

DOI: 10.11830/ISSN.1000-5013.201703109



采用无监督学习算法与卷积的 图像分类模型

王改华^{1,2}, 李涛², 吕朦², 袁国亮²

(1. 湖北工业大学 太阳能高效利用湖北省协同创新中心, 湖北 武汉 430068;
2. 湖北工业大学 电气与电子工程学院, 湖北 武汉 430068)

摘要: 为了提高图像分类精度,降低训练复杂度,提出一种采用无监督学习算法与卷积构造的图像分类模型. 首先,从输入无标签图像中随机抽取大小相同的图像块构成数据集,进行预处理. 其次,将预处理后的图像块通过两次 K -means 聚类算法提取字典,并采用离散卷积操作提取最终图像特征. 最后,采用 Softmax 分类器对提取的图像特征进行分类,得出准确率. 将该模型与卷积神经网络(CNN),Dropout CNN 网络进行比较,结果表明:在对大规模高维图像分类上,文中模型具有分类精确度高、简单、训练参数少、适应度高等优点.

关键词: K -means 聚类; 图像分类; 卷积; 卷积神经网络; Dropout

中图分类号: TP 391.41 **文献标志码:** A **文章编号:** 1000-5013(2018)01-0146-06

Image Classification Model Using Unsupervised Learning Algorithm and Convolution

WANG Gaihua^{1,2}, LI Tao², LÜ Meng², YUAN Guoliang²

(1. Hubei Collaborative Innovation Centre for High-Efficiency Utilization of Solar Energy,
Hubei University of Technology, Wuhan 430068, China;
2. School of Electrical and Electronic Engineering, Hubei University of Technology, Wuhan 430068, China)

Abstract: To improve image classification accuracy and reduce the training complexity, a image classification model based on unsupervised learning algorithm and convolution is proposed. Above all, the dataset is composed of randomly extracted image patch of the same size from the input unlabeled images, and the dataset was pre-processing. Next, the preprocessed image patch is fed into twice K -means clustering algorithm to extract the dictionary, and the final image feature is extracted by the discrete convolution operation. Finally, the accuracy of image classification was obtained by classifying the extracted image features using Softmax classifier. Compared with the convolutional neural net (CNN) and Dropout CNN. The results shows that the proposed model has the advantages of high classification accuracy, simplicity, less training parameters and high adaptability for large-scale high-dimensional image classification.

Keywords: K -means clustering; image classification; convolution; convolutional neural net; Dropout

卷积神经网络因其具有独特的二维数据处理方式及便捷性,在图像分类领域得到了快速的发展和广泛的应用,是目前应用较为成熟的一种结构^[1-5]. Krizhevsky 等^[6]提出深度卷积神经网络,虽然将 Imagenet 的分类错误率降低了 9%,但存在计算成本高和过拟合的情况. Zeiler 等^[7-8]采用反卷积方法对

收稿日期: 2017-03-04

通信作者: 王改华(1979-),女,讲师,博士,主要从事模式识别与图像处理的研究. E-mail: wanggh@aliyun.com.

基金项目: 湖北省教育厅科学技术研究青年项目(Q20161405);湖北省教育厅科学技术研究指导性项目(B2015045)

Krizhevsky 的网络结构进行优化,其采用的学习算法存在训练参数多、时间长、训练过程复杂等问题. Sermanet 等^[9]提出了 Overfeat 模型,但同样存在相同的问题. Jie 等^[10]提出 cross-level LLC-CNN 模型以提高场景分类的准确率,但存在最终全局特征向量对于几何变换不具有不变性的缺点. 此后,各种优化分类模型相继被提出^[11]. 吕娜^[12]提出采用 K-means 聚类算法提取特征的深层网络结构,虽然其在分类性能略优于单层网络结构^[13],但对实验平台要求较高,且采用的 DAG-SVM 分类器存在层次结构“误差累积”的弊端. 何鹏程^[14]提出基于 K-means 的卷积神经网络,其分类效果略优于单层 K-means 网络结构,但其训练过程繁琐且存在分类时间长、效率低等问题. 基于此,本文将 K-means 聚类算法与卷积结合,构建多层网络结构 KDL(K-means and deep learning)模型.

1 相关工作

1.1 卷积神经网络

卷积神经网络通过多个交叠的卷积层和子采样层从自然图像中提取图像特征,并采用全连接神经网络对提取的图像特征完成分类. 文中采用 Matlab 工具箱自带的一种卷积神经网络模型^[15],如图 1 所示. 在特征提取阶段,输入大小为 $n \times n$ 的图像,经过 6 个大小为 5×5 卷积核卷积,其运算过程为

$$a_j = f(\sum_{i \in I_j} x_i \cdot k_{i,j} + b_j). \tag{1}$$

式(1)中: i 为图像中对应一点取值; j 为特征图的编号; $k_{i,j}$ 为第 j 个特征对应的卷积核中的取值 i ; I_j 为输出特征图所对应的当前输入图像,每一个输出图对应一个偏置 b . 输入图像经过卷积后,添加对应的偏置 b ,通过激活函数 f 输出 6 个大小为 $(n-5+1) \times (n-5+1)$ 的特征图,构成卷积层 C1. 然后,将 C1 层 6 个特征图通过 2×2 的子采样生成 6 个大小为 $((n-4)/2) \times ((n-4)/2)$ 的特征图,构成采样层 S2. 通过对特征图中 2×2 区域中的像素值取均值 y 表示该区域像素值,即

$$y = \sum_x x_i / 4. \tag{2}$$

式(2)中: x 为子采样的区域; x_i 为该区域中的神经元; y 为子采样后生成的特征. 重复上述步骤,生成卷积层 C3 及采样层 S4. 之后,将采样层 S4 中的特征图经过光栅化处理,生成一个向量,光栅层与输出层间采用全连接,构成全连接网络,将该向量送入输出层,得出识别结果.

1.2 Dropout CNN 网络

Hinton 等^[16]提出一种能防止网络模型过拟合的方法 Dropout,有效提高了神经网络的泛化能力. 由于深度网络要求有足够多的训练样本,当训练样本较少时,会导致网络出现严重的过拟合现象. Dropout 是在网络训练过程中,以概率 p 随机选取部分神经元,使其激活输出值为 0,使每一批次所训练的网络结构都不同,避免网络中每两个神经元同时产生作用,防止某些特征只在固定组合下才生效. 在测试过程中不采用 Dropout,为了保证采用 Dropout 隐层的下一层的输入和训练时具有相同的数量,使用均值网络的方法对所有激活输出值乘以概率 $(1-p)$. 根据经验,Dropout 的比例 p 通常设为 0.5 时,网络模型取得的效果最好. Dropout 的结构,如图 2 所示. 图 2 中: v 为输入值; W 为权值; $a(\cdot)$ 为激活函数; $m(\cdot)$ 为随机取 0 的概率.

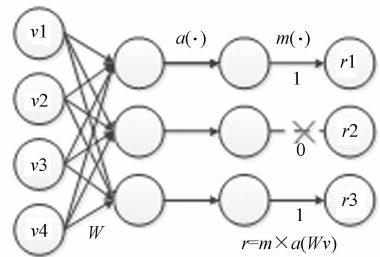


图 2 Dropout 结构图

Fig. 2 Dropout structure

2 KDL 模型

该模型由输入层、双隐层、输出层组成,隐层均采用 K-means 聚类算法对其进行训练,提取字典,再分别由特征映射函数和卷积操作提取图像特征,将提取出的图像特征送入 Softmax 分类器完成图像分类. KDL 模型的结构,如图 3 所示.

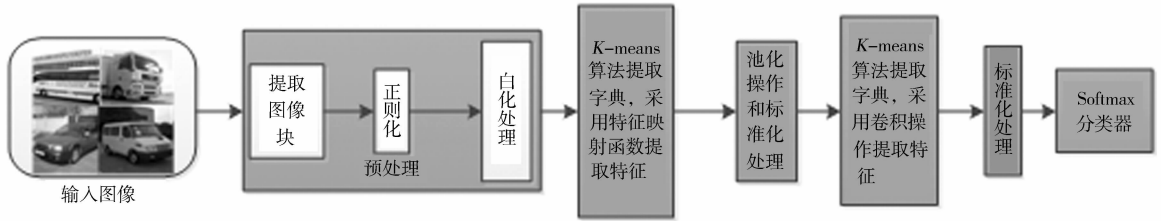


图 3 KDL 模型结构图
Fig. 3 KDL model structure

2.1 图像预处理

首先,采用大小为 $n \times n$ 无标签图像作为输入图像,并从中随机抽取一定量大小图像块构成 $w \times w$ 的无标签图像集 $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}, x^{(i)} \in \mathbb{R}^n$. 光照、遮挡等不利因素会影响从自然图像中得到的输入图像集的分类效果. 为了改善图像分类效果,提高分类精度,有必要对图像进行预处理. 采取正则化操作,增强图像对比度,以减少光线的影响. 正则化处理后的图像为

$$\hat{x}^{(i)} = (x^{(i)} - \text{mean}(x^{(i)})) / \sqrt{\text{var}(x^{(i)}) + \delta}. \tag{3}$$

式(3)中:var,mean 分别为方差和均值; δ 为常量,避免分母为 0 和对图像去噪. 由于图像数据间存在一定的相关性,为了消除图像间的冗余性,对图像采取主成分分析(PCA)白化处理.

2.2 图像特征提取

将预处理后的无标签图像集作为 K -means 聚类数据,通过 K -means 聚类算法训练网络,得到聚类中心,即字典. 首先, K -means 聚类算法设定初始聚类中心为 $\{\mu_1, \mu_2, \mu_3, \dots, \mu_k\}$. 由文献[13]可知,随着聚类中心(特征)数量的增加,该模型的性能也得到相应的提升,故设定聚类中心个数(字典维度)为 1 600. K -means 聚类算法通过迭代方式不断修改聚类中心,直到准则函数收敛. 设预处理后的无标签图像集为 $\{\hat{x}^{(1)}, \hat{x}^{(2)}, \dots, \hat{x}^{(m)}\}, \hat{x}^{(i)} \in \mathbb{R}^n$, 则准则函数为

$$j(c^{(1)}, \dots, c^{(m)}; \mu_1, \dots, \mu_k) = \frac{1}{m} \sum_{i=1}^m \|\hat{x}^{(i)} - \mu_{c^{(i)}}\|. \tag{4}$$

式(4)中: c 为类别; μ 为聚类中心.

对于任何一个图像块 i , 计算其属于的类别,即 $c^{(i)} = \arg \min \|\hat{x}^{(i)} - \mu_j\|$. 对于每一个类 j , 重新计算其聚类中心,即

$$\mu_j := \sum_{i=1}^m 1\{c^{(i)} = j\} \hat{x}^{(i)} / \sum_{i=1}^m 1\{c^{(i)} = j\}. \tag{5}$$

准则函数取最小值时,得到最优的聚类中心. 通过字典构建特征映射函数 $f_k(\hat{x}) = \max\{0, \mu(z) - z_k\}$. 式中: $z_k = \|\hat{x} - \mu_j\|$; $\mu(z)$ 为每一类所有样本到聚类中心距离和的平均值. 输入 $w \times w$ 的图像块 $\hat{x}^{(i)} \in \mathbb{R}^n$, 通过特征映射函数映射为一个新的特征表达 $y' = f_f(\hat{x}), y' \in \mathbb{R}^k$, 将 $n \times n$ 无标签图像转换成 $(n-w+1) \times (n-w+1) \times k$ 的图像特征表达,更准确地描述图像内容. 当特征映射函数输出值为 0 时,该特征到聚类中心的距离大于平均值. 这样保证输出大多数特征值为 0,使其具有稀疏性.

通过上述步骤提取输入无标签图像特征,由于得到的图像特征维数过高,故对其采取池化处理,降低特征向量的维数和整合图像特征. 待图像特征池化后,为了平衡特征的各个分量的影响,以改善后续 Softmax 分类器对图像特征的分类效果,该模型对已池化的图像特征进行标准化处理,即

$$\tilde{x}^{(i)} = (\hat{x}^{(i)} - \text{mean}(\hat{x}^{(i)})) / \sqrt{\text{var}(\hat{x}^{(i)}) + \delta}. \tag{6}$$

式(6)中:var,mean 分别为方差和均值; δ 为常量. 通过式(6)得到标准化处理后的图像特征为 $\tilde{x}^{(i)}$. 然后,将标准化的图像特征作为隐层 2 的输入,采用 K -means 聚类算法对其提取字典,并采取均值池化处理. 该过程与隐层 1 中采用 K -means 聚类算法提取字典的操作相同,通过 K -means 聚类算法提取聚类中心(即字典). 此时,设定初始聚类中心为 $\mu = \{\mu_1, \mu_2, \mu_3, \dots, \mu_k\}$, 由文献[12-13]可知,隐层 2 初始的聚类中心设定为 2 000 时,模型的识别率得到较好的提升,故设定聚类中心个数(字典的维度)为 2 000. 提取出字典后,采取离散卷积操作提取图像特征,有

$$y = \tilde{x} \cdot \mu'. \tag{7}$$

将输入的图像块与提取的字典做卷积,提取出最终图像特征,并采用式(6)对其标准化处理,以再次平衡特征的各分量的影响.对输入图像提取图像特征的过程,如图 4 所示.

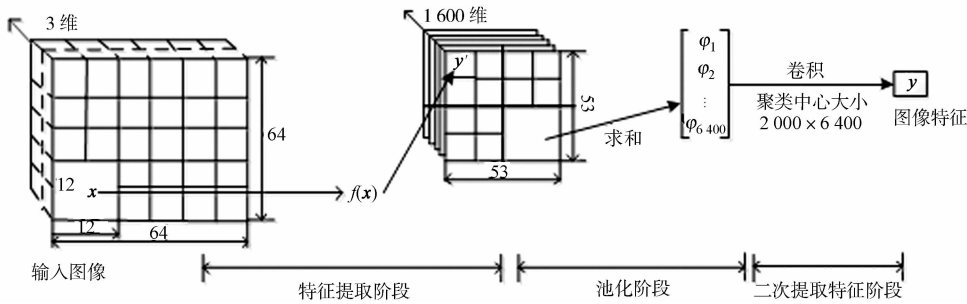


图 4 对输入 64×64 大小的彩色图片提取图像特征过程图

Fig. 4 Process diagram of 64×64 size color image extraction feature

2.3 图像特征分类

采用 Softmax 分类器完成图像分类,假设存在一个包含 m 个样本的图像数据集 $\{(\mathbf{x}^{(1)}, y^{(1)}), \dots, (\mathbf{x}^{(m)}, y^{(m)})\}$, $\mathbf{x}^{(i)} \in \mathbf{R}^n$, 对其进行多分类. 当类标签 $y^{(i)}$ 可以取 $k=10$ 个不同值时,采用 Softmax 分类器对其进行分类. 此时,假设函数表示为

$$h_{\theta}(\mathbf{x}^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | \mathbf{x}^{(i)}; \theta) \\ p(y^{(i)} = 2 | \mathbf{x}^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | \mathbf{x}^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k \exp(\theta_j^T \mathbf{x}^{(i)})} \begin{bmatrix} \exp(\theta_1^T \mathbf{x}^{(i)}) \\ \exp(\theta_2^T \mathbf{x}^{(i)}) \\ \vdots \\ \exp(\theta_k^T \mathbf{x}^{(i)}) \end{bmatrix}. \quad (8)$$

式(8)中: θ 为模型参数; $\exp(\theta_j^T \mathbf{x}^{(i)})$ 确保向量各元素间的和为 1. Softmax 分类器的代价函数为

$$J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{\exp(\theta_j^T \mathbf{x}^{(i)})}{\sum_{j=1}^k \exp(\theta_j^T \mathbf{x}^{(i)})} \right]. \quad (9)$$

式(9)中: $j=1, 2, 3, \dots, k$, k 为类别标签; $1\{y^{(i)} = j\}$ 为示范性函数,当 $y^{(i)} = j$ 成立时,其值取 1, 否则,取 0; θ 为模型参数, $\theta = [\theta_1^T \quad \theta_2^T \quad \dots \quad \theta_k^T]^T$.

由于目前还没有闭式解法求解 $J(\theta)$ 最小值,则通过求解其梯度公式,即

$$\nabla_{\theta} J(\theta) = \frac{1}{m} \sum_{i=1}^m [\mathbf{x}^{(i)} (1\{y^{(i)} = j\} - p(y^{(i)} = j | \mathbf{x}^{(i)}; \theta))]. \quad (10)$$

采用梯度下降法求解其最小值,并更新模型参数 θ 优化代价函数,有

$$\theta_j = \theta_j - \alpha \nabla_{\theta} J(\theta). \quad (11)$$

3 实验结果与分析

3.1 实验数据集

3.1.1 MNIST 数据集 MNIST 数据集包含 70 000 张样本图片,其中,训练样本图片 60 000 张,测试样本图片 10 000 张. 样本图片的内容包含阿拉伯数字 0~9, 每张样本图片为 28×28 的灰度图像.

3.1.2 Cifar-10 数据集 Cifar-10 数据集包含 60 000 张样本图片,其中,训练样本图片 50 000 张,测试样本图 10 000 张片. 样本图片内容包含飞机、卡车、汽车、船、马、猫、鹿、狗、鸟类、蛙类等 10 类物体. 其中,每一类图片 6 000 张,每张样本图片为 32×32 的彩色图像. 该数据集训练样本由 5 个部分组成,每部分包含 10 000 张图片;训练样本中每一类图片 5 000 张;测试样本每一类图片 1 000 张.

3.1.3 The vehicle 数据集 The vehicle 数据集包含 14 840 张样本图片,其中,训练样本图片 13 491 张,测试样本图片 1 349 张. 样本图片内容包含巴士图片 3 733 张,轿车图片 3 674 张,面包车图片 3 695 张,卡车图片 3 738 张. 每一张样本图片为 64×64 的彩色图像.

3.2 实验结果

通过上述 3 个数据集测试该模型的性能,将其分类结果与 CNN 网络及 Dropout CNN 的分类

结果进行比较,同时,与文献[12]中模型进行分析比较,实验结论,如表 1~3 所示.

由表 1~3 可知:KDL 模型对 MNIST 数据集的分类准确率(η)最高为 97.52%;CNN,Dropout CNN 对 MNIST 数据集的分类准确率最高可达 97.94%,94.31%.由此可见,虽然 CNN 网络、Dropout CNN 网络在 MNIST 取得不错的效果,但 KDL 模型在 MNIST 数据集上的分类效果与之相差不大.

由表 2,3 还可知:随着迭代次数(n)增加,CNN,Dropout CNN 在 Cifar-10 数据集上的分类准确率分别达到 44.23%,33.64%.由表 1 可知:随着迭代次数增加,KDL 模型对 Cifar-10 数据集的分类识别准确率逐渐提高到 61.72%.由此可见,KDL 模型在 Cifar-10 数据集上的分类效果优于 CNN 网络及 Dropout CNN 网络.此外,查找文献资料发现,CNN 网络在迭代 500 次左右时,准确率可达 48.46%;迭代 4 000 次时,准确率可达 70.49%;迭代 5 000 次时,准确率最高可达 75%,但其对数据集的适应度和分类效率不如 KDL 模型.此外,由表 1 还可知:KDL 模型对 The vehicle 数据集的分类识别准确率随着迭代次数增加,基本保持在 80.80%左右;而 CNN 网络、Dropout CNN 网络对 The vehicle 数据集的分类识别准确率分别达到 45.81%,52.64%.由此可见,KDL 模型在 The vehicle 数据集上的分类效果优于 CNN 网络及 Dropout CNN 网络.

表 1 KDL 模型在数据集上分类准确率

Tab. 1 Classification accuracy of KDL models on datasets

%

数据集	η				
	$n=100$	$n=200$	$n=300$	$n=400$	$n=1\ 000$
MNIST	97.40	97.32	97.52	97.30	97.22
Cifar-10	61.34	61.28	61.39	61.67	61.72
The vehicle	80.87	80.80	80.80	80.80	80.73

表 2 CNN 模型在数据集上分类准确率

Tab. 2 Classification accuracy of CNN models on datasets

%

模型	η				
	$n=1$	$n=5$	$n=10$	$n=15$	$n=20$
MNIST	80.80	95.94	97.14	97.80	97.94
Cifar-10	15.07	29.12	33.50	39.91	42.23
The vehicle	25.21	39.18	45.37	45.45	45.81

表 3 Dropout CNN 模型在数据集上分类准确率

Tab. 3 Classification accuracy of Dropout CNN models on datasets

%

模型	η				
	$n=1$	$n=5$	$n=10$	$n=15$	$n=20$
MNIST	46.29	91.09	92.35	93.98	94.31
Cifar-10	10.24	24.26	29.29	31.98	33.64
The vehicle	39.66	47.66	48.63	51.40	52.64

为了进一步更好地体现 KDL 模型的优势,将其与文献[12]中的模型进行分析对比.虽然文献[12]中的模型采用多层次 K -means 聚类算法提取图像特征,并通过大数据处理平台 Hadoop 并行计算取得了较好的分类效果,然而,该模型采用的 Hadoop 平台并不是适用于所有的数据库,且对实验设备要求较高;另一方面,该模型存在训练时间长,且采用的 DAG-SVM 分类器在多层次结构中存在“误差累积”的弊端,从而降低图像分类精确度.与文献[12]中的模型相比,KDL 模型不但具有分类效率高的优点,且对实验平台要求较低,同时对各类数据集都具有良好的分类性能.由此可见,KDL 模型的分类性能及适应度优于文献[12]中的模型.

综合上述分析,由于 CNN 网络及 Dropout CNN 网络采用有监督学习算法,训练过程复杂、训练参数较多导致图像分类效率偏低,而文中采用的无监督学习算法 K -means 算法在训练过程中只需训练字典,该模型训练过程简单、训练参数较少,具有简单、高效的特点,能更好地处理大规模无标签数据集.相比于文献[12]中的模型存在实验平台要求较高和误差累积的缺陷,KDL 模型具有分类性能高、适应度高和实验平台要求较低等优点.除此之外,KDL 模型也很好地将卷积独特的二维数据处理方式及便

捷性, 具有更好的适应度, 能更精准地提取图像特征.

4 结束语

将 K -means 算法与卷积构造 KDL 模型, 通过多次运用 K -means 聚类算法提取特征字典, 并分别通过构建的特征映射函数和卷积操作提取图像特征, 最终采用 Softmax 分类器对提取的图像特征分类. 相较于 CNN 网络、Dropout CNN 网络及文献[12]中的模型, KDL 模型具有分类精确度高、适应度高、训练过程简单、学习参数少等优点.

参考文献:

- [1] CIRESAN D, MEIER U, SCHMIDHUBER J. Multi-column deep neural networks for image classification[C]// IEEE Conference on Computer Vision and Pattern Recognition. Rhode Island: IEEE Press, 2012: 3642-3649.
- [2] SHI Baoguang, BAI Xiang, YAO Cong. Script identification in the wild via discriminative convolutional neural network[J]. Pattern Recognition, 2016, 52: 448-458. DOI: 10. 1016/j. patcog. 2015. 11. 005.
- [3] CHATFIELD K, SIMONYAN K, VEDALDI A, *et al.* Return of the devil in the details: Delving deep into convolutional nets[C]// Proc of the British Machine Vision Conference. Nottingham: BMVA Press, 2014: 1-12. DOI: 10. 5244/C. 28. 6.
- [4] SIMARD P Y, STEINKRAUS D, PLATT J C. Best practices for convolutional neural networks applied to visual document analysis[C]// Proc of the 17th International Conference on Document Analysis and Recognition. Edinburgh: IEEE Press, 2013: 958-963.
- [5] SUN Yi, WANG Xiaogang, TANG Xiaou. Deep convolutional network cascade for facial point detection[C]// Proc of the 26th IEEE Conference on Computer Vision and Pattern Recognition. Oregon: IEEE Press, 2013: 3476-3483.
- [6] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]// Proc of the 19th International Conference on Neural Information Processing. New York: ACM Press, 2012: 1097-1105. DOI: 10. 1145/3065386.
- [7] ZEILER M D, TAYLOR G W, FERGUS R. Adaptive deconvolutional networks for mid and high level feature learning[C]// Proc of the International Conference on Computer Vision, Barcelona: IEEE Press, 2011: 2018-2025.
- [8] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]// Proc of the European Conference on Computer Vision. Berlin: Springer, 2014: 818-833.
- [9] SERMANET P, EIGEN D, ZHANG Xiang, *et al.* Overfeat: Integrated recognition, localization and detection using convolutional networks[J]. Eprint Arxiv, 2014(V4): 1-16.
- [10] JIE Zequn, YAN Shuicheng. Robust scene classification with cross-level LLC coding on CNN features[M]. Bangalore: Springer, 2014.
- [11] BARAT C, DUCOTTET C. String representations and distances in deep convolutional neural networks for image classification[J]. Pattern Recognition, 2016, 54(C): 104-115. DOI: 10. 1016/j. patcog. 2016. 01. 007.
- [12] 吕娜. 基于深度层次特征学习的大规模图像分类研究[D]. 成都: 电子科技大学, 2015.
- [13] COATES A, NG A Y, LEE H. An analysis of single-layer networks in unsupervised feature learning[J]. Journal of Machine Learning Research, 2011, 15: 215-233.
- [14] 何鹏程. 改进的卷积神经网络模型及其应用研究[D]. 大连: 大连理工大学, 2015.
- [15] PALM R B. Prediction as a candidate for learning deep hierarchical models of data[D]. Denmark: Technical University of Denmark, 2012.
- [16] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, *et al.* Improving neural networks by preventing co-adaptation of feature detectors[J]. Computer Science, 2012, 3(4): 213-223.

(责任编辑: 钱筠 英文审校: 吴逢铁)