a nonlinear generalization of PCA in the sense that it performs PCA in the feature spaces of arbitrary large dimension. KPCA can generally provide a better recognition rate than ordinary PCA due to the following two reasons: 1) KPCA uses an arbitrary number of nonlinear components, while ordinary PCA uses only a limited number of linear principal components; 2) KPCA has more flexibility than ordinary PCA since KPCA can choose different kernel functions (for example, Gaussian kernel, Polynomial kernel, etc.) for different recognition tasks, while ordinary PCA uses only linear kernel functions.

In this paper, we propose a new deep learning network named kernel principal component network (KPCANet), which cascades two KPCA stages and one pooling stage. When the kernel function is linear, the proposed KPCANet degrades to the PCANet[3]. Experimental results show that the proposed KPCANet is invariant to illumination and stable to slight non-rigid deformation, and it generally outperforms PCANet in both face recognition and object recognition tasks.

1 KPCANet

Fig. 1 shows the whole structure of the proposed KPCANet, which consists of two KPCA stages and one pooling stage. Suppose that the patch size is $k_1 \times k_2$ at all stages, and all the input images are of size $m \times n$.

1.1 The first stage of KPCANet

We input N images $\boldsymbol{I}_i$ ($i = 1, 2, \dots, N$) that belong to c classes, and take a patch $\boldsymbol{p}_{i,j} \in \mathbf{R}^{k_1 \times k_2}$ centered in the j-th

Received 2015-05-04.

Biographies: Wu Dan (1990—), female, graduate; Shu Huazhong (corresponding author), male, doctor, professor, shu.list@seu.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. 61201344, 61271312, 61401085, 11301074), the Research Fund for the Doctoral Program of Higher Education (No. 20120092120036), the Program for Special Talents in Six Fields of Jiangsu Province (No. DZXX-031), Industry-University-Research Cooperation Project of Jiangsu Province (No. BY2014127-11), "333" Project (No. BRA2015288), High-End Foreign Experts Recruitment Program (No. GDT20153200043), Open Fund of Jiangsu Engineering Center of Network Monitoring (No. KJR1404).

Citation: Wu Dan, Wu Jiasong, Zeng Rui, et al. Kernel principal component analysis network for image classification[J]. Journal of Southeast University (English Edition), 2015, 31(4): 469 - 473. [doi: 10.3969/j.issn.1003 - 7985.2015.04.007]

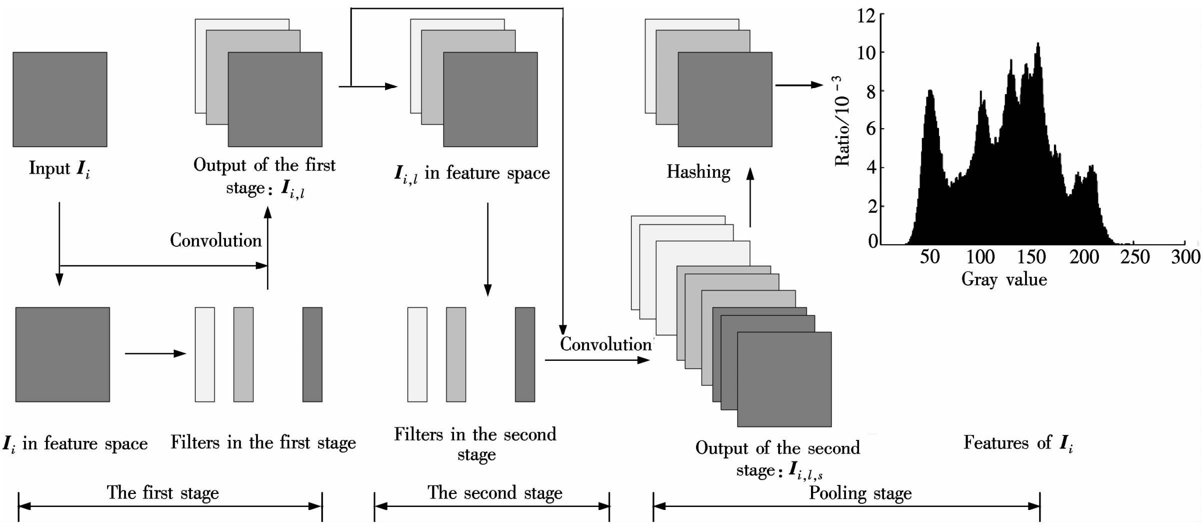

Fig. 1 The detailed block diagram of the proposed KPCANet

$(j=1, 2, \ldots, mn)$ pixel of image $\boldsymbol{I}_i$ and vectorize the patch as $\boldsymbol{x}_{i,j} \in \mathbf{R}^{k_1k_2}$. Collecting all the vectorized patches $\boldsymbol{x}_{i,j}(j=1, 2, \ldots, mn)$ of $\boldsymbol{I}_i$, we obtain a matrix $\boldsymbol{X}_i = [\boldsymbol{x}_{i,1}, \boldsymbol{x}_{i,2}, \ldots, \boldsymbol{x}_{i,mn}] \in \mathbf{R}^{k_1k_2 \times mn}$. Constructing the same matrix for all input images and putting them together, we obtain $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_N] \in \mathbf{R}^{k_1k_2 \times Nmn}$. For convenience, the p-th column of $\boldsymbol{X}$ is denoted as $\boldsymbol{x}'_p(p=1, 2, \ldots, Nmn)$. We then map $\boldsymbol{X}$ from the input space $\mathbf{R}^{k_1k_2 \times k_1k_2}$ to the feature space $\mathscr{F}$ by

$$\boldsymbol{T}: \mathbf{R}^{k_1k_2 \times k_1k_2} \to \mathscr{F}, \quad \boldsymbol{X} \mapsto \boldsymbol{X}^{\mathscr{F}} \tag{1}$$

To find the principal component of $\boldsymbol{T}(\boldsymbol{x}'_p)$, we need to diagonalize the covariance matrix $\boldsymbol{C}$:

$$\boldsymbol{C} = \frac{1}{Nmn} \sum_{p=1}^{Nmn} \boldsymbol{T}(\boldsymbol{x}'_p) \boldsymbol{T}(\boldsymbol{x}'_p)^{\mathrm{T}} \tag{2}$$

To simplify the diagonalization of $\boldsymbol{C}$, we can diagonalize $\boldsymbol{K}$ instead, where $\boldsymbol{K}_{pq} = (\tilde{\boldsymbol{T}}(\boldsymbol{x}'_p) \cdot \tilde{\boldsymbol{T}}(\boldsymbol{x}'_q))_{pq}$, $\tilde{\boldsymbol{T}}(\boldsymbol{x}'_p)$ denotes the centralized $\boldsymbol{T}(\boldsymbol{x}'_p)$ and symbol "$\cdot$" denotes the dot product. Since the dimension of $\mathscr{F}$ can be arbitrarily large even infinite[4-5], it will be difficult to compute dot product $(\tilde{\boldsymbol{T}}(\boldsymbol{x}'_p) \cdot \tilde{\boldsymbol{T}}(\boldsymbol{x}'_q))$ directly. Therefore, we substitute dot product with kernel function k and obtain $\boldsymbol{K}_{pq} = (k(\boldsymbol{x}'_p, \boldsymbol{x}'_q))_{pq}$. After that, $\boldsymbol{K}$ is centralized with $\boldsymbol{K}' = \boldsymbol{K} - \mathbf{1}_{Nmn}\boldsymbol{K} - \boldsymbol{K}\mathbf{1}_{Nmn} + \mathbf{1}_{Nmn}\boldsymbol{K}\mathbf{1}_{Nmn}$ and $\boldsymbol{K}'$ is diagonalized to obtain the principal eigenvectors $\boldsymbol{W}_l^1$ $(l=1, 2, \cdots, L_1)$, which are the KPCA filters in the first stage, where $(\mathbf{1}_{Nmn})_{ij} = \dfrac{1}{Nmn}$.

Zero-padding the boundary of $\boldsymbol{I}_i$ and convolving it with $\boldsymbol{W}_l^1$, we obtain the l-th filter output of the first stage $\boldsymbol{I}_{i,l} = \boldsymbol{I}_i * \boldsymbol{W}_l^1 \in \mathbf{R}^{m \times n}$ $(i=1, 2, \cdots, N; l=1, 2, \cdots, L_1)$, where "$*$" denotes 2D convolution and L_1 denotes the amount of filters in the first stage.

1.2 The second stage of KPCANet

By repeating the same process as in the first stage on $\boldsymbol{I}_{i,l}(i=1, 2, \cdots, N; l=1, 2, \cdots, L_1)$, we obtain L_2 kernel PCA filters $\boldsymbol{W}_s^2$ $(s=1, 2, \cdots, L_2)$ of the second stage. Convolving $\boldsymbol{I}_{i,l}$ with $\boldsymbol{W}_s^2$, we obtain the output of the second stage $\boldsymbol{I}_{i,l,s} = \boldsymbol{I}_{i,l} * \boldsymbol{W}_s^2$ $(i=1, 2, \cdots, N; l=1, 2, \cdots, L_1; s=1, 2, \cdots, L_2)$.

1.3 The pooling stage of KPCANet

Every L_2 input images are binarized and converted to an image with

$$\boldsymbol{P}_{i,l} = \sum_{s=1}^{L_2} 2^{s-1} H(\boldsymbol{I}_{i,l,s}) = \sum_{s=1}^{L_2} 2^{s-1} H(\boldsymbol{I}_{i,l} * \boldsymbol{W}_s^2)$$
$$i=1, 2, \cdots, N; l=1, 2, \cdots, L_1 \tag{3}$$

where H is the Heaviside step (like) function[3].

Each of the L_1 images $\boldsymbol{P}_{i,l}(l=1, 2, \cdots, L_1)$ is then partitioned into B blocks. We compute the histogram of the decimal values in each block, and concatenate all the B histograms into one vector denoted as $\mathrm{Bhist}(\boldsymbol{P}_{i,l})$. Finally, the KPCANet features of $\boldsymbol{I}_i$ are given by

$$\boldsymbol{f}_i = [\mathrm{Bhist}(\boldsymbol{P}_{i,1}), \mathrm{Bhist}(\boldsymbol{P}_{i,2}), \cdots, \mathrm{Bhist}(\boldsymbol{P}_{i,L_1})]^{\mathrm{T}} \in \mathbf{R}^{(2^{L_2})L_1B} \tag{4}$$

Since deep architectures are composed of multiple levels of nonlinear operations, such as in complicated propositional formulae re-using many sub-formulae[6], the first two stages of KPCANet are set to be the same in this paper, we can re-use the whole structure of the first stage as well.

2 Experimental Results

We evaluate the performance of the proposed KPCANet on various databases including MNIST, USPS, Yale face dataset, COIL-100 objects dataset, and AR dataset. Besides, we compare KPCANets that cascade various (from one to three) stage(s) of the KPCA layer in this paper. All the features learned by KPCANet are classified with a SVM classifier.

2.1 Comparison of KPCANet in different recognition tasks

In this section, we use various kernel functions to evaluate the performance of the proposed KPCANet in recognition tasks including handwritten digit recognition, face recognition and object recognition. Kernel functions that are used in this paper are presented in Tab. 1.

MNIST[7] and USPS[8] are used to evaluate the performance of KPCANet on handwritten images. MNIST contains 60 000 train images and 10 000 test images, and all images are of size 28 × 28 pixel. USPS contains 9 298 images of size 16 × 16 pixel in total, 5 000 of which are chosen randomly to train KPCANet and the rest are for testing. The Yale face database[9] is used to evaluate the performance of the proposed KPCANet on face images. It contains 165 grayscale images of 15 individuals in GIF format, and each individual contains 11 images with different facial expressions or configurations: center-light, wearing glasses, happy, left-light, wearing no glasses, normal, right-light, sad, sleepy, surprised, and winking. All images of this database are cropped to size 64 × 64 pixel, 90 of which are chosen randomly to train the proposed KPCANet and the rest are for testing. COIL-100 (Columbia Object Image Library)[10] is a database of the color images of 100 objects. The images of the objects are taken at pose intervals of 5°, and they correspond to 72 poses per object. All images are transformed into gray images and cropped to size 32 × 32 pixel. Half images of each object are chosen randomly to train KPCANet and the others are for testing.

Tab. 1 Various kernel functions used in this paper

Kernel function	Expression	Value of parameters
Linear	$k(\boldsymbol{x},\boldsymbol{y})=\boldsymbol{xy}+c$	$c=0$
Gaussian	$k(\boldsymbol{x},\boldsymbol{y})=\exp\left(-\frac{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert^2}{2\sigma^2}\right)$	$\sigma=1$
Polynomial	$k(\boldsymbol{x},\boldsymbol{y})=(\boldsymbol{xy}+1)^d$	$d=3$
Exponential	$k(\boldsymbol{x},\boldsymbol{y})=\exp\left(-\frac{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert}{2\sigma^2}\right)$	$\sigma=1$
Laplacian	$k(\boldsymbol{x},\boldsymbol{y})=\exp\left(-\frac{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert}{\sigma}\right)$	$\sigma=1$
Sigmoid	$k(\boldsymbol{x},\boldsymbol{y})=\tanh(\alpha\boldsymbol{xy}+c)$	$\alpha=\frac{1}{2}, c=-1$
Rational quadratic	$k(\boldsymbol{x},\boldsymbol{y})=\exp\left(1-\frac{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert^2}{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert^2+c}\right)$	$c=1$
Inverse multiquadric	$k(\boldsymbol{x},\boldsymbol{y})=\frac{1}{\sqrt{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert^2+c^2}}$	$c=1$
Circular	$k(\boldsymbol{x},\boldsymbol{y})=\begin{cases}\left(\frac{2}{\pi}\operatorname{acos}\left(\frac{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert}{\sigma}\right)-\frac{2}{\pi}\Vert\boldsymbol{x}-\boldsymbol{y}\Vert\sqrt{\frac{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert^2}{\sigma^2}}\right) & \Vert\boldsymbol{x}-\boldsymbol{y}\Vert<\sigma\\ 0 & \text{otherwise}\end{cases}$	$\sigma=0.2$
Spherical	$k(\boldsymbol{x},\boldsymbol{y})=\begin{cases}\left(1-\frac{3}{2}\frac{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert}{\sigma}+\frac{1}{2}\left(\frac{\Vert\boldsymbol{x}-\boldsymbol{y}\Vert}{\sigma}\right)^3\right) & \Vert\boldsymbol{x}-\boldsymbol{y}\Vert<\sigma\\ 0 & \text{otherwise}\end{cases}$	$\sigma=0.2$

The performances of different kernel functions on datasets including MNIST, USPS, Yale face dataset and COIL-100 dataset are presented in Tab. 2. Both the patch size and the block size are set to be 8 × 8 pixel, and the filter number is set to be 8 at all stages. The overlapping ratio of block is 0.5.

Tab. 2 Comparison of error rates of KPCANet with various kernel functions on different datasets %

Kernel function	MNIST	USPS	Yale face dataset	COIL-100 objects dataset
Linear	0.85	2.37	5.33	1.36
Gaussian	1.07	2.70	4.00	0.89
Polynomial	1.06	2.68	2.67	1.50
Exponential	0.97	2.51	8.00	1.19
Laplacian	1.06	2.84	4.00	1.14
Rational quadratic	1.08	2.61	4.00	1.42
Sigmoid	0.98	2.40	2.67	1.53
Inverse multiquadric	0.77	2.61	4.00	1.61
Circular	1.03	2.58	4.00	1.64
Spherical	0.88	2.56	5.33	1.50

It can be seen from Tab. 2 that the performance of PCANet performs better than KPCANet in handwritten digit recognition generally, while the latter outperforms the former in face recognition and object recognition.

2.2 Face recognition on AR face dataset

The properties of KPCANet are tested by performing KPCANet on the AR dataset[11]. The AR dataset contains about 4 000 color images of size 165 × 120 pixel from

126 individuals. The subset of the data that contains 100 individuals including 50 males and 50 females is chosen. The color images are converted to gray scale ones. Each individual consists of two images with frontal illumination and neutral expression, which is used as the training samples. The other images including 24 images varying from illumination to disguise are used for testing.

The patch size and the block size are set to be 7×7 pixel and 8×8 pixel, respectively. The overlapping ratio of the block is 0.5. We compare the proposed KPCANet with LBP[12] and P-LBP[13] in Tab. 3. KPCANet with linear kernel function and Laplacian kernel function is used in this experiment. From Tab. 3, one can see that when the images only undergo the change of illumination, the testing accuracy rate achieves 100% with both linear kernel KPCANet and Laplacian kernel KPCANet. It is demonstrated that KPCANet is invariant to illumination. Besides, KPCANet outperforms LBP[12] and P-LBP[13] on different expressions and disguises under various illumination conditions, showing that KPCANet is robust to small deformation and occlusion.

Tab. 3 Comparison of accuracy rates of the methods on the AR face database %

Test sets	Illumination	Expression	Occlusion with illumination
LBP[12]	93.83	81.33	83.50
P-LBP[13]	97.50	80.33	90.05
KPCANet (linear kernel)	100.00	95.67	99.50
KPCANet (Laplacian kernel)	100.00	94.33	99.59

2.3 KPCANet with various stages in AR face dataset

KPCANet, which cascades different numbers of the KPCA filter bank layer and a pooling layer, is performed with the AR face dataset used in Section 2.2, and all images are cropped to size 32×32 pixel. Linear kernel, sigmoid kernel and circular kernel are chosen here in order to simplify the results. The patch size and the block size are set to be 7×7 pixel and 8×6 pixel, respectively. The overlapping ratio of the block is 0.5. The results are shown in Tab. 4.

Tab. 4 Comparison of accuracy rates of KPCANet with different number of stages on the AR face dataset %

Filter bank layer number	Linear kernel	Sigmoid kernel	Circular kernel
1	94.67	94.67	92.00
2	96.00	94.67	94.67
3	96.00	97.33	96.00

From Tab. 4, we can see that the accuracy rate increases as the number of KPCA filter bank layers increases in the KPCANet, however, the training time grows exponentially at the same time.

3 Conclusion

In this paper, we propose the KPCANet, which is an extension of PCANet, for image classification. The proposed KPCANet cascades kernel PCA, binary hashing and block-wise histogram. Experiments prove that KPCANet with different kernel functions is stable in general and also is invariant to illumination and stable to slight deformation and occlusion. Moreover, KPCANet is suitable for the recognition of handwritten images, face images and object images.

References

[1] LeCun Y, Kavukcuoglu K, Farabet C. Convolutional networks and applications in vision[C]//*Proceedings of 2010 IEEE International Symposium on Circuits and Systems*. Paris, France, 2010: 253 − 256.

[2] Bruna J, Mallat S. Invariant scattering convolution networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, **35**(8): 1872 − 1886.

[3] Chan T H, Jia K, Gao S, et al. PCANet: a simple deep learning baseline for image classification? [J]. *arXiv preprint arXiv*: 1404.3606, 2014.

[4] Schölkopf B, Smola A, Müller K R. Kernel principal component analysis [C]//*International Conference on Artificial Neural Networks*. Lausanne, Switzerland, 1997: 583 − 588.

[5] Schölkopf B, Smola A, Müller K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. *Neural Computation*, 1998, **10**(5): 1299 − 1319.

[6] Bengio Y. Learning deep architectures for AI[J]. *Foundations and Trends® in Machine Learning*, 2009, **2**(1): 1 − 127.

[7] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, **86**(11): 2278 − 2324.

[8] Hull J J. A database for handwritten text recognition research[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994, **16**(5): 550 − 554.

[9] Georghiades A S, Belhumeur P N, Kriegman D J. From few to many: illumination cone models for face recognition under variable lighting and pose[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(6): 643 − 660.

[10] Nene S A, Nayar S K, Murase H. Columbia object image library (COIL-20), CUCS-005-96 [R]. New York: Department of Computer Science, Columbia University: 1996.

[11] Martinez A M, Benavente R. The AR face database, CVC technical report #24[R]. CVC, 1998.

[12] Ahonen T, Hadid A, Pietikäinen M. Face description with local binary patterns: application to face recognition [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(12): 2037 − 2041.

[13] Tan X, Triggs B. Enhanced local texture feature sets for face recognition under difficult lighting conditions[J]. *IEEE Transactions on Image Processing*, 2010, **19**(6): 1635 − 1650.

面向图像分类的核主成分分析网络

吴　丹[1,4]　伍家松[1,2,3,4]　曾　瑞[1,4]　姜龙玉[1,4]　Lotfi Senhadji[2,3,4]　舒华忠[1,4]

([1] 东南大学计算机网络和信息集成教育部重点实验室, 南京 210096)
([2]Institut National de la Santé et de la Recherche Médicale U 1099, Rennes 35000, France)
([3]Laboratoire Traitement du Signal et de l'Image, Université de Rennes 1, Rennes 35000, France)
([4] 中法生物医学信息研究中心, 南京 210096)

摘要:为了能够用线性分类器对非线性特征进行分类,同时提高图像的分类正确率,提出了一种核主成分分析网络(KPCANet). 首先通过核主成分分析算法将数据映射到高维空间中,使得数据线性可分,然后建立一个2层的KPCANet,提取出图像的主特征,最后将图像的主特征输入线性分类器中进行分类. 实验结果表明,KPCANet对于人脸识别、物体识别以及手写数字识别效果良好,其分类效果优于现存的主成分分析网络(PCANet). 同时,KPCANet的成分提取效果不受光照条件变化的影响,且对于遮挡以及微小的形变提取效果稳定.

关键词:深度学习;核主成分分析网络;主成分分析网络;人脸识别;物体识别;手写数字识别
中图分类号:TP391