

融合图嵌入的光滑主成分分析网络图像识别算法*

陈飞玥¹, 朱玉莲², 田甲略¹, 蒋珂¹

(1. 南京航空航天大学 计算机科学与技术学院, 江苏 南京 211106;

2. 南京航空航天大学 公共实验教学部, 江苏 南京 211106)

摘要:主成分分析网络(principal component analysis network, PCANet)是一种简单的深度学习算法,在图像识别领域具有优秀的性能。将图嵌入思想融入 PCANet,提出一种新的图像识别算法光滑主成分分析网络(Smooth-PCANet)。为了验证 Smooth-PCANet 算法的有效性,在人脸、手写体字符以及图片等不同数据集上构建实验,并将 Smooth-PCANet 与多种基于深度学习的图像识别算法作了对比。实验结果证明,Smooth-PCANet 算法比 PCANet 获得了更高的识别性能,并且更有效地避免了过拟合,在小样本训练时具有显著优势。

关键词:图像识别;主成分分析网络;图嵌入;深度学习;小样本训练集

中图分类号:TP391 文献标志码:A 文章编号:1001-2486(2022)03-016-07

Smooth principal component analysis network image recognition algorithm with fusion graph embedding

CHEN Feiyue¹, ZHU Yulian², TIAN Jialue¹, JIANG Ke¹

(1. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China;

2. Fundamental Experiment Teaching Department, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract: PCANet (principal component analysis network) is a simple deep learning algorithm with excellent performance in the field of image recognition. Integrating the idea of graph embedding into PCANet, a new image recognition algorithm Smooth-PCANet was proposed. In order to verify the effectiveness of the Smooth-PCANet algorithm, adequate experiments were performed on different data sets such as face, handwritten characters, and images. Compared with several image recognition algorithms based on deep learning, the experiments demonstrated that the Smooth-PCANet achieves higher recognition performance than the PCANet and avoids overfitting more effectively, with a significant advantage in small samples training.

Keywords: image recognition; principal component analysis network; graph embedding; deep learning; small training set

图像识别是计算机视觉的重要分支,在安全验证^[1-2]、电子商业^[3]、医学影像^[4]、目标检测^[5-6]等领域均有广泛的应用。通常,图像识别分为四个步骤,即图像采集、图像预处理、特征提取和分类。图像采集是指通过成像设备获取图像的过程;图像预处理是将图像信息转换为计算机可以操作的数字信息,并进行简单的编辑;特征提取是从图像中提取出利于样本类别判别分析的信息;而分类则是对提取的特征进行判断,以确定该样本所属的类别。所以,特征提取是图像识别的关键步骤之一。特征提取方法主要包括两类:基于手工设计的特征提取方法和基于学习(数据驱动)的特征提取方法。其中,基于人工设计的特征提取方法主要有尺度不变特征变换^[7](scale-

invariant feature transform, SIFT)、方向梯度直方图^[8](histogram of oriented gradient, HOG)等;基于学习的特征提取包括基于子空间学习的特征提取(如主成分分析^[9](principal components analysis, PCA)、线性判别分析^[10](linear discriminant analysis, LDA)等)和基于深度学习的特征提取(如 LeNet^[11]、AlexNet^[12]、VGG^[13]等)。

以卷积神经网络^[11](convolutional neural networks, CNN)为代表的基于深度学习的特征提取方法得到迅速发展,并在图像识别领域得到了广泛的应用。CNN 通常由多层网络组成,而每一层网络又由卷积层、非线性处理层和池化层构成。其中,卷积层是 CNN 的核心,训练 CNN 的过程实

* 收稿日期:2021-06-20

基金项目:国家自然科学基金资助项目(61703206)

作者简介:陈飞玥(1995—),女,陕西榆林人,硕士研究生,E-mail:chenfeiyue@nuaa.edu.cn;

朱玉莲(通信作者),女,副教授,博士,硕士生导师,E-mail:lianyi_1999@nuaa.edu.cn

实际上就是学习各层卷积核参数的过程。为了学习卷积核, CNN 一般使用反向传播(back propagation, BP)算法进行训练。BP 算法是一个迭代算法,它包含数据的正向传播和梯度的反向传播。正向传播是根据已有参数从输入层开始依次计算每个隐含层直至输出层的过程;而反向传播则是利用梯度下降法从输出层到输入层依次更新参数的过程。两个过程交替进行,直至满足停止条件(如误差最小值)。由于 CNN 网络结构通常非常复杂,有大量的参数需要学习,所以一方面,需要有大量的训练样本进行训练以避免过拟合现象出现;另一方面,采用 BP 迭代求解参数的过程是非常耗时的。此外,复杂的网络结构通常需要一些额外的训练技巧才能获得良好的性能。

为了解决这些问题, Chan 等提出了一种简化的深度网络——主成分分析网络 PCANet^[14]。PCANet 由输入层、两层卷积层和输出层组成。由于 PCANet 的卷积核直接通过 PCA 方法获得,有效避免了 CNN 中调参的过程,因此不仅大大降低了计算代价而且也避免了过拟合问题。大量的实验证实 PCANet 在很多方面取得了接近于 CNN 的性能^[14-16]。鉴于 PCANet 的成功,人们基于 PCANet 提出了一些新的算法,如 Low 等通过引入池化层实现了 PCANet 中两个卷积层之间的非线性映射进而提出了 stacking PCANet +^[16]; Xi 等使用局部二值模式(local binary patterns, LBP)描述符替换了 CNN 卷积核,从而提出了局部二值模式网络(LBPNet)^[17]; Ng 等则利用二维离散余弦变换(two-dimensional discrete cosine transform, 2D-DCT)与 PCA 高度近似这一特点,使用 2D-DCT 基作为卷积核提出了二维离散余弦变换网络(DCTNet)^[15]。

然而, PCANet 算法也存在不足。由于在训练过程中将所有局部块^[14]视为彼此独立的,该算法忽略了局部块的中心像素点之间隐含的位置关系。设 p_i, p_j 是输入图像中相邻的两个像素点,以 p_i, p_j 为中心的局部块分别为 x_i, x_j 。由于图像中相邻像素 p_i, p_j 之间通常具有很高的相似性,因此局部块 x_i, x_j 之间也应具有较高的相似性,但是 PCANet 在利用 PCA 获得卷积核时却忽略了这一重要信息,致使最终所学习的特征未能很好地保持其空间特性。通常,一个较好的空间变换应该能有效地保持原空间样本点之间的近邻关系及相似性。亦即,在原空间位置数据值相近的两个近邻样本点投影至目标空间后,仍然应满足位置相邻且数据值接近的空间特性。基于此,提出了各

种图嵌入方法^[18-21]。图嵌入方法一般首先根据原空间样本建立一个图 $G = (Q, E)$, 其中 Q 是图节点的集合,每个图节点对应着每个样本; E 是图节点之间带有权值的边的集合,而边的权值表示两个图节点之间的邻接关系。根据邻接关系求得一空间变化矩阵,使得图中的节点映射至目标空间后,样本间的位置关系与原空间尽可能一致。在图嵌入方法中,图节点既可以是整幅图像也可以是图像中的一个像素点。传统的流形学习方法如局部保持投影^[18](locality preserving projections, LPP)、局部线性嵌入^[19](locally linear embedding, LLE)等均以整幅图像为图节点。其中, LPP 通过寻找样本点间拉普拉斯特征映射的最优线性逼近实现图嵌入的目标; LLE 采用保持近邻样本点之间线性关系的思想保持样本点间的位置关系。而空间光滑的子空间学习方法^[20](spatially smooth subspace learning, SSSL)则以一幅图像的像素点作为图节点。SSSL 方法通过对子空间的基向量做光滑约束,使基向量中相邻位置元素值的变化趋于平缓进而获取图像的位置关系。在 SSSL 的基础上,通过变换约束函数提出了基于高斯拉普拉斯算子(Laplacian of Gaussian, LOG)和高斯导数(derivative of Gaussian, DOG)惩罚的空间光滑子空间学习^[21]等方法。大量的实验结果证实,在特征提取过程中保持局部位置信息,有利于后续图像识别性能的提升。

为了充分利用图像的位置信息,本文将图嵌入的思想融入 PCANet 中,提出基于图嵌入的光滑 PCANet (Smooth-PCANet)。在求解 PCANet 卷积核时,除了保证各个局部块在新空间的特征(向量化操作后)具有最大的协方差,同时还要保证新特征之间的光滑性,即相邻的局部块在新空间的特征应该具有相似的特征值。因此,相比于 PCANet, Smooth-PCANet 利用了更多的位置信息。为验证 Smooth-PCANet 的有效性,分别在人脸数据集(以 AR 和 Extended Yale B 数据集为例)、手写体文字数据集(以 MNIST 数据集为例)和图片数据集(以 CIFAR 10 数据集为例)上进行了实验。实验结果表明, Smooth-PCANet 获得了比 PCANet 更好的性能;同时,和一些经典的神经网络算法相比,当训练样本数量相对较少时 Smooth-PCANet 表现更为出色。

1 融合图嵌入的 Smooth-PCANet

1.1 PCANet

PCANet 为简易的深度神经网络,由输入层、

两个卷积层和输出层组成。对各层进行简要介绍。

输入层:设图像训练集 $I = \{I_1, I_2, \dots, I_N\}$ 中包含 N 个大小为 $m \times n$ 的图像。对于 I^l 中每幅图像 $I_i \in \mathbf{R}^{m \times n}$ 做如下预处理:首先以每个像素点为中心取一个大小为 $k_1 \times k_2$ (即核大小) 的局部块,并将其向量化为一个列向量,然后将所有像素点对应的列向量放在一起形成一个矩阵 $X_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,\tilde{m}}] \in \mathbf{R}^{k_1 k_2 \times \tilde{m}}$,其中 $\tilde{m} = m - k_1 + 1, \tilde{n} = n - k_2 + 1$ 。最后将 X_i 的每个列向量减去 X_i 的列平均值,进而得到去均值化后的矩阵 $\bar{X}_i = [\bar{\mathbf{x}}_{i,1}, \bar{\mathbf{x}}_{i,2}, \dots, \bar{\mathbf{x}}_{i,\tilde{m}}]$ 。把所有的 $\bar{\mathbf{x}}_i (i = 1, 2, \dots, N)$ 放在一起,可以得到一个大矩阵 $X = [\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \dots, \bar{\mathbf{x}}_N]$ 。最后将 X 作为 PCANet 第一阶段卷积层的输入。

第一个卷积层:设第一卷积层卷积核的个数为 L_1 。首先使用奇异值分解 (singular value decomposition, SVD) 分解求解 XX^T 的前 L_1 个特征值所对应的特征向量 $\{V_1^1, V_2^1, \dots, V_{L_1}^1\}$,然后将每个特征向量转化为 $k_1 \times k_2$ 的矩阵,而每个矩阵就被视为 PCANet 的卷积核。因此 $W_l^1 = \text{mat}_{k_1 k_2}(V_l^1) \in \mathbf{R}^{k_1 \times k_2} (l = 1, 2, \dots, L_1)$ 为第一卷积层的卷积核,其中 mat 是将向量 $V \in \mathbf{R}^{k_1 k_2}$ 转换为矩阵 $W \in \mathbf{R}^{k_1 \times k_2}$ 的函数。获得卷积核后就可以对每个训练图像进行卷积操作,相应得到针对 I_i 的特征图像 $I_{i,l}^1 = I_i * W_l^1 (l = 1, 2, \dots, L_1)$ 。所有图

像卷积后的特征图像构成的集合 $\{I_{i,l}^1\} (i = 1, 2, \dots, N; l = 1, 2, \dots, L_1)$ 将作为下一个卷积层的输入图像。

第二个卷积层:设第二卷积层的卷积核数为 L_2 。首先将第一阶段卷积层的输出 $\{I_{i,l}^1\}$ 进行输入层的预处理,生成第二层的输入矩阵,然后进行几乎与第一个卷积层相同的步骤,得到 L_2 个卷积核,最后分别将每个卷积核与 $\{I_{i,l}^1\}$ 中的每个图像元素进行卷积操作并将卷积结果输送至输出层。

输出层:经过两层卷积层的操作后,每个输入图像 I_i 都能得到 $L_1 \times L_2$ 个特征图像,随后在输出层进行非线性处理、分块直方图和特征拼接等一系列操作后,最终得到图像 I_i 的特征向量表示。由于本文算法只涉及卷积核的构建,并不涉及输出层的改进,因此关于输出层的操作此处不做详细介绍,具体操作参见文献[14]。

1.2 Smooth-PCANet

在 PCANet 的基础上,通过引入像素点之间的近邻信息,本文提出了 Smooth-PCANet。Smooth-PCANet 和 PCANet 算法的流程对比如图 1 所示。从图 1 可以看出,相对于 PCANet,Smooth-PCANet 在计算卷积核时融入了拉普拉斯矩阵信息。下面,本文将从构建拉普拉斯矩阵及求解卷积核两方面对 Smooth-PCANet 进行详细介绍。

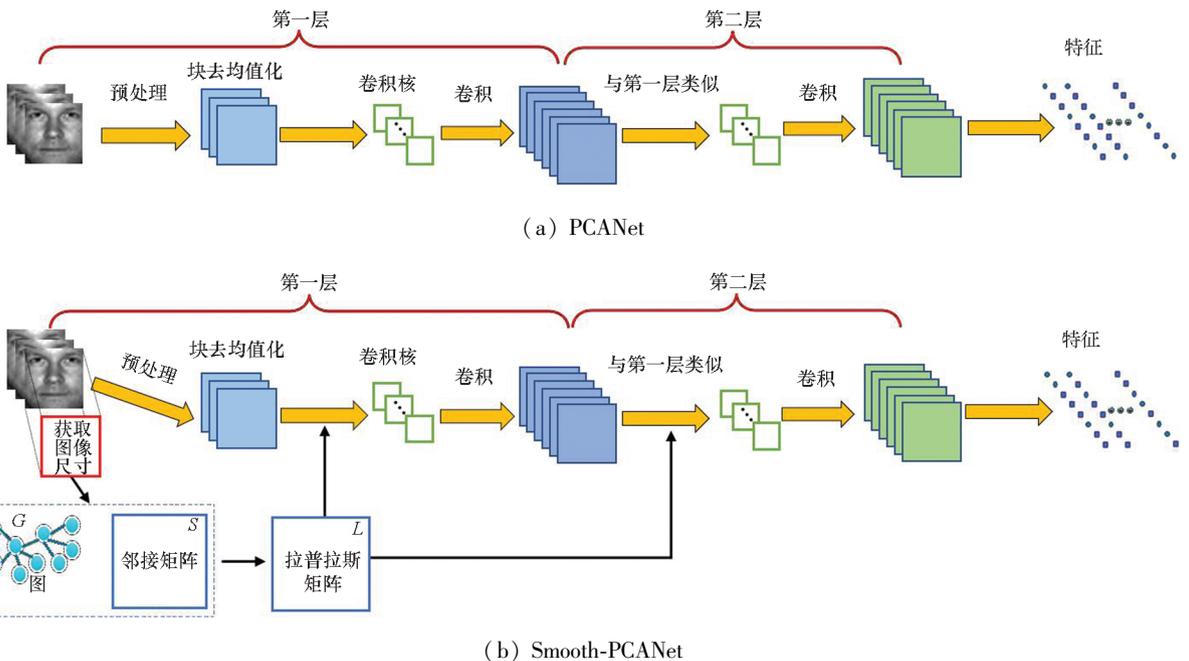


图 1 PCANet 和 Smooth-PCANet 算法的流程对比示意

Fig. 1 Flow charts comparison of PCANet and Smooth-PCANet

1.2.1 构建拉普拉斯矩阵

设输入图像的大小为 $m \times n$ 。首先构建一个全连通图 $G = (Q, E)$, 其中 Q 为图像所有像素点所构成的集合, E 为边集, 其元素边权值 $S_{ij} =$

$$\begin{cases} e^{-\frac{(x_{p_i} - x_{p_j})^2 + (y_{p_i} - y_{p_j})^2}{2}} & p_i \neq p_j, \text{其中 } x_{p_i}, y_{p_i} \text{ 和 } x_{p_j}, y_{p_j} \text{ 表} \\ 0 & \text{else} \end{cases}$$

示像素点 p_i, p_j 在图像上的纵横坐标。显然, 当两个像素距离较近时边权值较大; 而距离较远时边权值较小。通过邻接矩阵 S 很容易获得图拉普拉斯矩阵 $L = D - S$, 其中 D 为一个对角矩阵且对角线上的元素为 S 每行(或每列)的元素之和, 即

$$D_{ii} = \sum_j S_{ij}.$$

1.2.2 求卷积核

在 PCANet 中, 先以每个像素点为中心取大小为 $k_1 \times k_2$ (与卷积核大小相同) 的局部块, 之后以局部块为单位利用 PCA 算法求出网络的卷积核, 最后通过卷积核将每个局部块映射至目标空间。设 p_i, p_j 是图像中两个近邻像素点, 以 p_i, p_j 为中心获取的局部块分别为 x_i, x_j 。很显然 x_i, x_j 之间应具有较强相似性, 同时也希望在经过卷积核投影后, x_i, x_j 在目标空间获得的特征 y_i, y_j 也应具有较强的相似性且位置关系尽量保持。但在 PCANet 中, 这种关系并没有得到很好的利用。鉴于图嵌入的成功, 本文尝试在学习卷积核的同时引入光滑约束正则化项。设经过预处理后的第一层输入为 X , 通过卷积核 $V \in \mathbf{R}^{k_1 k_2 \times L_1}$ 获得的特征投影为 $Y = V^T X$, 为了保证任意局部块 x_i, x_j 在空间变换后的 y_i, y_j 仍然保持近邻关系, 应使 $\frac{1}{2} \sum_{i,j} \|y_i - y_j\|^2 S_{ij} = YLY^T = V^T X L X^T V$

尽量小。进一步结合 PCANet 中 $YY^T = V^T X X^T V$ 最大的目标, 本文提出了 Smooth-PCANet (d) 和 Smooth-PCANet (m), 对应的目标函数如式(1)~(2)所示。

$$\begin{aligned} V_d^* &= \arg \max_V \text{tr} \left(\frac{V^T X X^T V}{V^T X L X^T V} \right) \\ \text{s. t. } &V^T V = I \end{aligned} \quad (1)$$

$$\begin{aligned} V_m^* &= \arg \max_V \text{tr} (V^T X X^T V - \lambda V^T X L X^T V) \\ \text{s. t. } &V^T V = I \end{aligned} \quad (2)$$

其中, I 为 L_1 阶单位矩阵, $\lambda \in (0, 1]$ 为超参数。

因为式(1)~(2)是(广义)特征值问题, 因此很容易得到闭解。在求解卷积核后, Smooth-PCANet 采用和 PCANet 完全一样的操作进行特征提取。

2 实验结果与分析

2.1 人脸识别任务

在人脸识别任务中, 本文使用 AR 数据集和 Extended Yale B 数据集进行了实验。在实验中, PCANet 和 Smooth-PCANet 的基本参数设置如下: 卷积核的大小和数目分别设置为 $k_1 = k_2 = 5, L_1 = L_2 = 8$ 、直方图块重叠率为 0.5, 直方图块大小为 8×8 。

2.1.1 AR 数据集

AR 人脸数据库由 126 个人(70 个男性和 56 个女性), 每人至少 26 幅共计 4 000 多幅正面图像组成。本文中采用 AR 的一个子集, 包含 100 个人(50 个男性、50 个女性)每人 26 幅共 2 600 幅图像, 并将图像分辨率调整为 66×48 。某个人的 26 幅人脸图像如图 2 所示。在图 2 中, 时间 1 和时间 2 分别表示图片拍摄时间在第一周内和第二周内, 下标 1~13 代表图像的序号。实验将时间 1 和时间 2 的人脸图像划分为 6 个数据子集, 即 s_1 : Illum & Exps(时间 1 的 1~7)、 s_1 : Sunglass(时间 1 的 8~10)、 s_1 : Scarf(时间 1 的 11~13)、 s_2 : Illum & Exps(时间 2 的 1~7)、 s_2 : Sunglass(时间 2 的 8~10)、 s_2 : Scarf(时间 2 的 11~13), 使用 s_1 : Illum & Exps 学习网络并进行训练, 其余子集作为测试集评估各算法的性能。

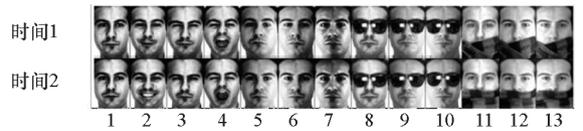


图 2 AR 数据库上某人的样本

Fig. 2 Samples of one person from AR database

实验结果如表 1 所示。由表 1 数据可以看出: ①从平均准确率上看, 只有 PCANet 及 Smooth-PCANet 算法能达到 90.00% 以上的准确率, 尤其是 Smooth-PCANet(m) 取得了最高结果 91.53%, 而传统神经网络的准确率均低于 50%。这是因为实验中仅以 700 幅无明显噪声的人脸图像作为训练集, 致使传统神经网络出现了不同程度的过拟合; 而 PCANet 和本文提出的 Smooth-PCANet 均是浅层网络, 有效避免了过拟合, 因此性能更为稳定。②从各个子集的实验结果来看, 含遮挡的人脸图像严重影响了传统神经网络的分类性能, 例如 LeNet 在无遮挡的 s_2 : Illum & Exps 上准确率约为 73.86%, 但在 s_2 : Scarf 上准确率仅有约 6.33%。值得注意的是, 虽然 PCANet 及 Smooth-PCANet 在遮挡人脸图

像上亦有性能下降,但是比起其他算法明显降幅最小,比如在 s_1 : Scarf 上,与 s_2 : Illum & Exps 的识别结果相比,PCANet 及 Smooth-PCANet 算法的性能降幅未超过 6.86%,远低于 LeNet 在该子集上的性能降幅 59.85%。③Smooth-PCANet(d) 和 Smooth-PCANet(m) 针对不同类型的人脸图像遮挡噪声分

别表现出了最佳性能。其中,Smooth-PCANet(d) 在墨镜遮挡人脸的性能具有明显优势,Smooth-PCANet(m) 则在围巾遮挡图像上表现更优。综上所述,Smooth-PCANet 不仅提升了 PCANet 在含遮挡人脸识别的准确性和鲁棒性,还在小样本训练集上发挥出比 PCANet 更大的优势。

表 1 AR 数据集上各算法的性能

Tab. 1 Performance of each algorithm on AR dataset

算法名称	s_1 :Sunglass	s_1 :Scarf	s_2 :Illum & Exps	s_2 :Sunglass	s_2 :Scarf	平均
CNN-2 layer	22.33	9.33	81.43	10.00	4.01	37.21
LeNet	17.67	14.01	73.86	12.67	6.33	35.21
AlexNet	6.34	4.33	57.01	5.33	3.33	24.06
VGG-19	4.01	5.33	71.57	4.67	3.67	29.16
PCANet	98.33	92.00	98.86	77.00	72.00	90.00
Smooth-PCANet(d)	99.67	94.33	97.86	81.33	73.67	91.16
Smooth-PCANet(m)	99.00	94.33	98.86	80.00	75.67	91.53

2.1.2 Extended Yale B 数据集

Extended Yale B 数据集包含 38 个个体的 2 414 幅人脸图像,每个个体拍摄约 64 幅不同光照下的照片^[22]。实验采用了该数据库的一个子集,包含 30 个个体的 1 920 幅人脸图像,分辨率为 96×84 ,并依照光照条件划分为 5 个子集(子集 1~子集 5)^[23],使用子集 1 进行学习训练,其余子集作为测试子集。某个个体的图像部分划分如图 3 所示。实验所得结果见图 4。从图 4 中可以看出:①从子集 2~子集 5,所有算法都出现了性能降低,说明光照是影响图像识别性能的重要因素。②PCANet 和 Smooth-PCANet 的性能在

所有子集上都超越了传统神经网络,且在子集 4 和子集 5 上该优势尤为明显,例如在子集 4 上 Smooth-PCANet(d) 和 Smooth-PCANet(m) 准确率分别为 68.57% 和 68.1%,而 AlexNet 仅有 9.05%。这是因为小样本训练集时神经网络再次出现过拟合问题。③在子集 5 上,Smooth-PCANet(m) 表现出更显著的光照鲁棒性,识别准确率比起第二名的 Smooth-PCANet(d) 亦至少高出 1.9%。因此,Smooth-PCANet(m) 更适用于光照条件极差的人脸识别场景。

2.2 手写体识别任务

本文使用 MNIST 数据集进行手写体识别。MNIST 是数字 0~9 手写体图像库,共包含 60 000 幅训练图像和 10 000 幅测试图像^[11,14,24],图像分辨率为 28×28 。实验设置 PCANet 与 Smooth-PCANet 的参数如下: $k_1 = k_2 = 7, L_1 = L_2 = 8$ 。在实验中发现不同网络上将 60 000 幅训练图像全部用于训练所得的结果几乎无差异,所以本文分别使用 2 000 和 20 000 幅训练样本做实验。实验结果如表 2 所示。从表 2 中可以看出:①训练样本为 20 000 幅时,所有的算法都达到了 93% 以上的准确率,VGG-19 和 LeNet 达到了最高准确率 98.72%; Smooth-PCANet(d) 略低于前二者,结果为 98.55%; PCANet 和 Smooth-PCANet(m) 达到 97.90% 的准确率,与第一名差别不超过 1%。当训练样本仅为 2 000 幅时,各算法出现了不同程度的性能降低,

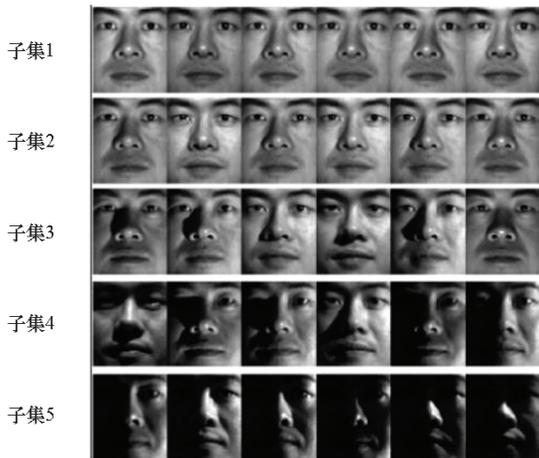


图 3 Extended Yale B 数据集的各子集示例

Fig. 3 Overview of each subset of Extended Yale B datasets

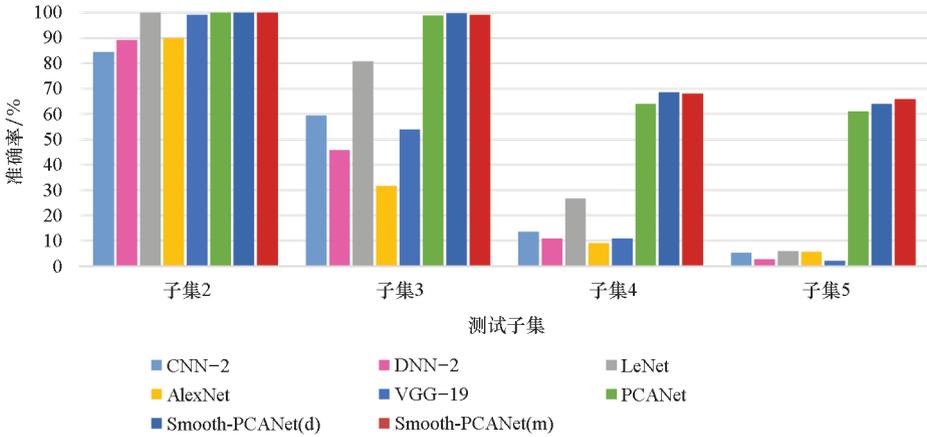


图 4 在 Extended Yale B 所有子集上各算法识别准确率柱状图

Fig. 4 Histogram of recognition accuracy of each algorithm on all subsets of Extended Yale B

表 2 各算法在不同训练样本规模的 MNIST 数据集上的性能

Tab. 2 Performance of each algorithm on MNIST datasets with different training sample sizes

算法名称	%	
	20 000 训练样本 时准确率	2 000 训练样本 时准确率
CNN	93.91	85.80
LeNet	98.72	94.64
AlexNet	98.62	97.10
VGG - 19	98.72	93.20
PCANet	97.90	93.70
Smooth-PCANet (d)	98.55	97.65
Smooth-PCANet (m)	97.90	94.10

如 AlexNet 性能下降了约 1.52%, VGG - 19 下降了 5.52%, PCANet 亦出现 4.2% 的降幅, 仅有 Smooth-PCANet (d) 的性能具有最小降幅 0.9%, 并且获得最高准确率 97.65%。②从整体上来说, 在手写体字符识别任务中 Smooth-PCANet 比 PCANet 获得了更好的性能; 同其他算法相比, Smooth-PCANet (d) 在小样本训练时具有更出色的表现。

2.3 图片识别任务

本文在 CIFAR 10 数据集进行了图片识别实验。CIFAR 10 数据集由 10 个类共 60 000 幅彩色 RGB 图像组成, 图像分辨率为 32×32 。在实验中, 将这彩色图像转为灰度图, 以 50 000 幅图像为训练集, 10 000 幅图像为测试集进行实验。实验中 PCANet 以及 Smooth-PCANet (d)、Smooth-PCANet (m) 的参数统一设置为 $k_1 = k_2 = 5, L_1 =$

$20, L_2 = 8$ 。所得结果如图 5 所示。从图 5 中可以看出, 在 CIFAR 10 数据集上所有的算法都无法达到较高的准确率, 这是因为经过灰度处理后图像的特征变得更难分辨与提取。故在图片集上提出的 Smooth-PCANet 与一般的神经网络相比并未展现出结果优势。值得关注的是, PCANet 和本文算法的实验结果与 LeNet 十分接近, 但是 LeNet 是较深层的神经网络, 除去输入输出层还包含了 3 个卷积层、2 个下采样层和 1 个全连接层。其他的神经网络层数更高。所以实验中传统的神经网络算法花费了远多于 PCANet 及 Smooth-PCA 的时间和硬件资源。如文献 [25] 所指出的, 在一定范围内深度学习网络的隐含层越多, 其分类性能越好。基于 PCANet 和 Smooth-PCANet 在实验中以最浅的网络层数达到图中所示性能水平的这一事实, 本文认为, 若对 Smooth-PCANet 进行合理的层数扩充, 其有望在图片识别任务上表现出更好的性能, 且在资源使用上仍具有较大优势。

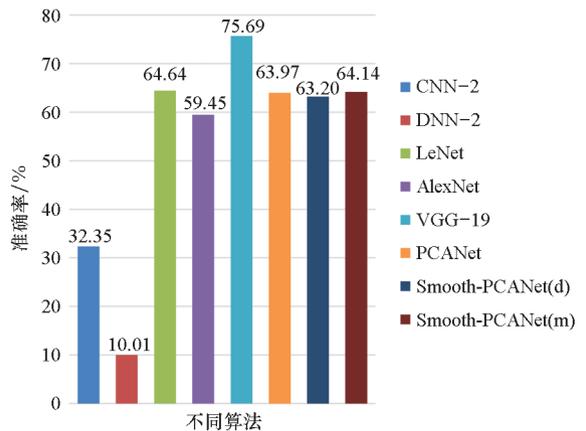


图 5 在 CIFAR 10 上各算法的识别准确率

Fig. 5 Recognition accuracy of each algorithm on CIFAR 10

3 结论

目前大多数深度学习存在着网络复杂训练难度大、硬件条件要求高、训练样本少时容易陷入过拟合等问题。本文在一种简易的深度学习网络 PCANet 上,通过融合图嵌入的光滑思想,提出了 Smooth-PCANet 图像识别算法,并分别在不同数据集上与其他方法进行对比实验。实验结果表明,Smooth-PCANet 实现了对 PCANet 的有效改进,对光照、遮挡等人脸图像噪声的鲁棒性更强,对手写字符的识别性能也更高;同 PCANet 一样,Smooth-PCANet 在小样本训练集上可以有效避免过拟合,与传统神经网络相比更有优势。

参考文献 (References)

- [1] SHI Y C, YU X, SOHN K, et al. Towards universal representation learning for deep face recognition [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6817–6826.
- [2] CHENG X L, XIONG H, FAN D P, et al. Implicit motion handling for video camouflaged object detection [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
- [3] JHA B K, SIVASABKARI G G, VENUGOPAL K R. E-commerce product image classification using transfer learning [C]// Proceedings of the 5th International Conference on Computing Methodologies and Communication, 2021: 904–912.
- [4] BONGRATZ F, RICKMANN A M, POLSTERL S, et al. Vox2Cortex: fast explicit reconstruction of cortical surfaces from 3D MRI scans with geometric deep neural networks [C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022.
- [5] KIM E, KIM S, SEO M, et al. XProtoNet: diagnosis in chest radiography with global and local explanations [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 15719–15728.
- [6] ZHANG S F, CHI C, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]// Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 9759–9768.
- [7] LOWE D G. Object recognition from local scale-invariant features [C]// Proceedings of the 7th IEEE International Conference on Computer Vision, 1999.
- [8] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005.
- [9] MARTINEZ A M, KAK A C. PCA versus LDA [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(2): 228–233.
- [10] BELHUMEUR P N, HESPANHA J P, KRIEGMAN D J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(7): 711–720.
- [11] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [C]// Proceedings of the IEEE, 1998, 86(11): 2278–2324.
- [12] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60(6): 84–90.
- [13] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]// Proceedings of International Conference on Learning Representations, 2015.
- [14] 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: a Publication of the IEEE Signal Processing Society, 2015, 24(12): 5017–5032.
- [15] NG C J, TEOH A B J. DCTNet: a simple learning-free approach for face recognition [C]// Proceedings of Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015: 761–768.
- [16] LOW C Y, TEOH A B J, TOH K A. Stacking PCANet+: an overly simplified ConvNets baseline for face recognition [J]. IEEE Signal Processing Letters, 2017, 24(11): 1581–1585.
- [17] XI M, CHEN L, POLAJNAR D, et al. Local binary pattern network: a deep learning approach for face recognition [C]// Proceedings of IEEE International Conference on Image Processing, 2016: 3224–3228.
- [18] HE X F, NIYOGI P. Locality preserving projections [C]// Proceedings of the 7th Annual Conference on Neural Information Processing Systems, 2003.
- [19] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500): 2323–2326.
- [20] CAI D, HE X F, HU Y X, et al. Learning a spatially smooth subspace for face recognition [C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2007.
- [21] ZUO W M, LIU L, WANG K Q, et al. Spatially smooth subspace face recognition using LOG and DOG penalties [C]// Proceedings of the 6th International Symposium on Neural Networks, 2009: 439–448.
- [22] ZHANG Q, LI B X. Discriminative K-SVD for dictionary learning in face recognition [C]// Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010: 2691–2698.
- [23] LEE K C, HO J, KRIEGMAN D J. Acquiring linear subspaces for face recognition under variable lighting [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(5): 684–698.
- [24] LAROCHELLE H, ERHAN D, COURVILLE A, et al. An empirical evaluation of deep architectures on problems with many factors of variation [C]// Proceedings of the 24th International Conference on Machine Learning, 2007: 473–480.
- [25] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504–507.