

DOI:10.16356/j.1005-2615.2021.05.008

基于无监督学习视觉特征的深度聚类方法

陈俊芬¹, 赵佳成¹, 翟俊海¹, 李艳²

(1. 河北大学数学与信息科学学院河北省机器学习与计算智能重点实验室, 保定 071002;

2. 北京师范大学珠海分校应用数学学院, 珠海 519087)

摘要: 基于自编码器的特征提取技术广泛应用于图像聚类分析, 在较简单的图像集上取得了令人满意的聚类结果, 但自编码器的特征表示能力有限, 很难捕捉到复杂低质图像的局部特征。本文提出一种基于非对称结构卷积自编码器 (Convolutional auto-encoder with an asymmetric structure, ASCAE) 的学习视觉特征的深度聚类方法, 其中非对称结构的卷积自编码器用于学习特征表示, 然后使用 K-means 算法对特征数据进行聚类分析。为进一步提高特征表示能力, ASCAE 方法的网络采用变步长的卷积层和全连接的重构误差正则约束网络的重构误差。在 7 个公开图像集上的实验结果表明该网络有很好的特征表示能力, 并且使得 K-means 算法能提供很好的聚类结果。在 COIL-20 和 MNIST 图像集上, 聚类方法 ASCAE 的聚类精度分别为 0.754 和 0.918, 优于同类型的 4 种深度聚类方法 (AEC、IEC、DEC 和 DEN)。

关键词: 图像聚类; 卷积自编码器; 特征表示; K-means 算法; 聚类精度

中图分类号: TP391

文献标志码: A

文章编号: 1005-2615(2021)05-0718-08

Deep Clustering Method Based on Unsupervised Visual Features Learning

CHEN Junfen¹, ZHAO Jiacheng¹, ZHAI Junhai¹, LI Yan²

(1. Hebei Key Laboratory of Machine Learning and Computational Intelligence,

College of Mathematics and Information Sciences, Hebei University, Baoding 071002, China;

2. School of Applied Mathematics, Beijing Normal University (Zhuhai), Zhuhai 519087, China)

Abstract: Despite recent progress in features extraction using deep conventional auto-encoders, which have greatly benefited image clustering analysis with the satisfying clustering results on several simple image-datasets. However, the representation ability of the conventional auto-encoders is quite limit when they capture local features of complex and low-quality images. Therefore, a novel deep clustering method combining with visual features learning is proposed, i.e., the convolutional auto-encoder with an asymmetric structure (ASCAE), in which an asymmetric convolutional auto-encoder is used to learn feature representation, and then K-means algorithm performs clustering analysis for these learned features. To further improve the representation suitability for downstream clustering tasks, ASCAE method adopts strided convolution layers, and minimizes the reconstruction error of whole network regularized by L_2 error between the left and right fully connected layers. Experimental results on seven public image-datasets illustrate that the network of ASCAE method usually offers better feature representation and brings promising clustering performance presented by K-means algorithm. Clustering accuracy of ASCAE method is 0.754 and 0.918 on databases COIL-20 and MNIST, respectively, which is better than four depth clustering methods of the same

基金项目: 河北省引进留学人员基金(C20200302)资助项目; 河北省科技重点研发计划(19210310D)资助项目; 广东省自然科学基金(2018A0303130026)资助项目。

收稿日期: 2020-02-15; **修订日期:** 2021-01-17

通信作者: 陈俊芬, 女, 副教授, E-mail: chenjunfen2010@126.com。

引用格式: 陈俊芬, 赵佳成, 翟俊海, 等. 基于无监督学习视觉特征的深度聚类方法[J]. 南京航空航天大学学报, 2021, 53(5): 718-725. CHEN Junfen, ZHAO Jiacheng, ZHAI Junhai, et al. Deep clustering method based on unsupervised visual features learning[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2021, 53(5): 718-725.

type (AEC, IEC, DEC and DEN).

Key words: image clustering; convolutional auto-encoder; feature representation; K-means algorithm; clustering accuracy

聚类是广泛应用于机器学习和数据挖掘领域的无监督学习方法。聚类结果依赖于聚类算法,也依赖于数据的特征表示,尤其是高维复杂数据(图像、语音和视频等)的特征表示显著影响着聚类性能。研究表明,在自动学习图像数据特征时,深层神经网络具有更强大的表达特征的能力^[1]。2006年,Hinton提出了深度(堆叠)自编码器(Autoencoder, AE),他们在对MNIST数据集的分类研究中发现,网络权值的调整主要依赖图像本身的信息量,而标签信息量只能微调网络权值^[2]。另外,对数据进行标注需要消耗大量的资源,例如2012年Hinton团队在ImageNet图像分类比赛中以极大的优势获得了胜利^[3],但是为ImageNet标注花费了2年多的时间^[4]。因此采用无监督学习方法对无标签图像自动提取特征具有重要的理论与实际意义。提取特征常用的深度方法有受限玻尔兹曼机^[5]、自编码器^[6]、卷积自编码器^[7]和生成式对抗网络^[8]等。自编码器是一种非线性特征提取方法,它的对称结构可以进行无监督学习,最终得到高维输入的低维特征表示。

很多研究结合自编码器对图像数据进行聚类,基于自编码器的图像聚类分析通常包括两阶段:(1)自编码器对图片进行特征提取;(2)聚类算法对特征进行聚类。由文献^[9-10]所提的深度聚类方法的实验发现,最终的聚类结果与自编码器的参数学习结果有很大关系,因此提取适合聚类的特征表示是研究的重点。与主成分分析(Principal component analysis, PCA)的线性特征相比,AE的非线性特征更具有辨识力,大大提高了聚类性能。尽管在人脸图像数据上取得了较满意的聚类结果,但由于AE网络的全连接导致网络参数的计算代价很高,从而限制了AE网络的深度。实验在较小的数据集上进行,例如在MNIST数据集上随机选取5000个图片进行实验。经典自编码器很难得到图像的局部细节特征,其特征表示能力有限。为了克服这些不足,进一步提高图像数据的聚类效果,本文提出一种基于新型卷积编码器(Convolutional auto-encoder with an asymmetric structure, ASCAE)的深度聚类方法,采用非对称的卷积编码器学习输入图片的特征表示,使用经典K-means算法对特征进行聚类划分。该算法的主要内容包括:

(1)设计了一个非对称卷积编码器。采用3×

3小卷积核提取局部特征,且编码器部分多一个全连接层。

(2)采用变步长卷积层替代池化层。第1个卷积层的卷积步长等于卷积核宽度,随后卷积层的卷积步长逐步减少。

(3)采用端到端的策略训练网络,其中随机初始化网络权值替代贪婪的层层预训练。

(4)通过最小化网络的重构误差和全连接的重构误差学习网络权值。

1 相关工作

无监督学习出现在各种图像处理任务中,例如:图像聚类^[9-14]、图像分类^[15-17]、图像分割^[18-19]及目标检测^[20]等。近年来,无监督图像聚类研究取得了很好的成绩。文献^[11]中采用生成对抗网络为无监督的图像特征提取带来新方向。其中,使用大步幅的卷积核替代池化层,删除全连接层和使用批处理;生成网络使用ReLU作为激活函数,而输出层使用Tanh函数;判别网络使用LeakyReLU函数;最后,使用判别网络的各层特征进行聚类分析,但其精度却不及传统的卷积神经网络。

相较于深度卷积生成对抗网络而言,自编码器和卷积神经网络在无监督聚类学习中的应用更为广泛。文献^[12]找到一种解决特征空间学习与聚类判别的联合方案。基本思路是:第一阶段训练自编码器进行特征空间的初步学习,然后使用K-means得到特征空间的K个聚类中心点;第二阶段计算所学特征与聚类中心点的相似性分布Q,并构造辅助目标分布P;最后通过最小化二者的KL(P||Q)散度调整网络参数和改进聚类结果。实验结果超越了以往的无监督聚类算法。用于聚类分析的无监督深度嵌入方法(Deep embedding method for clustering, DEC)框架如图1所示^[12]。

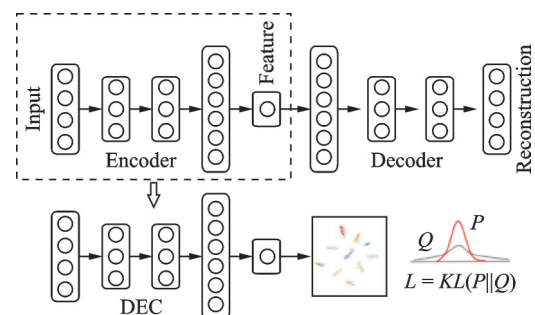


图1 用于聚类分析的无监督深度嵌入方法^[12]

Fig.1 Unsupervised DEC analysis^[12]

联合特征空间学习与聚类进行统一学习,为无监督聚类提供了新方向。例如,文献[9]提出了联合卷积自动编码器嵌入和相对熵进行深度聚类的方法。文献[10]提出了全卷积自动编码器的判别增强图像聚类方法。这两类方法分别做了不同方向的改进,获得了比原方法更优的结果。

文献[14]中用于视觉特征的无监督聚类方法的核心思路是:对卷积神经网络所学特征进行聚类生成伪标签,网络使用伪标签数据进行监督学习并更新网络参数,二者交替执行直至收敛。训练过程通过限制新簇的样本数、重新选择簇类中心以及根据伪标签进行均匀采样等措施避免聚类的平凡解。

文献[21]在杂乱数据上使用无监督图像特征预训练方法,对聚类与深度相结合的思路进行了优化。在初步聚为 m 类的基础上先进行一次 $4m$ 类别分类,再进行多次 k 类别分类,这些分类器使用相同的特征提取网络。循环往复训练,所训练出的特征质量要优于其他非监督学习方法。不同于以往的深度聚类研究思路,文献[22]中提出了另一类无监督图像分类和分割的不变信息聚类方法:根据图像经过平移、旋转等变换所学特征的聚类软分配结果应保持不变的思路,对卷积神经网络进行权重更新。卷积神经网络上的小改动使得聚类结果达到最优。该方法为无监督聚类学习学到更有表达能力的特征提供了新思路。

文献[9-10,12]中的实验表明,深度聚类的最终结果很大程度上依赖于卷积自动编码器的初始生成特征的 K -means 聚类结果。也就是说,联合训练之前网络生成的特征决定了最终聚类结果的好坏。本文对深度卷积自动编码器网络结构进行了改进,对卷积方式进行了适当调整,并在原 $L2$ 损失函数中添加关于全连接层的正则项。

2 深度聚类方法 ASCAE

本文所提深度聚类框架分开执行特征提取和聚类分析,旨在通过改进卷积自动编码器的结构和目标损失函数提升对图像的特征提取能力,从而最大化聚类分析的收益。

2.1 非对称特征提取器

为了得到辨识力更强的特征表示,本文提出了一种新颖的网络结构,如图2所示。网络细节描述如下:编码器部分由4个卷积层(CNN- $C_1 \sim C_4$)和3个全连接层(Full- $D_1 \sim D_2$)组成;解码器部分包括2个全连接层(Full- $D_2 \sim D_3$)和4个反卷积层(DCNN- $T_4 \sim T_1$)。由此,所设计的网络又称为“非对称结构的卷积编码器”,即 ASCAE 网络。该网络没有池

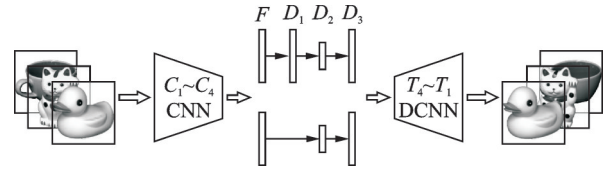


图2 非对称(上)和对称(下)的卷积自动编码器网络框架图
Fig.2 Framework of asymmetric(upper) / symmetric(lower) convolutional auto-encoder networks

化层, C_1 和 T_1 有 25 个卷积核,其余层均为 50 个。

受 VGG16 网络^[23]的启发,除 C_4 层外所有卷积层均采用 3×3 的卷积核。 C_1 层的卷积步长 $S=3$ 代表着滑动窗之间没有重叠,相当于把输入图片分割成多个小片后分别提取特征。这种运算加快了网络的学习速度,在分辨率高的大数据集上有更明显的优势。为了捕捉更丰富的特征信息, C_2 层上 $S=2$ 而 C_3 和 C_4 层上 $S=1$ 。

考虑到灰度值的极大差异性有助于卷积核获取关键特征,而归一化会缩小这种差异从而影响最终的特征表示,所以本文不对灰度图片进行归一化。另外,ASCAE 网络采用 ReLU 激活函数。

2.2 ASCAE 方法

设图像数据集为 $X = \{x_1, x_2, \dots, x_N\}$, 图片 x_i 通过 ASCAE 网络重构为 $\hat{x}_i = f_\theta(x_i)$, F 层和 D_3 层的输出分别记为 b_i 和 $g_i (0 \leq i < N)$ 。通过最小化网络输入和输出的差异性来训练网络参数,因此损失函数为

$$L(\theta; X) = \frac{1}{N} \sum_{i=1}^N \|x_i - f_\theta(x_i)\|^2 \quad (1)$$

式中 $\theta = \{\theta_e, \theta_d\}$ 为网络参数。

在最小化网络的重构误差过程中,通过减少 F 层与 D_3 层的差异性来降低 ASCAE 网络中全连接部分的误差。目标函数变为

$$J(\theta; X; \lambda) = \frac{1}{N} \sum_{i=1}^N \|x_i - f_\theta(x_i)\|^2 + \frac{\lambda}{N} \sum_{i=1}^N \|b_i - g_i\|^2 \quad (2)$$

式中: λ 为平衡参数,由于本文的图像像素值范围为 $[0, 255]$ 且未归一化,后文实验中 λ 的取值范围为 $[0, 40]$ 。

非对称卷积自动编码器提取特征算法的伪代码描述如下:

算法1 ASCAE 学习特征表示算法

输入:数据集 $X = \{x_1, x_2, \dots, x_N\}$

输出: F 层的特征 $B = \{b_1, b_2, \dots, b_N\}$

初始化:用服从高斯分布 $G(0, 0.1^2)$ 的随机数初始化网络权值;将 D_2 层的神经元个数设置为 k 个

迭代过程:最小化目标损失函数 $J(\theta; X; \lambda)$ 来训练非对称卷积自动编码器

$$x_i \xrightarrow{\theta_e} b_i(\text{features}) \xrightarrow{\theta_d} \hat{x}_i$$

通过反向传播算法迭代更新网络参数值
 停机条件:目标损失不再下降或达到最大迭代次数

输出特征: $f_{\theta_e}(X) \rightarrow B$

2.3 对称特征提取器

为验证非对称卷积自编码网络的特征表示能力,本文设计了一个对称卷积自编码网络(Convolutional auto-encoder with a symmetric structure, SCAE),如图 2 所示。SCAE 网络不同于 ASCAE 网络之处在于前者少了 D_1 层。二者的损失函数相同,都是通过最小化 $J(\theta; X; \lambda)$ 调整网络权值直至收敛。

2.4 特征表示的聚类分析

由图 2 可知,输入层到 F 层组成一个卷积神经网络来获取输入图像的局部特征。低层的特征图通常表示简单的特征,而高层则表示复杂的特征。随着网络层数的增加,特征图上像素点的感受野变大,所以 C_4 层上 1×1 特征图的感受野是整个图片,如图 3 所示。 C_4 的一个特征图对应 F 层的一个神经元,且连接权值表示输入的某个局部特征^[24]。

虽然全连接 D_2 层能对 F 层的局部特征进行组合,减少特征的位置对分类任务的影响,却不适用于对位置敏感的图像分割、聚类模式任务。因此本文使用 F 层的特征而未使用 D_2 层的特征完成聚类任务。下文的实验将验证 F 层特征优于 D_2 层。另外,2 个全连接层能更好地解决非线性问题,实验部分也将验证 ASCAE 网络的 F 层特征最优。

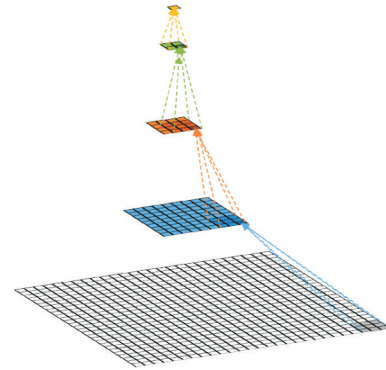


图 3 不同层特征图的感受野对应关系图

Fig.3 Corresponding relation between receptive fields of feature maps in different layers

3 实验与分析

本节通过 K-means 算法对特征进行聚类分析以验证 ASCAE 网络提取特征的能力。所有实验均在数据量较小的 7 个图像数据集(表 1)上进行。为了减少随机初始化对 K-means 算法的影响,每组实验都重复 50 次,对最好的聚类性能进行分析,并与文献[10]提供的聚类性能进行对比。评价聚类性能的数值指标有聚类精度(Accuracy, ACC)和标准互信息(Normalized mutual information, NMI)。ACC 表示聚类结果的正确率。NMI 衡量了样本标签的预测分布与真实分布的相关程度。两个指标值越接近 1,说明聚类准确度越高。

实验环境: IntelCorei5 - 6300HQ 处理器, NVIDIA 2.0 GB 显存, 8.0 GB 内存;基于开源的 Keras 库搭建 ASCAE 网络。

表 1 7 个图像数据集的信息描述

Table 1 Information description of seven image datasets

Dataset	Sample	Dimension	Label	Note
COIL-20 ^①	1 440	128 × 128	20	日常生活物品,有角度无形变,灰度图片
MNIST ^②	70 000	28 × 28	10	手写数字,有形变,灰度图片
USPS ^③	9 298	16 × 16	10	手写数字,有形变,灰度图片
CAS-PEAL-R1 ^④	200	480 × 360	40	正面视图,有表情变化,纯色背景,每人 5 幅灰度图片
IMM-Face ^⑤	240	640 × 480	40	视图有侧面和正面,有表情变化,纯色背景,每人 6 幅彩色/灰度图片
BioID-Face ^⑥	156	384 × 286	26	正面视图,较大姿态变化,有表情变化,室内背景,灰度图片
UMISTS ^⑦	564	220 × 220	20	视图有侧面和正面,较大姿势,纯色背景灰度图片

① <http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>.

② <http://yann.lecun.com/exdb/mnist/>.

③ <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/datasets/multiclass.html>.

④ <http://www.jdl.ac.cn/peal/index.html>.

⑤ <http://www.imm.dtu.dk/~aam/aamexplorer/>.

⑥ <https://www.bioid.com/About/BioID-Face-Database>.

⑦ <https://blog.csdn.net/garfielder007/article/details/51480770>.

3.1 COIL-20数据集

第1组对比实验验证最优参数对SCAE和ASCAE网络性能的影响。若设置 D_2 层的激活函数为Softmax,通过最小化损失函数 $L(\theta; X)$ 得到特征表示,然后使用K-means算法对 F 层特征进行聚类,仅得到0.356的聚类精度,且损失函数不易收敛。此 F 层特征的二维展示如图4所示。可以看出 F 层特征没有类簇的形状分布,依旧混乱交织在一起而无法辨识。



图4 数据集COIL-20上SCAE网络 F 层特征的二维展示
Fig.4 2-D visualization of F -layer features of SCAE network on dataset COIL-20

为了优化该问题,设置 D_2 层激活函数为ReLU,最小化目标函数 $J(\theta; X; \lambda)$ 来训练SCAE网络,使用K-means算法对 F 层特征进行聚类,聚类精度达到0.749。此时算法的收敛速度快,损失函数的收敛值更低,图片的重建度也更好, F 层特征的聚类精度更高。

针对同样的参数设置,通过最小化目标函数 $J(\theta; X; \lambda)$ 训练ASCAE网络,此时ASCAE方法的聚类精度提高到0.754。另外,图5展示了参数 λ 的不同取值对聚类性能的影响,其具有先增加后减小的趋势,在 $\lambda=20$ 时ACC和NMI达到了最优。后续实验中均设置 $\lambda=20$ 。图6给出了 F 层特征的二维可视化图。至此,ASCAE网络构建完成。从图6可以看到,ASCAE网络学习到的特征有了类簇的形状分布,使得K-means算法提供了较高的聚类精度,但仍有少量类别的特征辨识度较低,因

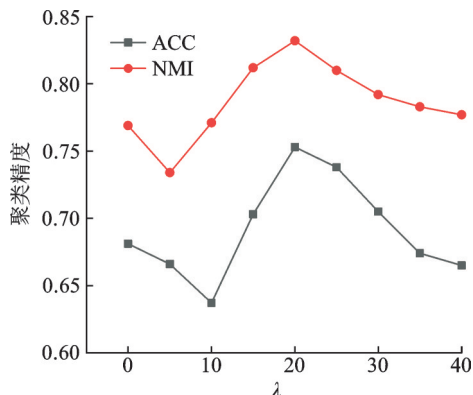


图5 数据集COIL-20上不同 λ 值的ASCAE聚类性能
Fig.5 Clustering performance of ASCAE method with different λ on dataset COIL-20



图6 数据集COIL-20上ASCAE网络的 F 层特征的二维展示
Fig.6 2-D visualization of F -layer features of ASCAE network on dataset COIL-20

此混杂在其他类别中。图7展示了不同拍摄角度的敞篷跑车,可以发现各个角度的敞篷跑车在大小和外形上有很大的差异,不易聚成一簇。

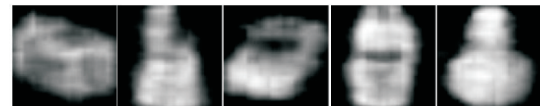


图7 不同拍摄角度的敞篷跑车
Fig.7 Roadster with different poses

另外,类别不同、外形相似的数据也很容易发生错误聚类,如图8所示。图8(a)中的物体从某个角度看均为长方形,即这些物体的全局特征差异性较小,这可能是导致这些数据的特征在二维平面内相互缠绕的重要原因。图8(b)中,重建图片的很多细节未被还原,这种模糊性使得重建图片间的相似性很大。



(a) Original images with similar shapes and different categories



(b) Reconstruction images generated by ASCAE network

图8 类别不同、外形相似的物品重建前后图片

Fig.8 Images before and after reconstruction of objects with similar shapes and different categories

表2给出了本文两种网络对应的聚类方法与文献[10]中KMS算法和DEN算法在数据集COIL-20上的聚类性能比较。对比发现,本文的对称网络SCAE提供了不错的特征表示,使得K-means的聚类精度高于文献[10]中的两种算法,而基于非对称网络ASCAE特征表示的聚类精度最高。由于DEN算法的目标函数包括局部结构和稀疏性,能扩大输入图像和特征表示的互信息,因而聚类结果的NMI值超越了本文方法,但ACC并不是很好。由此验证了本文所提方法的SCAE和

表 2 4 种算法在数据集 COIL-20 上的聚类性能比较
Table 2 Comparison of clustering performances of four algorithms on dataset COIL-20

Method	ACC	NMI
KMS ^[10]	0.592	0.767
DEN ^[10]	0.725	0.870
SCAE	0.749	0.788
ASCAE	0.754	0.811

ASCAE 网络在该数据集上有很好的特征表示能力。

3.2 手写数字数据集

本组实验在手写数字集 MNIST 和 USPS 上进行。在这 2 个数据集上,SCAE 和 ASCAE 方法均是通过最小化目标函数 $J(\theta; X; \lambda)$ 更新网络权值,然后使用 K-means 算法进行聚类,其与文献 [10] 中 4 种算法的聚类性能比较如表 3 所示。

表 3 6 种算法在数据集 MNIST 及 USPS 上的聚类性能比较

Table 3 Comparison of clustering performances of six algorithms on datasets MNIST and USPS

Method	MNIST		USPS	
	ACC	NMI	ACC	NMI
KMS ^[10]	0.535	0.531	0.535	0.531
AEC ^[10]	0.760	0.669	0.715	0.651
IEC ^[10]	0.609	0.542	0.767	0.641
DEC ^[10]	0.843		0.743	0.724
SCAE	0.766	0.731	0.663	0.673
ASCAE	0.918	0.837	0.729	0.663

由表 3 可知,在 MNIST 数据集上,聚类方法 ASCAE 比 SCAE 学习到了更利于聚类的特征表示,使得 K-means 提供了 0.918 的最高聚类精度。在 USPS 数据集上,ASCAE 较 SCAE 具有很大优势,但均不及深度聚类算法 IEC 和 DEC,ASCAE 落后 1.4%~3.8% 的聚类精度。

在 USPS 数据集上,ASCAE 网络的 F 层特征的聚类精度和单个数字的聚类精度如图 9 所示。数字 0 和 6 的聚类精度最高,达到 0.907 和 0.91; 而数字 2,3,5 和 9 的聚类精度均小于全部数字的聚类精度 0.729。

为进一步探究 ASCAE 方法在 USPS 数据集上聚类性能不高的原因,图 10 中给出了手写数字 2,5 和 9 的部分图片。观察图 10 可以发现,一些手写数字本身具有类别不确定性,例如,有些手写数字 2 很像数字 3 或 5,有些手写数字 5 很像 6 或 3,有些手写数字 9 跟 0 或 7 的相似性很高。另外书写顺序的不确定性,例如,从左上到左下书写的数字 5

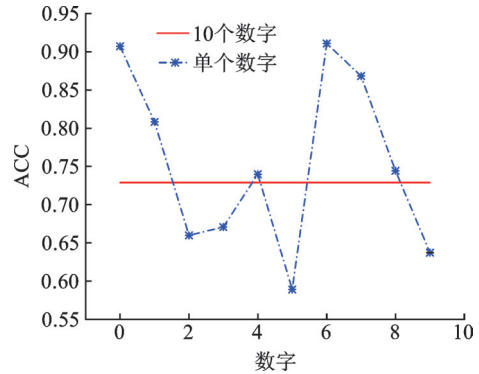


图 9 数据集 USPS 上单个数字和全部数字的聚类精度
Fig.9 Comparison of clustering accuracy of all digits and individual digit on dataset USPS

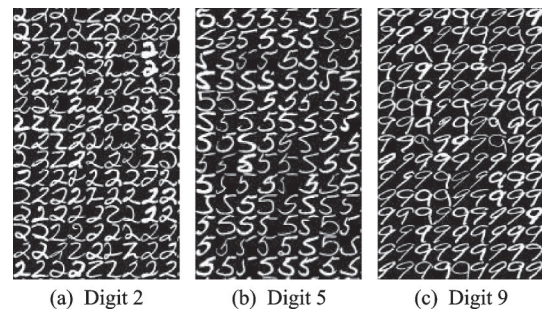


图 10 数据集 USPS 中手写数字 2,5 和 9 的部分图片
Fig.10 Some images of handwritten digits 2, 5 and 9 from dataset USPS

对比从右上到左下书写的数字 5 有很大的形变。这些从根源上就有类别不确定性的图像,使得学习到的特征很难聚合在一起,给聚类带来了巨大的挑战。

图 11, 12 给出了 ASCAE 网络在数据集



图 11 数据集 MNIST 上 ASCAE 网络的 F 层特征的二维展示

Fig.11 2-D visualization of F -layer features of ASCAE network on dataset MNIST



图 12 数据集 USPS 上 ASCAE 网络的 F 层特征的二维展示

Fig.12 2-D visualization of F -layer features of ASCAE network on dataset USPS

MNIST 和 USPS 的 F 层特征的二维展示。图 11 中区域之间有比较明显的边界,但是每个区域都有不同颜色的散点。而图 12 的下部分有 2 个子区域的边界不太明显,辨识度下降。

3.3 人脸数据集

本组实验在 4 个人脸数据集(CAS-PEAL-R1, BIO-ID, IMM 和 UMISTS)上进行,部分图片见图 13。基于 ASCAE 和 SCAE 方法的聚类结果见表 4。

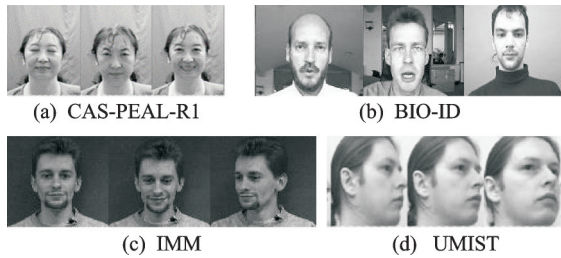


图 13 4 个人脸数据集上的部分图片

Fig.13 Some images of four human face datasets

表 4 ASCAE 和 SCAE 方法在 4 个人脸数据集上的聚类性能

Table 4 Clustering performance of ASCAE 和 SCAE methods on four human face datasets

Dataset	SCAE		ASCAE	
	ACC	NMI	ACC	NMI
CAS-PEAL-R1	0.740	0.895	0.880	0.947
BIO-ID	0.901	0.954	0.864	0.922
IMM	0.542	0.754	0.621	0.804
UMISTS	0.485	0.668	0.485	0.668

由表 4 可知,在数据集 CAS-PEAL-R1 和 IMM 上,聚类方法 ASCAE 比 SCAE 具有更优的聚类性能。图像背景简单,大部分为正脸且光线较亮,特征表示对眼睛、嘴巴和鼻子的形状有较高的辨识度。在 UMIST 上,这 2 个方法的聚类性能远远小于其他 3 个数据集。原因可能是人脸视图有侧面和正面,姿势变化较大。但在 BIO-ID 上,SCAE 方法带来的聚类结果优于 ASCAE。原因可能是图像有较复杂的室内背景,ASCAE 网络的 3 个全连接层整合局部特征后,使得特征表示的辨识度反而下降了。

3.4 代价分析

ASCAE 网络多一个全连接层,其网络参数增加 $50 \times 50 + 50 = 300$ 个,即空间存储量增加不大。以数据集 MNIST 为例,训练次数设为 200。两个网络学习特征的目标损失和运行时间如表 5 所示,可以看出 ASCAE 的运行时间较长,但目标损失收敛值也较小。

表 5 ASCAE 和 SCAE 网络在数据集 MNIST 上迭代 200 次的计算代价

Table 5 Computational cost of ASCAE and SCAE networks iterating 200 times on dataset MNIST

Network	Cost	Running time
SCAE	838.1	26 min 50 s
ASCAE	809.0	28 min 31 s

4 结 论

由于经典自编码器的特征表示能力有限,使得基于自编码器的深度聚类方法在复杂图像数据集上的聚类性能不高。为此,本文提出了一种新的深度聚类方法 ASCAE。本文的深度聚类方法通过改善网络结构和目标损失函数,提取出更有辨识力的聚类特征来提高聚类性能。实验结果表明,在 7 个图像数据集上尤其是图片简单、易区分类别的数据集上,聚类方法 ASCAE 更能提取到适合聚类的特征表示。但是对姿态形变较大或类别不确定性较大的数据集,其聚类效果仍有提升空间。

参考文献:

- [1] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. Nature, 2015, 521(7553): 436-44.
- [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 [J]. Communications of the ACM, 2017, 60(6): 84-90.
- [4] DENG Jia, DONG Wei, SOCHER R, et al. ImageNet: A large-scale hierarchical image database [C]// Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2009.
- [5] FISCHER A, IGEL C. An introduction to restricted Boltzmann machines [C]// Proceedings of Iberoamerican Congress on Pattern Recognition. [S.l.]: Springer, 2012.
- [6] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [J]. Nature, 1986, 323(6088): 533-536.
- [7] MASCI J, MEIER U, CIREŞAN D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction [C]// Proceedings of International Conference on Artificial Neural Networks. [S.l.]: Springer, 2011.
- [8] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C]// Proceedings of Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2014.
- [9] GHASEDI D K, HERANDI A, DENG Cheng, et

- al. Deep clustering via joint convolutional autoencoder embedding and relative entropy minimization [C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.]:IEEE, 2017.
- [10] LI Fengfu, QIAO Hong, ZHANG Bo. Discriminatively boosted image clustering with fully convolutional auto-encoders[J]. Pattern Recognition, 2018, 83: 161-173.
- [11] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks [C]//Proceedings of the 4th International Conference on Learning Representations. San Juan, Puerto Rico : [s.n.] , 2016: 97-108
- [12] XIE Junyuan, GIRSHICK R, FARHADI A. Unsupervised deep embedding for clustering analysis[C]//Proceedings of 33rd International Conference on Machine Learning. New York, USA: [s.n.], 2016.
- [13] YANG Jianwei, PARIKH D, BATRA D. Joint unsupervised learning of deep representations and image clusters[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.] : IEEE, 2016.
- [14] CARON M, BOJANOWSKI P, JOULIN A, et al. Deep clustering for unsupervised learning of visual features [C]//Proceedings of the European Conference on Computer Vision .Munich, Germany: ECCV, 2018.
- [15] WU Zhirong, XIONG Yuanjun, YU S X, et al. Unsupervised feature learning via non-parametric instance discrimination [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.[S.l.] : IEEE, 2018:3733-3742.
- [16] KANG Guoliang, JIANG Lu, YANG Yi, et al. Contrastive adaptation network for unsupervised domain adaptation [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.] : IEEE, 2019:4893-4902.
- [17] ZHANG Liheng, QI Guojun, WANG Liqiang, et al. AET vs. AED: Unsupervised representation learning by auto-encoding transformations rather than data [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.] : IEEE, 2019:2547-2555.
- [18] HE Kaiming, FAN Haoqi, WU Yuxin, et al. Momentum contrast for unsupervised visual representation learning [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.] : IEEE, 2020: 9726-9735.
- [19] WANG Wenguan, SONG Hongmei, ZHAO Shuyang, et al. Learning unsupervised video object segmentation through visual attention[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.] : IEEE, 2019:3064-3074.
- [20] VO H V, BACH F, CHO M, et al. Unsupervised image matching and object discovery as optimization[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.] : IEEE, 2019: 8287-8296.
- [21] CARON M, BOJANOWSKI P, MAIRAL J, et al. Unsupervised pre-training of image features on non-curated data [C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.] : IEEE, 2019: 2959-2968.
- [22] JI Xu, HENRIQUES J F, VEDALDI A. Invariant information clustering for unsupervised image classification and segmentation [C]//Proceedings of the IEEE International Conference on Computer Vision. [S.l.] : IEEE, 2019: 9865-9874.
- [23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]//Proceedings of the 3rd International Conference on Learning Representations. San Diego, CA: [s.n.], 2015: 1-14.
- [24] HARIHARAN B, ARBELÁEZ P, GIRSHICK R, et al. Hypercolumns for object segmentation and fine-grained localization [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.] : IEEE, 2015:447-456.

(编辑:孙静)