

# 基于主成分分析网络的改进图像分类算法

赵小虎<sup>1,2</sup>, 尹良飞<sup>1,3</sup>, 朱亚楠<sup>4</sup>, 刘鹏<sup>1,2\*</sup>, 王学奎<sup>1,3</sup>, 沈雪茹<sup>1,3</sup>

<sup>1</sup> 矿山互联网应用技术国家地方联合工程实验室, 江苏 徐州 221008;

<sup>2</sup> 中国矿业大学物联网(感知矿山)研究中心, 江苏 徐州 221008;

<sup>3</sup> 中国矿业大学信息与控制工程学院, 江苏 徐州 221116;

<sup>4</sup> 微软(中国)有限公司, 北京 100080

**摘要** 针对深层卷积神经网络模型的训练复杂、调参技巧和经验要求高、理论分析难等问题,提出一种训练效率高、可解释性强以及理论分析简单的图像分类算法。基于主成分分析网络(Principal Component Analysis Network, PCANet)进行特征提取,并采用宽度神经网络(Flat Neural Network, FNN)分类图像,最后通过直接计算得到模型参数。根据训练数据集自适应决定宽度神经网络节点数目,增加节点时不需要重新训练,只需要调整局部参数。实验表明,该模型能够快速训练,较其他非监督分类算法以及传统深层神经网络,该模型在识别准确率方面具有较强的竞争力。

**关键词** 图像处理; 主成分分析网络; 宽度神经网络; 快速训练; 可解释性

中图分类号 TP391.4

文献标识码 A

doi: 10.3788/LOP56.021004

## Improved Image Classification Algorithm Based on Principal Component Analysis Network

Zhao Xiaohu<sup>1,2</sup>, Yin Liangfei<sup>1,3</sup>, Zhu Yanan<sup>4</sup>, Liu Peng<sup>1,2\*</sup>, Wang Xuekui<sup>1,3</sup>, Shen Xueru<sup>1,3</sup>

<sup>1</sup> National and Local Joint Engineering Laboratory of Internet Application Technology on Mine, Xuzhou, Jiangsu 221008, China;

<sup>2</sup> Internet of Things Perception Mine Research Centre, China University of Mining and Technology, Xuzhou, Jiangsu 221008, China;

<sup>3</sup> School of Information and Control Engineering, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China;

<sup>4</sup> Microsoft (China), Beijing 100080, China

**Abstract** Aiming at the known deficiencies with complex training, strict parameter-tuning skills and experiences, difficult theoretical analysis of deep neural networks, an improved image classification algorithm with high training efficiency, strong interpretability and simple theoretical analysis is proposed, in which the principal component analysis network (PCANet) is used for feature extraction and the flat neural network (FNN) is for classification. In addition, the model parameters can be obtained by direct calculation and the flat neural network adaptively determines the number of nodes according to the training dataset. When the nodes increase, it is not necessary to retrain the model and only the parameters need to be adjusted locally to update the model. The experimental results show that the proposed model can acquire rapid training. Moreover, it possesses more competition in recognition accuracy compared with other unsupervised classification algorithms and traditional deep neural networks.

**Key words** image processing; principal component analysis network; flat neural network; fast training; interpretability

**OCIS codes** 100.2960; 150.1135; 100.4996

收稿日期: 2018-06-27; 修回日期: 2018-07-25; 录用日期: 2018-07-30

基金项目: 国家重点研发计划(2017YFC0804400)

\* E-mail: 13814538110@163.com

# 1 引言

近十年来,深度学习在计算机视觉领域取得了巨大成功。它能够从原始的像素出发,通过多层结构自动学习有效特征,用于图像识别或实现其他视觉任务。在图像识别算法中,最常用的深度学习结构是卷积神经网络(Convolution Neural Network, CNN)。最为著名的是Lecun等<sup>[1]</sup>提出的卷积神经网络模型之一,即LeNet,模型简单,仅仅包含卷积层以及池化层。2006年Hinton等<sup>[2]</sup>提出了一种深度学习结构,将图像识别带入一个全新的阶段。

2012年,Krizhevsky等<sup>[3]</sup>提出了AlexNet,通过加深CNN模型的深度,构造出8层网络结构,并采用ReLU和dropout技术,获得了当年ILSVR(Large Scale Visual Recognition Challenge, ILSVR)的冠军。相比于AlexNet,Szegedy等<sup>[4]</sup>极大地增加了CNN的深度,构建了一个22层的深度卷积网络——GoogLeNet,在ImageNet上的识别率将top5错误率降到了6.67%,获得了ILSVRC 2014的冠军。2015年,微软构造了一个具有152层的深度残差图像识别网络(Residual Networks, ResNet)<sup>[5]</sup>,在ImageNet上错误率降到了3.57%,获得了当年ILSVRC的冠军。

随着卷积神经网络层数的不断增加,模型能力逐渐增强,识别效果也逐渐变好。但模型越深,训练越不易收敛,并且对计算性能要求也极高,训练效率低下;反之,当层数比较少的时候识别率降低。且模型层数越多则模型复杂度越高,训练时对数据集要求也较高。另外,对于复杂的模型而言,除了训练效率极低外,模型理论分析较为困难,训练时往往依靠丰富的调参技巧和经验。

Coates等<sup>[6]</sup>指出,即使是简单的双层网络,也可以通过增加模型的宽度代替模型的深度,进而获得满意的效果。Chen等<sup>[7]</sup>提出宽度神经网络模型(Flat Neural Network, FNN),通过自适应增加模型宽度来调整模型网络结构,以达到优秀的预测效果。其网络结构包括输入层、输出层和隐含层共3层,其中隐含层由特征映射节点和增强节点组成。FNN的网络结构特点使得输入输出具有线性形式,其权值通过计算一步(One-Step)更新,训练迅速,广泛应用于分类任务<sup>[8-10]</sup>和时间序列预测<sup>[11-12]</sup>。

2014,Chan等<sup>[13]</sup>提出新的深度特征提取框架——主成分分析网络(Principal Component Analysis Network, PCANet)。该方法由主成分分

析(Principal Component Analysis, PCA)、哈希运算和分块直方图3个基本操作组成。PCA是一种特征抽取的方法,将原特征进行线性变换,映射至低纬度空间中,具有简化复杂问题的优点,广泛地应用于各类光学图像的处理任务<sup>[14-18]</sup>。PCANet通过将PCA进行两级级联构造多级滤波器,然后由二进制哈希算法进行索引,最后用分块直方图进行合并,输出提取特征。通过PCA选择卷积核的参数,直接计算得到模型参数,模型训练简单且非常易于进行理论分析。该方法在许多图像分类任务中表现优异<sup>[19-23]</sup>,已成为图像识别领域的一个研究热点。

本文针对上述深层卷积神经网络模型的训练复杂、调参技巧和经验要求高、理论分析难等缺点,结合PCANet和FNN,提出一种可以快速调参并且易于进行理论分析的神经网络图像分类模型。该模型利用PCANet进行分类图像的特征提取,并将PCANet的输出作为宽度神经网络的输入,进而通过宽度神经网络分类。所提模型具备以下优点:1)改善了传统深层卷积神经网络模型复杂的参数更新机制,使其调参简单,模型训练速度极快;2)可以自适应动态调节宽度神经网络的节点个数,以调节网络结构;3)通过直接计算得到参数,理论分析简单。

## 2 基于主成分分析网络的改进图像分类模型

### 2.1 PCANet

PCANet旨在设计一个可通过计算得到模型参数的深度神经网络,由3部分组成:输入层、PCA滤波器层和输出层。其中,PCA滤波器层由两级级联的卷积神经网络组成,卷积核的参数通过PCA直接得到;输出层由二进制哈希算法和分块直方图组成。利用输出层得到的特征向量来训练宽度神经网络模型,最后用于图像分类,具体模型架构如图1所示。

对于输入图片(大小为 $m \times n$ ),首先在每个像素周围进行一次 $k_1 \times k_2$ 的块采样,用所有的采样块级联表示该图片,第 $i$ 张图片可以表示为 $\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,mn} \in \mathbf{R}^{k_1 k_2}$ ,其中 $\mathbf{x}_{i,j}$ 表示第 $i$ 张图片的第 $j$ 个块(patch)的向量化, $k_1 k_2$ 表示 $k_1 \times k_2$ 的值;然后对采样的块进行去均值化处理: $\bar{\mathbf{X}}_i = [\bar{\mathbf{x}}_{i,1}, \bar{\mathbf{x}}_{i,2}, \dots, \bar{\mathbf{x}}_{i,mn}]$ ,其中 $mn$ 表示 $m \times n$ 的值;最后对训练集中的其他图片作同样处理,得到处理后的训练样本矩阵表示为

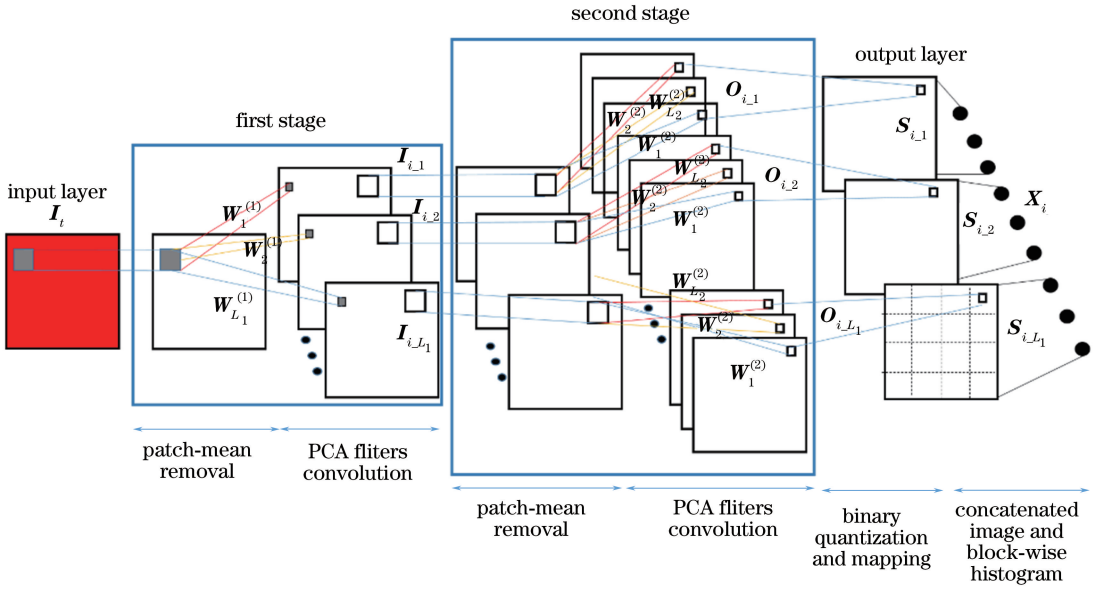


图1 PCANet 结构图

Fig. 1 Structural diagram of PCANet

$$\mathbf{X} = [\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \dots, \bar{\mathbf{X}}_N] \in \mathbf{R}^{k_1 k_2 \times Nmn}, \quad (1)$$

式中:  $Nmn$  表示  $N \times m \times n$  的值。

假定在第  $i$  层的滤波器数量为  $L_i$ , PCA 算法的目的是通过寻找一系列的标准正交矩阵来最小化重构误差, 即

$$\min_{\mathbf{V} \in \mathbf{R}^{k_1 k_2 \times L_1}} \|\mathbf{X} - \mathbf{V}\mathbf{V}^T \mathbf{X}\|_F^2, \text{ s.t. } \mathbf{V}^T \mathbf{V} = \mathbf{I}_{L_1}, \quad (2)$$

式中:  $\|\cdot\|_F^2$  表示矩阵的 Frobenius 范数。对应的 PCA 滤波器可表示为

$$\mathbf{W}_l^{(1)} = \text{mat}_{k_1, k_2} [q_l(\mathbf{X}\mathbf{X}^T)] \in \mathbf{R}^{k_1 \times k_2}, \quad (3)$$

式中:  $\text{mat}_{k_1, k_2}(\mathbf{v})$  为将  $\mathbf{v} \in \mathbf{R}^{k_1 k_2}$  映射到  $\mathbf{w} \in \mathbf{R}^{k_1 \times k_2}$ ;  $q_l(\mathbf{X}\mathbf{X}^T)$  为  $\mathbf{X}\mathbf{X}^T$  的第  $l$  个主特征向量,  $l = 1, 2, \dots, L_i$ 。

第 2 层的映射和第 1 层的映射机制基本相同, 第 1 层 PCA 映射输出可表示为

$$\mathbf{I}_{i,l} = \mathbf{I}_i * \mathbf{W}_l^{(1)}, i = 1, 2, \dots, N, \quad (4)$$

式中:  $*$  表示卷积运算。在卷积映射之前需要对样本进行边缘补零操作, 以保证映射的结果与源图像尺寸相同。与第 1 层所作的分块操作相同, 对像素  $\mathbf{I}_{i,l}$  取块、向量化、去均值化, 得到  $\bar{\mathbf{Y}}_{i,l} = [\bar{y}_{i,1,l}, \bar{y}_{i,2,l}, \dots, \bar{y}_{i,mn,l}] \in \mathbf{R}^{k_1 k_2 \times mn}$ 。对输入矩阵进行上述操作, 最终得到第 2 层输入块采样:  $\mathbf{Y}_l = [\bar{\mathbf{Y}}_{1,l}, \dots, \bar{\mathbf{Y}}_{N,l}] \in \mathbf{R}^{k_1 k_2 \times Nmn}$ 。假定在第 2 层的滤波器数量为  $L_2$ , 第 2 层 PCA 滤波器统一通过选取协方差矩阵对应的特征来组成, 即

$$\mathbf{W}_l^{(2)} = \text{mat}_{k_1, k_2} [q_l(\mathbf{X}\mathbf{X}^T)] \in \mathbf{R}^{k_1 \times k_2}, \quad (5)$$

式中:  $l = 1, 2, \dots, L_2$ 。由于第 1 层具有  $L_1$  个滤波

器核, 每张图像经过第 1 层会产生  $L_1$  个输出矩阵, 第 1 层输出的每一个特征矩阵在第 2 层中都将产生  $L_2$  个特征输出。最终对于每一张样本, 二阶 PCANet 都会产生  $L_1 \times L_2$  个输出的特征矩阵:

$$\mathbf{O}_{i,l} = \{\mathbf{I}_{i,l} * \mathbf{W}_l^{(2)}\}_{l=1}^{L_2}. \quad (6)$$

对第二层的输出矩阵进行二值化处理, 得到只包含整数和零的结果, 然后对其进行二值化哈希编码, 编码位数为  $L_2$ , 即

$$\mathbf{S}_{i,l} = \sum_{l=1}^{L_2} 2^{l-1} H(\mathbf{S}_{i,l} * \mathbf{W}_l^{(2)}), \quad (7)$$

式中:  $H(\cdot)$  是单位阶跃函数。经过上述处理, 在  $L_2 = 8$  时, 每个像素值被编码成 0 至 255 之间的整数。

将每个矩阵分成  $B$  块, 统计计算每个直方图信息, 表示为  $B_h(\cdot)$ , 即将  $B$  个直方图组合成一个向量。最后将各个块的直方图特征进行级联, 最终得到块扩展的直方图特征:

$$f_i = [B_h(\mathbf{S}_{i,1})],$$

$$B_h(\mathbf{S}_{i,2}), \dots, B_h(\mathbf{S}_{i,L_1})] \in \mathbf{R}^{(2^{L_2})L_1 B}. \quad (8)$$

## 2.2 宽度神经网络

宽度神经网络模型结构如图 2 所示。该网络除了原始输入经过一层神经网络映射外还包含许多增强节点, 组成网络输出层的增强输入。如图 2(a) 所示, 首先对输入节点进行线性组合, 然后应用非线性激活函数  $\xi(\cdot)$  构造增强节点。K 个增强层节点的宽度神经网络可以表示为

$$Y = [x \mid \xi(xW_h + \beta_h)]W, \quad (9)$$

式中： $W_h$  为随机生成的增强层权重矩阵； $W$  为训练得到的权重矩阵； $Y$  为输出矩阵； $x$  为输入矩阵； $\beta_h$  为  $-1$  或者  $+1$  的常数偏差节点。

若用矩阵  $A$  表示  $[x \mid \xi(xW_h + \beta_h)]$ ，则整个模型如图 2(b)所示，可以表示为

$$Y = AW. \quad (10)$$

本研究提出的宽度神经网络模型示意图如图 3 所示。宽度神经网络主要包括特征映射层和增强层，PCANet 输出  $X$  作为宽度神经网络的输入。特征映射层由  $p$  个节点组成，可表示为

$$Z = \varphi(XW_1 + \alpha), \quad (11)$$

式中： $\varphi(\cdot)$  为非线性激活函数； $\alpha$  为常数偏置； $W_1$  权重矩阵。特征映射层的输出作为增强层的输入，可表示为

$$T = \varphi(ZW_2 + \beta), \quad (12)$$

式中： $\varphi(\cdot)$  为非线性激活函数； $\beta$  为常数偏置； $W_2$  权重矩阵。特征映射层动态增加  $s$  个节点，使得训练模型性能达到最优。特征映射层节点和增强层节点共同组成输入矩阵  $A = [Z_1, Z_2, \dots, Z_p, T_1, \dots, T_q]$ ，更新权重  $W$ ，利用 (10) 式求解输出  $Y$ ，最终得到每个样本的类别输出。

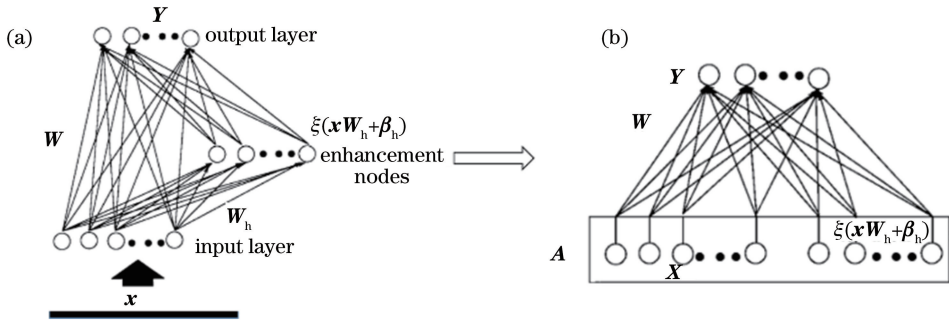


图 2 宽度神经网络示意图

Fig. 2 Schematic of flat neural network

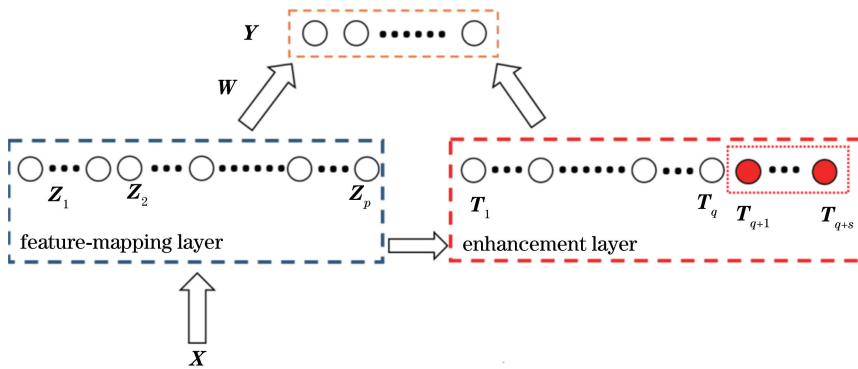


图 3 宽度神经网络模型更新示意图

Fig. 3 Schematic of flat neural network model update

### 2.3 整体流程

算法的整体流程主要分为 PCA 扫描和宽度神经网络训练。输入图像经过 PCA 扫描生成输入特征向量，然后进行宽度神经网络模型训练，以得到该图像所属的类别。具体如图 4 所示。

假设输入  $N$  张大小为  $m \times n$  的训练图片，扫描窗口大小为  $k_1 \times k_2$ ，通过 Padding 操作，扫描生成  $m \times n \times N$  个  $k_1 \times k_2$  维的特征向量，对每个特征向量去平均后进行 PCA，提取  $L_1$  个主成分特征向量作为第 1 阶段  $L_1$  个卷积核的参数（卷积核的大小即为  $k_1 \times k_2$ ），输入图片经过卷积得到  $N \times L_1$  张大

小为  $m \times n$  的输出图片，第 1 阶段扫描完成。第 2 阶段在第 1 阶段基础上同样进行扫描、主成分分析计算，得到  $L_2$  个主成分特征向量，并作为第 2 阶段卷积核的参数，第 2 个卷积层输出  $N \times L_1 \times L_2$  张大小为  $m \times n$  的输出，将这些输出分为  $N$  组（每张原始图片对应一组），并对每组输出进行 Hashing 运算以及直方图统计（Histogram），得到长度为  $2^{L_2} \times L_1 \times B$  的特征向量输出（其中  $B$  为对每张图片分块作 Hashing 运算以及直方图统计时分块的数目）。对于  $N$  张输入图片，得到  $N$  个长度为  $2^{L_2} \times L_1 \times B$  的特征向量。

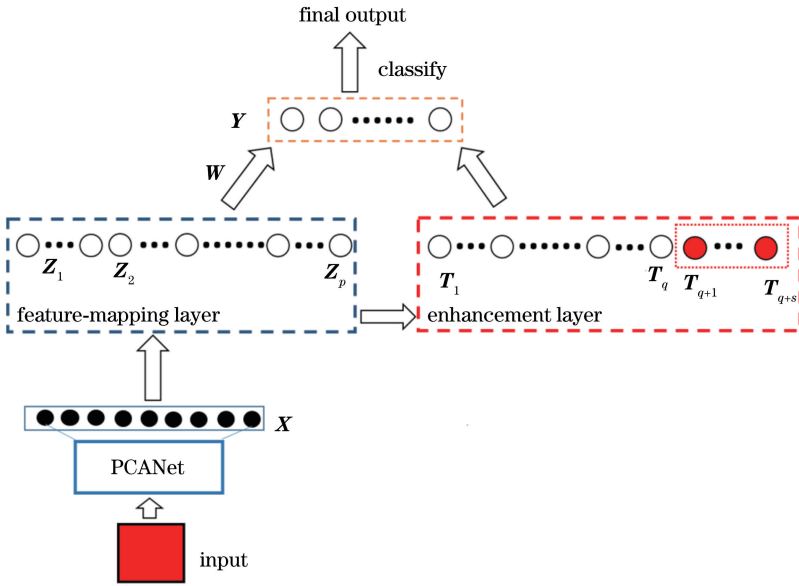


图4 整体流程图

Fig. 4 Overall flow chart

将训练样本经过 PCANet 得到的特征向量作为特征,原输入标签作为标签,训练宽度神经网络。在测试阶段,每个输入样本先后经过 PCANet、特征映射层以及增强层得到输出层输入  $\mathbf{A} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_p, \mathbf{T}_1, \dots, \mathbf{T}_q]$ , 并利用 softmax 得到最终的预测向量  $\mathbf{Y}$  以及类别输出。

### 3 模型训练和参数更新

#### 3.1 模型的训练过程

输入:训练样本集经过 PCANet 特征提取层(图1所示)生成的特征矩阵  $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N]$ , 其中  $\mathbf{X}_i (i=1, \dots, N)$  为每个样本经过 PCANet 所得特征向量,均为列向量, $N$  为样本个数。

输出:输出权重向量  $\mathbf{W}$  以及 softmax 分类。

第1步:将特征矩阵  $\mathbf{V}$  的每一列  $\mathbf{V}_i (i=1, \dots, N)$  作为一个样本生成包含  $p$  个节点的特征映射层,然后求解每个节点的本地权重  $\mathbf{W}_\alpha$  和偏置  $\alpha$ , 得到  $\mathbf{Z} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_p]$ 。

第2步:将特征映射层输出  $\mathbf{T}$  的每行同输入特征向量一起作为一个样本生成包含  $q$  个节点的增强层,求解每个节点的本地权重  $\mathbf{W}_\beta$  和偏置  $\beta$ , 得到  $\mathbf{T} = [\mathbf{T}_1, \mathbf{T}_2, \dots, \mathbf{T}_q]$ 。

第3步:根据每个节点的本地权重和偏置求解矩阵  $\mathbf{A} = [\mathbf{Z}_1, \mathbf{Z}_2, \dots, \mathbf{Z}_p, \mathbf{T}_1, \dots, \mathbf{T}_q]$ , 输出权重向量  $\mathbf{W} = [\mathbf{W}_1, \mathbf{W}_2, \dots, \mathbf{W}_{p+q}]^T$ 。

第4步:利用第2步增加  $s$  个增强层节点,更新  $\mathbf{W}$ 。

第5步:对每个样本求解输出层输入矩阵  $\mathbf{A}$ , 利用式  $\mathbf{Y} = \mathbf{A}\mathbf{W}$  求解每个样本最终类别向量输出,并计算得到预测准确率,准确率不再上升则停止训练,否则返回第4步。

#### 3.2 输出权重更新方法

对权重矩阵的逐步更新可以通过对输入矩阵  $\mathbf{A}_n$  进行 Q-R 分解。假设存在矩阵  $\mathbf{Q}$  和  $\mathbf{R}$  使得  $\mathbf{A}_n \triangleq \mathbf{Q}\mathbf{R}$ , 其中  $\mathbf{Q}$  是一个正交矩阵, $\mathbf{R}$  是一个上三角矩阵。当新增加节点时, $\mathbf{R}$  和  $\mathbf{Q}$  可以通过 Givens rotations 得到, 则  $\mathbf{A}_{n+1} \triangleq \hat{\mathbf{Q}}\hat{\mathbf{R}}$ , 其中  $\hat{\mathbf{Q}}$  仍然是正交矩阵,而  $\hat{\mathbf{R}}$  也是上三角矩阵,分别由  $\mathbf{Q}$  和  $\mathbf{R}$  通过 Givens rotations 得到。而  $\mathbf{A}_{n+1}$  的伪逆矩阵可以表示为

$$\mathbf{A}_{n+1}^+ \triangleq \hat{\mathbf{R}}^+ \hat{\mathbf{Q}}', \quad (13)$$

最终可以求得  $\mathbf{W}_{n+1}$ , 具体步骤如下。

输入: $\mathbf{A}_n \triangleq \mathbf{Q}\mathbf{R}$ , 向量  $\mathbf{a}$ , 权重矩阵  $\mathbf{W}_n$ , 其中  $\mathbf{A}_n$  是  $n \times m$  的矩阵, $\mathbf{Q}$  是  $n \times n$  正交方阵, $\mathbf{R}$  是  $n \times m$  的上三角矩阵, $\mathbf{a}'$  是一个  $m \times 1$  的行向量。

输出: $\mathbf{A}_{n+1} \triangleq \begin{bmatrix} \mathbf{A}_n \\ \mathbf{a}' \end{bmatrix} \triangleq \hat{\mathbf{Q}}\hat{\mathbf{R}}$ , 权重矩阵  $\mathbf{W}_{n+1}$ , 其中

$\mathbf{A}_{n+1}$  是  $(n+1) \times m$  的矩阵, $\hat{\mathbf{Q}}$  是  $(n+1) \times (n+1)$  的正交矩阵,而  $\hat{\mathbf{R}}$  是  $(n+1) \times m$  的上三角矩阵。

第1步:扩展  $\hat{\mathbf{Q}}$  和  $\hat{\mathbf{R}}$ , 即

$$\hat{\mathbf{Q}} \leftarrow \text{diag}(\mathbf{Q}, 1) = \begin{bmatrix} \mathbf{Q} & 0 \\ 0 & 1 \end{bmatrix}, \hat{\mathbf{R}} \leftarrow \begin{bmatrix} \mathbf{R} \\ \mathbf{a}' \end{bmatrix}, \quad (14)$$

第 2 步: for  $i=1$  to  $m$  do,

$$\mathbf{J}_i = \text{Rot}(\mathbf{r}_i, \mathbf{e}_i), \quad (15)$$

$$\hat{\mathbf{Q}} \leftarrow \hat{\mathbf{Q}} * \mathbf{J}'_i, \quad (16)$$

$$\hat{\mathbf{R}} \leftarrow \mathbf{J}_i * \hat{\mathbf{Q}}, \quad (17)$$

第三步:  $\hat{\mathbf{R}}$  是上三角矩阵, 那么  $\mathbf{W}_{n+1}$  很容易求得

$$\hat{\mathbf{R}}\mathbf{W}_{n+1} = \hat{\mathbf{Q}}' \begin{bmatrix} \mathbf{Y}_n \\ \mathbf{y}' \end{bmatrix}. \quad (18)$$

由此可见, 不同于传统卷积神经网络模型, 通过梯度下降法等复杂传播机制求解模型参数, 本算法的模型训练完全通过前向计算求解模型参数, 因此模型可以快速得到训练。

## 4 实验与分析

为验证本文算法的有效性, 在不同的训练集: MNIST 数据集<sup>[24]</sup>、Cifar-10 数据集<sup>[25]</sup>、AR 数据集<sup>[26]</sup>以及 FERET 数据集<sup>[27]</sup>上进行实验, 以验证所提出算法的分类性能; 并将其与无监督分类算法以及传统深度卷积神经网络的模型进行比较, 以验证本文模型除了具有快速训练的优点外, 还可以得到较高的识别准确率。另在 MNIST 数据集上, 对比不同的分类算法, 比较算法的有效性。实验结果均为多次实验后得到的平均识别率, 以上实验在 Inter (R)Core(TM)i5-6700, CPU 主频 3.4 GHz, 16 G 内存的 windows 操作系统, Matlab2017b 以及 python 集成环境 anaconda3 环境下运行的。

### 4.1 参数设置

参数设置实验主要基于 MNIST 数据集, 主要包括 PCANet 以及宽度神经网络的参数设置。对于 PCANet 可设置的参数主要包括两层滤波器个数  $L_1$ 、 $L_2$ , 滤波器大小  $k_1$ 、 $k_2$  等, 其中两层滤波器个数对分类准确率的提升相关度较大; 宽度神经网络主要包括特征映射层节点个数。

对于 PCANet 两层滤波器个数进行参数设置实验, 首先进行第 1 层的参数实验, 对于第 1 层滤波器个数  $L_1$  从 2 到 12 依次变化进行准确率比较实验; 在确定第 1 层滤波器个数最佳的情况下对第 2 层滤波器个数  $L_2$  从 2 到 12 依次变化选择最佳参数, 如图 5、6 所示。由实验结果可知, 设置  $L_1=8$ 、 $L_2=4$  情况下分类效果最佳。

在宽度神经网络参数设置实验中, 将特征映射层节点个数从 4 至 20 依次变化观察准确率变化。如图 7 所示, 当特征映射层的节点数量不够大时, 通过增加数量可以提高准确率。但是, 当节点数达到

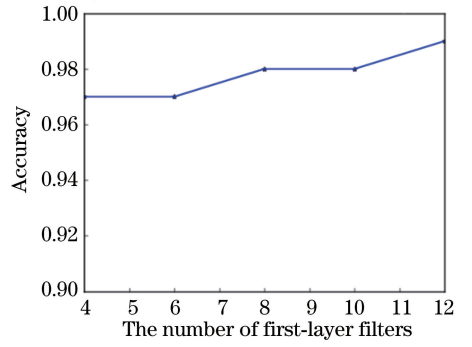


图 5 第 1 层滤波器个数对分类效果影响

Fig. 5 Influence of number of first-layer filters on classification effect

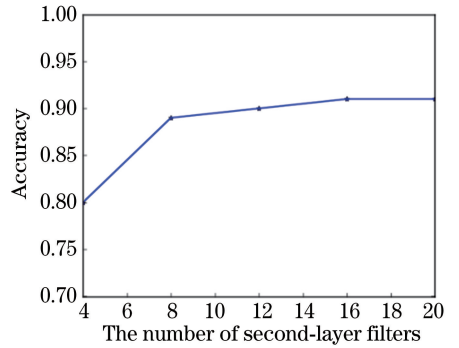


图 6 第 2 层滤波器个数对分类效果影响

Fig. 6 Influence of number of second-layer filters on classification effect

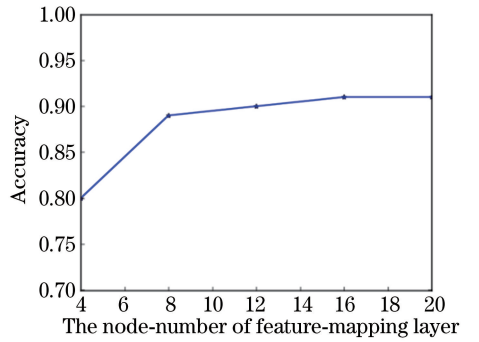


图 7 特征映射层节点对分类准确率影响

Fig. 7 Influence of node-number of feature-mapping layer on classification accuracy

阈值时, 准确率提高不再明显。由实验结果可知, 特征映射层节点个数达到 12 时效果较佳。

### 4.2 识别率实验

本文在 MNIST 数据集、Cifar-10 数据集、AR 数据集以及 FERET 数据集进行识别率测试实验, 将本文算法与非监督分类算法以及传统深度学习模型的识别准确率作对比。

#### 4.2.1 MNIST 数据集实验

MNIST 数据集是一个手写字识别数据集, 图

片大小为  $28 \times 28$ , 其中包含 60000 个训练样本以及 10000 个测试样本。像素值为 0 表示白色, 像素值为 255 表示黑色, 示例图像如图 8 所示。

表 1 给出在 MNIST 数据集上, 本方法与不同的分类算法进行字体识别率对比。其中 PCANet 模型参数在不同数据集取最佳参数, 卷积核大小  $k_1 = k_2 = 5$ , 设置滤波器参数分别为  $L_1 = 8, L_2 = 4$ 。由实验可以看出, 相对于其他模型, 本文算法的识别准确率在 MNIST 数据集上优于其他算法, 达 99.27%。

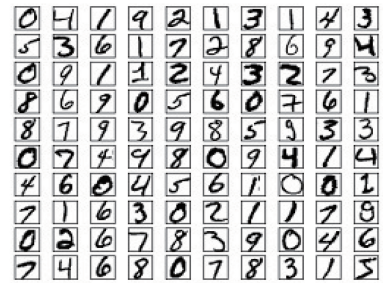


图 8 MNIST 数据集示例

Fig. 8 Examples in MNIST dataset

表 1 MNIST 数据集中数字识别率

Table 1 Recognition accuracies of numbers in MNIST dataset

Method	HSC <sup>[28]</sup>	K-NN-SCM <sup>[29]</sup>	CDBN <sup>[30]</sup>	ConvNet <sup>[31]</sup>	Stochastic pooling ConvNet <sup>[32]</sup>	Conv. Maxout+ Dropout <sup>[33]</sup>	ScatNet-2 (SVMrbf) <sup>[34]</sup>
Accuracy / %	77.00	63.00	82.00	53.00	47.00	45.00	43.00
Method	PCANet <sup>[35]</sup>	DNN (Le-Net5) <sup>[35]</sup>	PCANet+	PCANet+	PCANet+	PCANet+	Proposed
Accuracy / %	94.00	95.05	88.24	72.68	70.18	98.48	99.27

4.2.2 Cifar-10 数据集实验

Cifar-10 数据集由 Alex Krizhevsky, Vinod Nair 和 Geoffrey Hinton 收集而来, 包含 50000 张  $32 \times 32$  大小的训练图片以及 10000 张  $32 \times 32$  大小的测试图片, 数据集图片包含 airplane, automobike 等 10 个类别的物体, 不同类别图像示例如图 9 所示。

表 2 给出在 Cifar-10 数据集上, 本方法与不同的分类算法进行字体识别率对比。其中 PCANet 模型参数在不同数据集取最佳参数, 卷积核大小  $k_1 = k_2 = 5$ , 设置滤波器参数分别为  $L_1 = 40, L_2 = 8$ 。由实验可以看出, 本文方法的识别准确率明显优于其他分类算法。



图 9 Cifar-10 数据集示例

Fig. 9 Examples in Cifar-10 dataset

表 2 Cifar-10 数据集中物体识别率

Table 2 Prediction accuracies of targets in Cifar-10 dataset

Method	Tiled CNN <sup>[36]</sup>	Improved LCC <sup>[37]</sup>	KDES- A <sup>[38]</sup>	Stochastic pooling ConvNet <sup>[32]</sup>	CNN+ Spearmint <sup>[39]</sup>	Conv. Maxout+ Dropout <sup>[33]</sup>	NIN <sup>[40]</sup>
Accuracy / %	73.10	74.50	76.00	84.87	85.02	88.32	89.59
Method	DNN (Le-Net5)	PCANet+	PCANet+	PCANet+	PCANet +SVM	Proposed	
Accuracy / %	83.39	79.67	75.20	74.01	87.50	99.01	

4.2.3 AR 数据集实验

AR 数据集是人脸数据集, 包括 126 人(70 男和 56 女) 超过 4000 张的彩色照片, 图片的大小为  $768 \times 576$ , 示例人脸图像如图 10 所示。

表 3 给出在 AR 数据集上, 本方法与不同的分类算法进行字体识别率对比。其中 PCANet 模型参数在不同数据集取最佳参数, 卷积核大小  $k_1 = k_2 = 5$ , 设置滤波器参数分别为  $L_1 = 8, L_2 = 8$ 。由实验可以看出, 相对于其他模型, 本文算法的准确率在

AR 数据集有所提高。

4.2.4 FERET 数据集实验

FERET 数据集是人脸数据集, 该姿态库由 1196 个人的 14051 幅图像组成。所有的图像被分成了 4 个部分, Fb 部分包含不同表情的 1195 张图片, 图像与目标集合中的图像是在同日、同光照条件下拍摄; Fc 部分包含不同光照程度的 194 张图片, 图像与目标集合中的图像是在同日、不同光照的条件下拍摄; Dup1 部分由 722 张图片组成, 图像与目

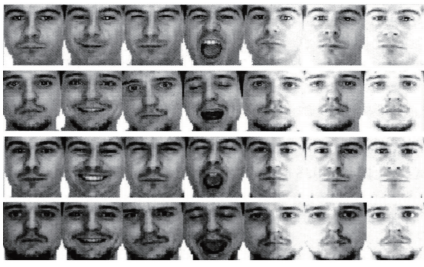


图 10 AR 数据集示例

Fig. 10 Examples in AR dataset

表 3 AR 数据集中人脸识别率

Table 3 Recognition accuracies of faces in AR dataset

Method	LBP <sup>[41]</sup>	P-LBP <sup>[42]</sup>	PCANet <sup>[35]</sup>	DNN (Le-Net5)	PCANet+ AdaBoost	PCANet+ RandomForest	PCANet+ DecisionTree	PCANet+ SVM	Ours
Accuracy / %	81.33	80.33	85.00	80.06	73.59	62.59	57.20	75.59	85.21

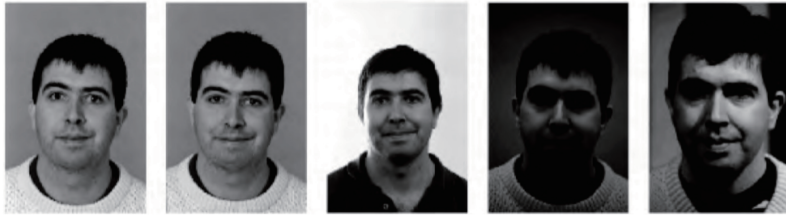


图 11 FERET 数据集示例

Fig. 11 Examples in FERET dataset

表 4 FERET 数据集中人脸识别率

Table 4 Recognition accuracies of faces in

FERET dataset %

Dataset	Fb	Fc	Dup1	Dup2
LBP <sup>[41]</sup>	93.00	51.00	61.00	50.00
P-LBP <sup>[42]</sup>	98.00	98.00	90.00	85.00
DMMA <sup>[43]</sup>	98.10	98.50	81.60	83.20
POEM <sup>[44]</sup>	99.60	99.50	88.80	85.00
PCANet+SVM	99.00	98.93	90.76	90.27
DNN (Le-Net5)	99.05	98.79	91.27	90.52
Proposed	99.40	99.90	93.29	89.90

### 4.3 训练效率实验

为突出算法可快速训练的优点,针对 MNIST 数据集,对比不同分类算法的识别准确率和训练时间,见表 5。由实验结果可知,本文算法相对于其他算法训练速度更快,且分类的识别率更高。

本文主要从算法的准确率和训练效率两个角度验证算法的优异性。分别在 MNIST、Cifar-10、AR 数据集以及 FERET 数据集上对比本文算法和其他分类模型的识别率,以及在 MNIST 数据集上对比不同模型的训练时间。结果表明,所提出的算法在保证较高的预测准确率的同时可以快速训练模型的优点。

标集合中的图像是在 234 d 后、不同照相机的条件下拍摄;Dup2 部分图像与目标集合中的图像是在 18 个月后、不同照相机的条件下拍摄。其中,训练数据集包含 433 个人,每个人含有两张照片。示例人脸图像如图 11 所示。

识别的准确率如表 4 所示。PCANet 的卷积核大小设置为  $k_1=k_2=5$ ,滤波器参数  $L_1=8, L_2=8$ 。由实验结果可知,相对于其他分类模型,本文算法在表情、光照和时间跨度均具有较好的稳健性。

表 5 MNIST 数据集上训练效率对比

Table 5 Training efficiency comparison in MNIST dataset

Method	Accuracy / %	Training time / s	Training time of PCANet / s	Training time of FNN / s
PCANet+	70.18	2219	1804	415
DecisionTree	72.68	2054	1726	328
RandomForest	88.24	2940	2143	797
PCANet+ AdaBoost	98.48	3605	2976	729
PCANet+ SVM	99.05	3195		
DNN (Le-Net5)	99.78	1739	1682	57
Proposed				

## 5 结 论

通过结合 PCANet 和 FNN 提出一种新的图像分类的方法,采用 PCANet 提取图像特征,进而利用提取得到的图像特征进行 FNN 训练,以得到分类预测。FNN 的增强层节点个数可以根据数据集进行动态调整,并且增强层的节点数能够根据数据集自适应调整。由实验结果可以看出,所提改进模



型较普通深度神经网络模型在提高图像分类准确率的同时,还具备模型训练快速的优点;此外,由于本文所研究网络均可通过计算得到模型参数,因此模型具有较强的可解释性,易于进行理论分析。

但对于各种参数的优化取值,由于计算量等原因,本文处理相对比较简单,未深入探讨。未来还需研究其他不同参数取值对分类效果的影响。同时也会尝试其他特征提取算法,如 KPCANet、ScatNet 等,以及更多的分类算法,在识别率和训练时间上优化算法,将其应用于更多类别的分类及更复杂的应用场景。

### 参 考 文 献

- [1] 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.
- [2] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786): 504-507.
- [3] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks [C] // *The Twenty-sixth Annual Conference on Neural Information Processing Systems (NIPS)*, December 3-8, 2012, Lake Tahoe, Nevada, USA. Cambridge, MA: MIT Press, 2012: 1097-1105.
- [4] Szegedy C, Liu W, Jia Y, *et al.* Going deeper with convolutions [C] // *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 7-12, 2015, Boston, MA, USA. New York: IEEE, 2015: 1-9.
- [5] He K, Zhang X, Ren S, *et al.* Deep residual learning for image recognition [J/OL]. *Computer Science: Computer Vision and Pattern Recognition*, 2016: 770-778. [2018-05-29]. <https://arxiv.org/abs/1512.03385>.
- [6] Coates A, Ng A Y Lee H. An analysis of single-layer networks in unsupervised feature learning [C] // *The fourteenth international conference on Artificial Intelligence and Statistics (AISTATS)*, April 11-13, 2011, Ft. Lauderdale, FL, USA. United States: JMLR, 2011, 15: 215-223.
- [7] Chen C L P, Wan J Z. A rapid learning and dynamic stepwise updating algorithm for flat neural networks and the application to time-series prediction[J]. *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, 1999, 29(1): 62-72.
- [8] Zhou Y C, Peng J T, Chen C L P. Extreme learning machine with composite kernels for hyperspectral image classification [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(6): 2351-2360.
- [9] Huang G B, Bai Z, Kasun L L C, *et al.* Local receptive fields based extreme learning machine [J]. *IEEE Computational Intelligence Magazine*, 2015, 10(2): 18-29.
- [10] Zhang L, Suganthan P N. A survey of randomized algorithms for training neural networks [J]. *Information Sciences*, 2016, 364-365: 146-155.
- [11] Qiu X, Suganthan P N, Amaratunga G A J. Electricity load demand time series forecasting with empirical mode decomposition based random vector functional link network[J]. *Proceedings of the IEEE*, 2016: 001394-001399.
- [12] Ren Y, Suganthan P N, Srikanth N, *et al.* Random vector functional link network for short-term electricity load demand forecasting [J]. *Information Sciences*, 2016, 367-368: 1078-1093.
- [13] Chan T H, Jia K, Gao S H, *et al.* PCANet: A simple deep learning baseline for image classification? [J]. *IEEE Transactions on Image Processing*, 2015, 24(12): 5017-5032.
- [14] Zuo C, Chen Q, Qu W J, *et al.* Phase aberration compensation in digital holographic microscopy based on principal component analysis[J]. *Optics Letters*, 2013, 38(10): 1724-1726.
- [15] Yeom S, Javidi B, Ferraro P, *et al.* Three-dimensional color object visualization and recognition using multi-wavelength computational holography [J]. *Optics Express*, 2007, 15(15): 9394-9402.
- [16] Jiao S M, Zhuang Z Y, Zou W B. Fast computer generated hologram calculation with a mini look-up table incorporated with radial symmetric interpolation [J]. *Optics Express*, 2017, 25(1): 112-123.
- [17] Du Q, Fowler J E. Hyperspectral image compression using JPEG2000 and principal component analysis [J]. *IEEE Geoscience and Remote Sensing Letters*, 2007, 4(2): 201-205.
- [18] Gao H, Cai J F, Shen Z, *et al.* Robust principal component analysis-based four-dimensional computed tomography [J]. *Physics in Medicine and Biology*, 2011, 56(11), 3181-3198.
- [19] Wu J S, Qiu S J, Zeng R, *et al.* Multilinear principal component analysis network for tensor object classification[J]. *IEEE Access*, 2017, 5: 3322-3331.
- [20] Han B, Jia Z H, Gao X B. Improved PCANet for

- aurora images classification [J]. Journal of Xidian University, 2017, 1: 83-88.
- 韩冰, 贾中华, 高新波. 改进的主成分分析网络极光图像分类方法[J]. 西安电子科技大学学报(自然科学版), 2017, 1: 83-88
- [21] Wu J, Qiu S, Zeng R, *et al.* PCANet for color image classification in various color spaces [C] // International Conference on Cloud Computing and Security (ICCCS), June 16-18, 2017, Nanjing, China. Cham, Switzerland: Springer, 2017: 494-505.
- [22] Wang Z, Bi Z, Wang C, *et al.* Traffic lights recognition based on PCANet [C] // 2015 Chinese Automation Congress (CAC), November 27-29, 2015, Wuhan, China. New York: IEEE, 2015: 559-564.
- [23] Cheng L Y, Mi G Y, Li S, *et al.* Quality diagnosis of joints in laser brazing based on principal component analysis-support vector machine model [J]. Chinese Journal of Lasers, 2017, 44(3): 0302004.  
程力勇, 米高阳, 黎硕, 等. 基于主成分分析-支持向量机模型的激光钎焊接头质量诊断[J]. 中国激光, 2017, 44(3): 0302004.
- [24] LeCun Y, Cortes C. MNIST handwritten digit database [EB/OL]. [2018-05-29]. <http://yann.lecun.com/exdb/mnist>, 2010.
- [25] Krizhevsky A. Learning multiple layers of features from tiny images [D]. Toronto, Canada: University of Toronto, 2009.
- [26] Sun Y, Chen Y, Wang X, *et al.* Deep learning face representation by joint identification-verification [C] // Twenty-eighth Conference on Neural Information Processing Systems (NIPS), December 8-11, 2014, Montreal, Canada. Cambridge, MA: MIT Press, 2014: 1988-1996.
- [27] Phillips P J, Wechsler H, Huang J, *et al.* The FERET database and evaluation procedure for face-recognition algorithms [J]. Image and Vision Computing, 1998, 16(5): 295-306.
- [28] Yu K, Lin Y, Lafferty J. Learning image representations from the pixel level via hierarchical sparse coding [J]. Proceedings of the IEEE, 2011: 1713-1720.
- [29] Belongie, Serge, Malik, *et al.* Shape context: a new descriptor for shape matching and object recognition [C] // The annual conference on Neural Information Processing Systems (NIPS), December 3-8, 2001, Vancouver, British Columbia, Canada. Cambridge, MA: MIT Press, 2001: 831-837.
- [30] Lee H, Grosse R, Ranganath R, *et al.* Unsupervised learning of hierarchical representations with convolutional deep belief networks [J]. Communications of the ACM, 2011, 54(10): 95-103.
- [31] Jarrett K, Kavukcuoglu K, LeCun Y. What is the best multi-stage architecture for object recognition? [C] // 2009 IEEE International Conference on Computer Vision (ICCV), September 29 - October 2, 2009, Kyoto, Japan. New York: IEEE, 2009: 2146-2153.
- [32] Zeiler M D, Fergus R. Stochastic pooling for regularization of deep convolutional neural networks [J]. Neural and Evolutionary Computing, Machine Learning, Preprint ArXiv, ArXiv: 1301.3557, 2013.
- [33] Goodfellow I J, Warde-Farley D, Mirza M, *et al.* Maxout Networks [Z/OL]. 2013 [2018-05-30]. <https://arxiv.org/pdf/1302.4389.pdf>.
- [34] Bruna J, Mallat S. Invariant scattering convolution networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1872-1886.
- [35] Chan T H, Jia K, Gao S H, *et al.* PCANet: A simple deep learning baseline for image classification? [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5017-5032.
- [36] Le Q V, Ngiam J, Chen Z, *et al.* Tiled convolutional neural networks [C] // The Twenty-Fourth Annual Conference on Neural Information Processing Systems (NIPS 2010), December 6-11, 2010, Vancouver, B. C., Canada. Cambridge, MA: MIT Press, 2010: 1279-1287.
- [37] Yu K, Zhang T. Improved local coordinate coding using local tangents [C] // International Conference on International Conference on Machine Learning (ICML), June 21-24, 2010, Haifa, Israel. New York: ACM, 2010: 1215-1222.
- [38] Bo L, Ren X, Fox D. Kernel descriptors for visual recognition [C] // The Twenty-Fourth Annual Conference on Neural Information Processing Systems (NIPS), December 6-11, 2010, Vancouver, Canada. Cambridge, MA: MIT Press, 2010: 244-252.
- [39] Snoek J, Larochelle H, Adams R P. Practical bayesian optimization of machine learning algorithms [C] // The Twenty-sixth Annual Conference on Neural Information Processing Systems (NIPS), December 6-11, 2012, Harrahs

- and Harveys, USA. Cambridge, MA: MIT Press, 2012: 2951-2959.
- [40] Pang Y, Sun M, Jiang X, *et al.* Convolution in Convolution for Network in Network [J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(5):1587-1597.
- [41] Ahonen T, Hadid A, Pietikainen 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.
- [42] Tan X Y, 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.
- [43] 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.
- [44] Vu N S, Caplier A. Enhanced patterns of oriented edge magnitudes for face recognition and image matching [J]. IEEE Transactions on Image Processing, 2012, 21(3): 1352-1365.