深度学习中的自编码器的表达能力研究

王雅思 姚鸿勋 孙晓帅 许鹏飞 赵思成
(哈尔滨工业大学计算机科学与技术学院 哈尔滨 150001)

摘 要 近年来,深度学习框架和非监督学习方法越来越流行,吸引了很多机器学习和人工智能领域研究者的兴趣。从深度学习中的"构造模块"入手,主要研究自编码器的表达能力,尤其是自编码器在数据降维方面的能力及其表达能力的稳定性。从深度学习的基础方法入手,旨在更好地理解深度学习。第一,自编码器和限制玻尔兹曼机是深度学习方法中的两种"构造模块",它们都可用作表达转换的途径,也可看作相对较新的非线性降维方法。第二,重点探究了对于视觉特征的理解,自编码器是否是一个好的表达转换途径。主要评估了单层自编码器的表达能力,并与传统方法PCA进行比较。基于原始像素和局部描述子的实验验证了自编码器的降维作用、自编码器表达能力的稳定性以及提出的基于自编码器的转换策略的有效性。最后,讨论了下一步的研究方向。

关键词 深度学习,表达转换,数据降维,单层自编码器
中图法分类号 TP391.4 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.9.012

Representation Ability Research of Auto-encoders in Deep Learning

WANG Ya-si YAO Hong-xun SUN Xiao-shuai XU Peng-fei ZHAO Si-cheng
(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract Deep learning frameworks and unsupervised learning methods have become increasingly popular and attracted the attention of many researchers in machine learning and artificial intelligence fields. This paper started from the "building-blocks" of deep learning methods and focused on the representation ability research of auto-encoders, especially on auto-encoders' ability to reduce dimensionality and the stability of their representation ability. We expected that starting from the basis of deep learning can help us understand it better. Firstly, auto-encoder and restricted Boltzmann machine are two "building-blocks" of deep learning methods, both of which can be used to transform representation and be seen as relatively new nonlinear dimensionality reduction methods. Secondly, we investigated whether auto-encoder is a good representation transformation method under the context of understanding visual features, including evaluating single-layer auto-encoder's representation ability compared with classic methodology principal component analysis. Experiments based on original pixels and local descriptors demonstrate auto-encoder's ability to reduce dimensionality, the stability of its representation ability and the effectiveness of proposed AE-based transformation strategy. Finally, future research direction was discussed.

Keywords Deep learning, Representation transformation, Dimensionality reduction, Single-layer auto-encoder

1 引言

1.1 特征表达及数据降维

真实世界中存在大量非常复杂的事物和现象,通常我们希望能够以一种更加简洁且完整的方式去表示一个事物或现象,这就需要去揭示隐藏在复杂表象下的事物或现象的客观规律。从某个事物或现象(例如天气状况)中抽象出一些数据(如温度、湿度、风力等),通过多个变量来表示或描述一个现象,这个多维变量叫做特征。

特征作为机器学习系统的原材料,对于最终模型的影响毋庸置疑。机器学习算法的性能在很大程度上取决于数据表达或特征表达的选择,当数据能够被很好地表达为特征时,即便使用简单的模型也可达到满意的精度。故在实际应用机器学习算法时,很重要的一个步骤是怎样预处理数据以得到一个好的特征表达。

真实世界中的数据通常是高维的。对高维数据的处理包括两点特性[1]。第一点是"维数灾难",它给后面的数据处理带来困难,是处理高维数据时遇到的最大问题之一;第二点是"维数福音",高维数据中包含着关于客观事物和现象的极为全面和丰富的信息,蕴含着解决问题的可能性,当然也含有很多冗余信息。

作为一类普遍存在的规律,在大多数情况下我们观察到

到稿日期:2014-08-22 返修日期:2014-10-22 本文受国家自然基金项目(61472103),国家自然基金重点项目(61133003)资助。
王雅思(1990-),女,硕士,主要研究方向为计算机视觉、机器学习,E-mail:wangyasi@hit.edu.cn;**姚鸿勋**(1965-),女,博士,教授,主要研究方向为图像处理与模式识别、多媒体信息处理与挖掘、自然人机交互技术,E-mail:h.yao@hit.edu.cn(通信作者);**孙晓帅**(1984-),男,博士,主要研究方向为计算机视觉、模式识别,E-mail:xiaoshuaisun@hit.edu.cn;**许鹏飞**(1984-),男,博士,主要研究方向为视觉特征提取和表达,E-mail:pfxu@hit.edu.cn;**赵思成**(1986-),男,博士,主要研究方向为情感计算、多媒体分析,E-mail:zsc@hit.edu.cn。

的从表面上看是高维的、复杂的事物或现象，实际上是可以用少量的简单变量来支配的。处理高维数据的关键是在众多的因素中找到事物的本质规律。在图像处理领域，原始像素值作为初级特征表达时通常维度很高，需要采用合适的降维方法对图像高维特征进行处理，以得到更加简洁而有效的特征表达。数据降维技术可以克服“维数灾难”，在模式识别、数据挖掘、机器学习等领域都具有广泛应用，相应研究方兴未艾。

1.2 人工神经网络及深度学习

人工神经网络是一种数学(计算)模型，由大量神经元构成，能够模仿生物神经网络的结构和功能。人工神经网络是一种自适应系统，能够根据外部输入信息改变内部结构，常用其来对输入和输出之间的复杂关系进行建模。人工神经网络的发展[2]道路是曲折的。感知机是现代神经计算的起始点，1962 年感知机的学习收敛定理被证明，引发了以感知机为代表的第一次神经网络研究发展的高潮；1965 年，M. Minsky 和 S. Papert 指出感知机的缺陷，对神经网络研究持悲观态度，使神经网络的研究由兴起进入停滞期；到 20 世纪 80 年代初，J. J. Hopfield 等人的相关工作显示出神经网络的潜力，使神经网络的研究从停滞期进入繁荣期；而到 90 年代中后期，随着支持向量机的出现，研究者们认识到人工神经网络的一些局限性，神经网络的研究再次陷入低潮。

人工神经网络又一次沉寂了多年，2006 年发表在 Science 杂志上的一篇文章[3]打破了这一僵局。自那时起，深度学习的框架和非监督的学习方法被很多机器学习和人工智能领域的研究者广泛学习和研究。深度学习研究的起源正是人工神经网络。深度学习方法实际上指的是一种深层的结构，而人工神经网络中含有多个隐含层的多层感知器，这恰恰就是一种深度学习的结构。深度学习方法学习的是一种多层次的非线性结构，能够对输入数据进行分布式表示，具有强大的从数据中抽取本质特征的能力，从而能得到更加抽象的特征表达。最近的研究结果证明了深度学习方法确实取得了目前最好的结果，无论是在图像、语音领域还是自然语言处理领域。特别地，深度学习在语音领域已经成功应用到工业界。在图像分析领域，也有很多好的结果。通过构造一个九层局部连接的稀疏自编码器，也就是将 9 个单层稀疏自编码器连接成一个深层结构，Le 等人[4]发现网络对高层语义概念敏感，例如猫脸和人的身体。Krizhevsky 等人[5]通过使用一个深度卷积神经网络在 2012 年的 ImageNet 比赛中取得了创纪录的结果。Zeiler 等人[6]则在 Caltech-101 和 Caltech-256 数据集上取得了最好的结果。

深度学习方法虽然取得了很好的结果，并且被广泛应用，但它实际上有点像一个“黑箱子”，没有非常充分并且严格的理论体系来支撑。所以存在一个问题：使用深度学习方法取得了好的结果，但是在理论上不知道为什么。在深度学习方向，人们通过采用越来越深度的模型和越来越复杂的非监督学习算法，取得了很多进步[4,5,7]，然而对于深度学习方法的构造模块，例如单层自编码器、单层限制玻尔兹曼机的研究，则相对较少。

作为深度学习方法中的“构造模块”，单层的自编码器和单层的限制玻尔兹曼机[8]都可以作为表达转换的途径，学到一个原始数据的新的表达，而当限制第二层的节点数小于第一层的节点数时，又可以同时达到对数据进行降维的效果。我们期望从深度学习的基础方法入手，着重研究单层自编码器的表达能力，以更好地理解深度学习。研究的课题背景如图 1 所示。

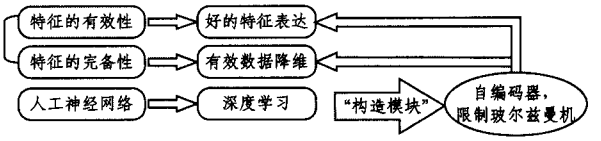


图 1 课题背景

2 深度学习方法基础研究

这一部分简要介绍了深度学习方法中的两种“构造模块”，包括自编码器和限制玻尔兹曼机，以及什么是表达转换。

2.1 深度学习“构造模块”之自编码器

自编码器(Auto-encoder，AE)是一个 3 层的神经网络，将输入表达 X 编码为一个新的表达 Y，然后再将 Y 解码回 X。这是一个非监督学习算法，使用反向传播算法来训练网络使得输出等于输入。

当向网络中添加一些限制时，可以学到一些关于输入表达的有趣结构。当隐含层节点数 d 比输入层节点数 n 小时，可以得到一个输入的压缩表达。当 d 比 n 大时，添加一些限制，比如稀疏限制，会得到类似于稀疏编码的结果。

2.2 深度学习“构造模块”之限制玻尔兹曼机

限制玻尔兹曼机(Restricted Boltzmann Machine，RBM)模型是一个二部图。当有输入表达 v 和条件概率 $p(h|v)$ 时，可以得到隐含层表达。反过来，当有隐含层表达 h 和条件概率 $p(v|h)$ 时，可以得到一个新的输入表达，如此反复下去，如图 2 所示。通过调整参数，我们想要的最终输出与原始的输入表达相同，那么此时隐含层表达就可以看作输入表达的一个新的表达。

由于自编码器和限制玻尔兹曼机这两种方法的大体思路是一致的，主要在于训练过程不同，因此在实验部分，着重考虑自编码器部分。

图 2 限制玻尔兹曼机的结构

2.3 表达转换

表达转换指的是将原始表达转换为另一个不同的表达，本文主要考虑新的表达维数小于原始表达维数的情形。在模式识别系统中，使用到的特征很重要，经常需要将高维的冗余的原始特征转换为低维的保留有效信息的特征，也就是特征转换。

特征转换可以分为两类：第一类是“特征选择”，即从原始表达中选择一个子集作为新的表达；第二类是“特征抽取”，即将原始表达投影到一个低维特征空间中以得到一个更加紧凑的表达。

3 自编码器降维作用及表达能力的稳定性

这一部分主要探究了对视觉特征的理解及自编码器是否是一个好的表达转换途径。评估了单层自编码器分别基于原始像素和局部描述子的降维作用及表达能力的稳定性，并将其与传统的主成分分析法进行比较。

3.1 自编码器基于原始像素的降维作用及表达能力的稳定性

在 MNIST 数据集上实现两种表达转换和一个 softmax

分类器，探究算法性能如何随着转换得到的新的表达维数的变化而变化，通过算法的性能（分类准确率）来评估单层自编码器基于原始像素的降维作用及表达能力的稳定性。

在 MNIST 数据集上的实验的基本思路如下：首先有训练数据库和测试数据库，库中图像大小为 28 * 28 像素，共 784 个像素，把这 784 个像素值拉成一个向量，作为原始特征。将原始特征进行表达转换，得到一个新的特征。表达转换包括特征选择和特征抽取两类，其中对于特征抽取，分别用自编码器和主成分分析法（Principal Component Analysis，PCA）降维。得到新的特征后，训练 softmax 分类器，用测试数据测试，得到分类准确率。当新的特征维数变化时得到不同的分类准确率，记录下来。在 MNIST 数据集上的分类实验概述如图 3 所示。

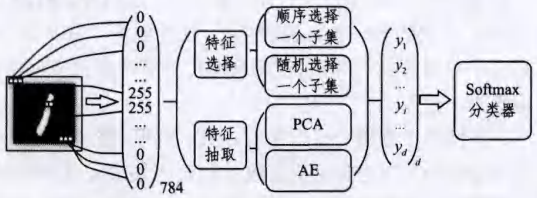


图 3　MNIST 数据集上分类的概述

3.1.1　特征选择

对于特征选择，考虑直接从原始像素表达中选择一个可变大小的子集来进行表达转换，其分为两小类：顺序地选择一个子集和随机地选择一个子集。得到新的特征之后，训练 softmax 分类器，然后进行测试。当随机选择子集时，运行程序 150 次，取平均结果。

3.1.2　特征抽取

对于特征抽取，分别使用自编码器和主成分分析法来进行表达转换，其中主成分分析法作为线性特征抽取方法，自编码器作为非线性特征抽取方法。得到新的特征之后，同样训练 softmax 分类器后进行测试。对于 AE 降维，运行程序 10 次，分别取平均结果和最好结果。

3.2　自编码器基于局部描述子的降维作用及表达能力的稳定性

通过 SIFT 匹配结果来评估单层自编码器基于局部描述子——SIFT 描述子的降维作用及表达能力的稳定性。

SIFT 匹配的基本思路如下：首先查询 SIFT 特征点和包含很多 SIFT 特征点的库，将查询 SIFT 特征点与库中的 SIFT 特征点进行匹配，得到一个匹配结果；之后将查询 SIFT 特征点及库中的 SIFT 特征点分别用自编码器和主成分分析法进行降维，降维后再进行同样的匹配过程，得到两种经过降维之后的匹配结果。通过比较 SIFT 匹配结果来检验表达转换过程是否有效。SIFT 匹配概述如图 4 所示。SIFT 匹配部分可分为两种情况：基于单幅图像的 SIFT 匹配和基于多幅图像的 SIFT 匹配。

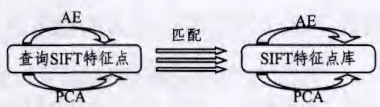
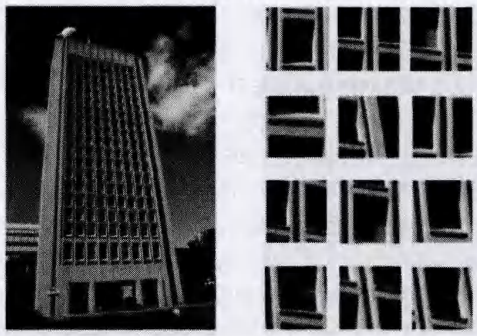


图 4　SIFT 匹配概述

3.2.1　基于单幅图像的 SIFT 匹配

第一种情况是基于单幅图像的 SIFT 匹配，在一幅内部存在重复结构的图像中进行 SIFT 匹配。

首先从这幅图像中得到 n 个 SIFT 特征点 $s_1, s_2, \cdots, s_n$。通过在合适的尺度上展示出以某个特征点为中心的块来可视化这个特征点。由于图像内部存在重复结构，将这 n 个 SIFT 特征点可视化，会得到很多重复结构。图 5 显示了原始图像和其中 12 个不同的 SIFT 特征点可视化后的结果。



（a）原始图像　　（b）12 个 SIFT 特征点

图 5

以第 i 个 SIFT 特征点（s_i）为例，把它作为查询特征点，其余特征点 $s_1, s_2, \cdots, s_{i-1}, s_{i+1}, \cdots, s_n$ 作为库。将查询特征点与库中特征点进行匹配，匹配后对库中特征点按相似度排序，由于匹配是基于 SIFT 描述子的，因此选用欧氏距离作为度量，选择前 k（比如 15）个最相似的特征点 $m_1, m_2, \cdots, m_k$。之后，分别用 PCA 和 AE 对原始 SIFT 描述子降维，从 128 维降到 d 维（将新的表达分别叫作“PCA-表达”和“AE-表达”），然后在低维空间中对新的描述子执行相同的 SIFT 匹配操作，具体过程如图 6 所示。

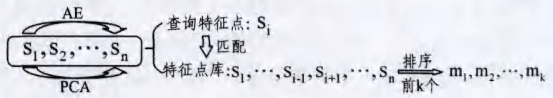
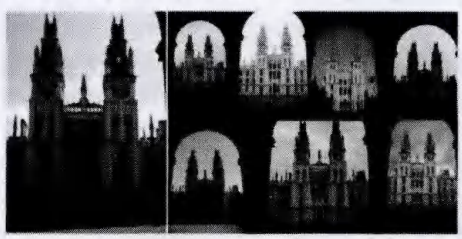


图 6　一幅图像内的 SIFT 匹配过程

3.2.2　基于多幅图像的 SIFT 匹配

第二种情况是基于多幅图像的 SIFT 匹配。给定同一场景的多幅图片，将其中一幅图片作为查询，其余作为库，同样执行 SIFT 匹配的操作。

从牛津建筑数据集中选择了同一建筑的 8 幅图像。把其中一幅作为查询图像，其余 7 幅作为库。从查询图像中得到 n 个 SIFT 特征点 $s_1, s_2, \cdots, s_n$，将其作为查询 SIFT 特征点库，从包含 7 幅图像的库中得到 m 个 SIFT 特征点 $d_1, d_2, \cdots, d_m$，将其作为 SIFT 特征点字典。由于这 7 幅图像包含的是同一栋建筑，同样将 SIFT 特征点可视化，当将查询 SIFT 特征点与 SIFT 特征点字典中的 SIFT 特征点进行匹配时，也会匹配到重复的结构。图 7 显示了牛津建筑数据集中的例子，包括查询图像和含有 7 幅图像的库。



（a）查询图像　　（b）包含 7 幅图像的库

图 7　牛津建筑数据集中的例子

以第 i 个查询 SIFT 特征点(s_i)为例，把它作为当前查询特征点。将查询特征点与字典中的 SIFT 特征点 $d_1, d_2, \cdots, d_m$ 匹配，匹配后对字典中的特征点按相似度(欧氏距离)排序，选择前 k(比如 30、50)个最相似的特征点 $m_1, m_2, \cdots, m_k$。为了使结果显示得更加清楚，只挑选出这 k 个特征点中尺度大于 5 的特征点，将其可视化。之后，同样分别用 PCA 和 AE 对查询 SIFT 特征点库和 SIFT 特征点字典降维，从 128 维降到 d 维，然后在低维空间中对新的 d 维描述子执行相同的 SIFT 匹配操作。具体过程如图 8 所示。

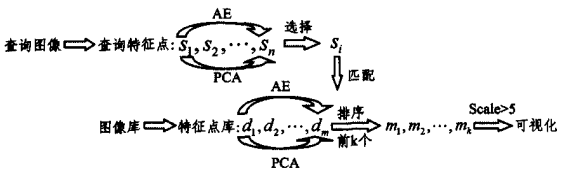


图 8　多幅图像内的 SIFT 匹配过程

4　实验结果与分析

4.1　自编码器基于原始像素的降维作用及表达能力的稳定性

通过比较 MNIST 数据集上的分类结果来检验表达转换(降维)过程是否有效。为了更简洁，只考虑数字“1”和“2”。总共有 12700 个训练样本(分别有 6742 个“1”和 5958 个“2”)和 2167 个测试样本(分别有 1135 个“1”和 1032 个“2”)。

4.1.1　特征选择部分的实验结果与分析

在特征选择部分，记录了当从原始表达中顺序或随机选择一个可变子集作为表达转换的途径，将选择的子集作为新的表达时，随着新的表达维数的变化，softmax 分类器性能也就是分类准确率的变化情况。

当随机选择子集时，对于每一维度，运行程序 150 次，得到平均结果。实验结果较直观，如图 9 所示。“O”代表顺序选择子集，“*”代表随机选择子集。

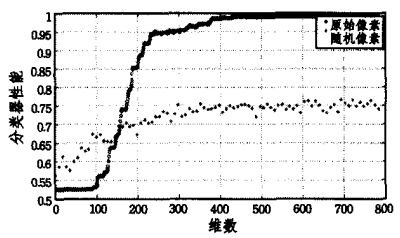


图 9　在特征选择部分分类器性能的变化趋势

从图 9 可以看出，对于顺序地选择子集，分类器性能单调迅速递增，具体来说：

(1)当维数小于 100 时，对应性能很低，而且几乎没有变化；

(2)当维数介于 100～240 之间，随着维数的增加，对应性能单调迅速递增，有点类似于阶梯式增长；

(3)当维数介于 240～400 之间，随着维数的增加，对应性能单调缓慢递增，增长比较缓慢且稳定；

(4)当维数大于等于 400 时，对应性能几乎接近于 1。

其原因可以通过观察 MNIST 中的图像看出：

(1)当维数小于 100 时，前 100 个像素几乎不能表达出任何信息，所以对应性能很低；

(2)当维数介于 100～400 时，随着维数的增加，表达的信息也逐渐递增，对应性能也相应递增，先迅速递增后缓慢递增，在整个阶段是单调增加的；

(3)当维数接近 400 时，数字的轮廓已经大致能够被完全表达出来，所以对应性能接近于 1。

同时，从图 9 也可以看出，对于随机地选择子集，分类器的性能保持相对稳定地增长，具体来说：

(1)当维数小于 300 时，随着维数的增加，对应性能非单调缓慢增长，总体呈增长趋势；

(2)当维数大于 300 时，随着维数的增加，对应的性能总体稳定保持不变。

其原因可以通过观察 MNIST 中的图像看出：

(1)当维数小于 300 时，随着维数的增加，虽然是随机地选择子集，但是表达的信息还是逐渐递增；同时由于是随机选择子集，对应性能不再是单调递增，但总体呈增长的趋势；

(2)当维数大于 300 时，随着维数的增加，由于是随机选择子集，相当于顺序选择子集后打乱顺序，对应分类器性能有上下浮动，总体保持不变，维数的增加实际上没有带来表达信息的提高。

总体来说，对于顺序选取子集，随着维数的增大，表达的信息逐渐增强，所以性能曲线呈单调递增；而对于随机选取子集，它实际上相当于顺序选取子集后打乱顺序，故图像中数字的轮廓信息被打乱，因此实验结果与顺序选取子集有较大区别。

4.1.2　特征抽取部分的实验结果与分析

在特征抽取部分，记录了当使用自编码器和主成分分析法对原始特征进行表达转换时，随着新的表达维数变化，softmax 分类器性能的变化情况。

对于 PCA-表达，当保留 99.9999%的方差变化时，PCA-表达维数为 541，记录当维数从 1 变化到 541 时的性能变化。对于 AE-表达，记录当维数从 1 变化到 1027 时的性能变化，对每一维度，运行程序 10 次，得到平均和最好结果。实验结果如图 10 所示，“▽”代表用 PCA 进行表达转换，“O”和“*”分别代表用 AE 进行表达转换对应的最好和平均结果。

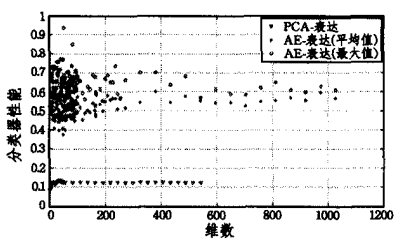


图 10　在特征抽取部分分类器性能的变化趋势

从图 10 可以看出，PCA 降维之后，分类器性能非常低，而 AE 降维之后对应的分类器性能比 PCA 降维之后要高出很多，这说明自编码器的降维作用优于主成分分析。当 AE-表达的维数变化时，其对应的性能没有变化很快，虽上下起伏，但保持相对稳定，这说明自编码器表达能力的稳定性。

本节通过分类器性能验证了单层自编码器基于原始像素的降维作用及表达能力的稳定性，以及提出的基于自编码器的转换策略的有效性。

4.2　自编码器基于局部描述子的降维作用及表达能力的稳定性

通过比较 SIFT 匹配结果来检验表达转换(降维)过程是否有效。

4.2.1　基于单幅图像的实验结果与分析

这一部分考虑基于单幅图像的 SIFT 匹配。图 11 展示

了一幅图像内的 SIFT 匹配的两个可视化结果。在可视化结果中，第一列为查询 SIFT 特征点可视化，第二列为匹配到的前 15 个最相似的 SIFT 特征点可视化；第一行为使用原始 SIFT 特征点进行 SIFT 匹配的结果，第二行和第三行为使用 PCA 对原始 SIFT 特征点降维后进行 SIFT 匹配的结果，其中第二行为使用 PCA 降维时保留 95%的方差变化(此时将原始 SIFT 特征点降到 16 维)，第三行为使用 PCA 降维时保留 99%的方差变化(此时将原始 SIFT 特征点降到 39 维)，第四行和第五行为使用 AE 对原始 SIFT 特征点降维后进行 SIFT 匹配的结果，其中第四行为使用 AE 降维时隐含层节点数为 16，第五行为使用 AE 降维时隐含层节点数为 39。

从结果可以看出，原始 SIFT 匹配的结果不是很好，这是因为直接使用欧氏距离作为匹配的准则。当使用相同的匹配准则时，PCA-表达和 AE-表达的匹配结果比原始 SIFT 特征点的匹配结果要好，这说明了使用 PCA 和 AE 进行表达转换的有效性。在使用相同的匹配准则并且将原始 SIFT 特征点降到相同维度(16 维和 39 维)的情形下，AE-表达的匹配结果比 PCA-表达的匹配结果更好，并且随着维数的变化，匹配结果保持稳定，这说明了自编码器的降维作用优于主成分分析法，并且自编码器的表达能力具有很好的稳定性。综上可知，AE-表达的匹配结果最好，这也证明了基于自编码器的转换策略的有效性。

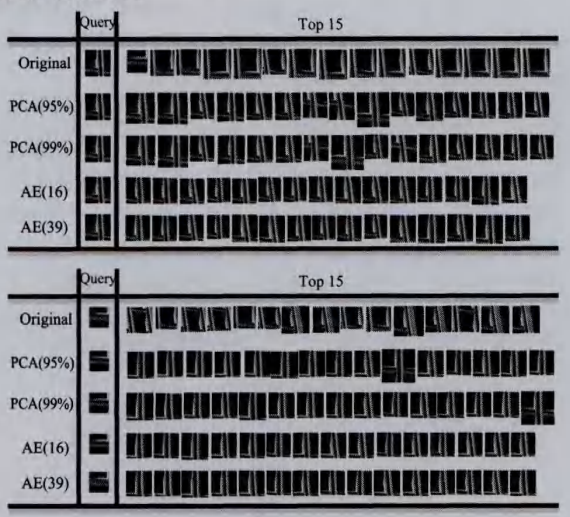


图 11　一幅图像内的 SIFT 匹配的两个可视化结果

4.2.2　基于多幅图像的实验结果与分析

这一部分考虑基于多幅图像的 SIFT 匹配。图 12 展示了在多幅图像中的 SIFT 匹配的一个可视化结果。我们仅可视化出尺度大于 5 并且匹配正确的 SIFT 特征点。在可视化结果中，第一列为查询 SIFT 特征点可视化，第二列为匹配到的前 30 个最相似的 SIFT 特征点可视化，仅显示出了匹配基本正确的 SIFT 特征点，第三列为匹配到的前 30－50 个最相似的 SIFT 特征点可视化，仅显示出了匹配基本正确的 SIFT 特征点；第一行为使用原始 SIFT 特征点进行 SIFT 匹配的结果，第二行为使用 PCA 对原始 SIFT 特征点降维后进行 SIFT 匹配的结果，使用 PCA 降维时保留 99%的方差变化，此时将原始 SIFT 特征点降到 39 维，第三行为使用 AE 对原始 SIFT 特征点降维后进行 SIFT 匹配的结果，使用 AE 降维时隐含层节点数为 39。

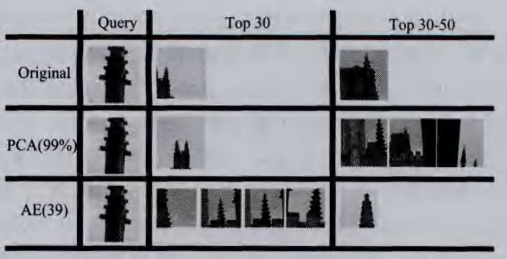


图 12　多幅图像中的 SIFT 匹配的可视化结果

从结果可以看出，由于直接使用欧氏距离作为匹配的准则，原始 SIFT 匹配的结果仍然不是那么的好。当使用相同的匹配准则时，PCA-表达和 AE-表达的匹配结果比原始 SIFT 特征点的匹配结果要好，这说明了使用 PCA 和 AE 进行表达转换的有效性。可以看到，不论是在前 30 还是前 50(尤其在前 30)匹配中，在使用相同的匹配准则并将原始 SIFT 特征点降到相同维度(39 维)的情形下，AE-表达的匹配结果比 PCA-表达更好，这说明了自编码器的降维能力优于主成分分析法。当使用自编码器将原始 SIFT 特征点降到不同维度时，匹配效果相近，这说明了自编码器表达能力的稳定性。综上可知，AE-表达的匹配结果最好。

本节通过匹配结果验证了单层自编码器基于局部描述子的降维作用及表达能力的稳定性，及提出的基于自编码器的转换策略的有效性。

结束语　本文从深度学习方法的“构造模块”入手，以更好地理解深度学习。自编码器和限制玻尔兹曼机都可看作表达转换的途径，同时也可看作相对较新的非线性降维方法。重点探究了对于视觉特征的理解，自编码器是否是一个好的表达转换途径。评估了单层自编码器的表达能力，并将其与传统方法主成分分析法进行比较。基于原始像素和基于局部描述子的实验都验证了单层自编码器的降维作用要优于主成分分析法以及单层自编码器表达能力具备一定的稳定性，同时也验证了提出的基于自编码器的转换策略的有效性。

我们实际上尝试了一些其他方式来评估自编码器的表达能力，也讨论了自编码器的一些变种以取得更好的性能。比如在对手写数字集进行分类时，将自编码器的隐含层节点设为 10 个，当图像为数字“1”时，隐含层的第一个节点被激活，其余节点被抑制，以此类推。再比如对原始特征进行降维时，先用主成分分析法对原始特征预先降维，得到 PCA-表达，将隐含层节点的激活值设为 PCA-表达，来训练自编码器以得到网络中各个权值等。但是这些都没有取得很好的效果，这也表明，单层自编码器和简单的非线性变换不具备足够的能力来很好地建模高维数据，这也是现在越来越多地使用深度模型的原因。

对于表达转换和数据降维问题，还有很多地方值得研究。未来的工作中将会继续这一课题，更加重视理论研究。比如 AE 和 RBM 中隐含层节点个数的选择，这是一个一直都没有被很好解决的问题，引入数学中估计高维数据的本征维数的方法，也许可以为隐含层节点个数的设计给出一些参考性建议。这也是下一步要解决的问题之一。

参 考 文 献

[1]　谭璐. 高维数据的降维理论及应用[D]. 长沙：国防科学技术大学，2005

(下转第 65 页)

小。因此二者的训练代价差别不大，而姿态实验的训练代价比较大。Huang 和 CLPM 方法是本文方法的 1～13 倍，而 CLPM 方法的训练代价的增长相对稳定。基于以上实验结果，本文方法达到了最好的识别效果和最小的训练代价。

表 2　时间复杂度

方法	表情实验	分辨率实验	姿态实验
DBN	0.42	0.41	0.49
Huang	1.38	1.33	1.97
CLPM	5.61	5.52	5.68

结束语　本文提出了一种基于深度神经网络的方法来消除包括表情变化、分辨率变化以及姿态等多种变化形态对人脸识别的影响。在 RBM 堆栈的顶层增加了一个回归层，用来在一个统一的深度学习框架下完成特征提取以及分类两种任务。网络中大量参数通过无监督的误差重建方法进行初始化，随后最小化网络的输出特征与标签之间的交叉熵，通过 BP 算法来调整网络参数。在 CMU-PIE 数据库上的实验结果表明 DBN 方法相比其他人脸识别方法有更优秀的表现。在未来的工作中，我们将基于深度神经网络的方法对姿态变化的人脸识别进行更加深入的研究。

参 考 文 献

[1] Huang H, Zeng X. Super-resolution method for multi-view face recognition from a single image per person using nonlinear mappings on coherent features [J]. IEEE Signal Processing Letters, 2012, 19(4): 195-198

[2] Zou W, Yuen P. Very low resolution face recognition problem [C]// IEEE Transactions on Image Processing, 2012

[3] Zhang X, Gao Y. Face recognition across pose. A review [J]. Pattern Recognition, 2009, 42(11): 2876-2896

[4] Huang H, He H. Super-resolution method for face recognition using nonlinear mappings on coherent features [J]. IEEE Transactions on Neural Networks, 2011, 22(1): 121-130

[5] Li B, Chang H, Shan S, et al. Low-resolution face recognition via coupled locality preserving mappings [J]. IEEE Signal Processing Letters, 2010, 17(1): 20-23

[6] Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507

[7] Salakhutdinov R, Hinton G E. Learning a nonlinear embedding by preserving class neighbourhood structure [J]. International Conference on Artificial Intelligence and Statistics, 2007, 3: 412-419

[8] Zhou S, Chen Q, Wang X. Discriminative deep belief networks for image classification [C]// ICIP. 2010: 1561-1564

[9] Mohamed A R, Dahl G, Hinton G. Deep belief networks for phone recognition [C]// NIPS Workshop on Deep Learning for Speech Recognition and Related Applications. 2009

[10] Sim T, Baker S, Bsat M. The CMU pose, illumination, and expression(pie) database [C]// Proceedings of International Conference on Automatic Face and Gesture Recognition. 2002: 46-51

[11] Hinton G E, Osindero S, Teh Y-W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554

[12] Huang G B, Lee H, Learned-Miller E. Hierarchical representations for face verification with convolutional deep belief networks [C]// Proceedings of International Conference on Computer Vision and Pattern Recognition(CVPR). 2012: 2518-2525

[13] 何加浪，张宏. 神经网络在软件多故障定位中的应用研究[J]. 计算机研究与发展，2013，50(3)：619-625
He Jia-lang, Zhang Hong. Application of artificial neural network in software multi-faults location [J]. Journal of Computer Research and Development, 2013, 50(3): 619-625

[14] 周旭东，陈晓红，陈松灿. 半配对半监督场景下的低分辨率人脸识别[J]. 计算机研究与发展，2012，49(11)：2328-2333
Zhou Xu-dong, Chen Xiao-hong, Chen Song-can. Low-Resolution face recognition in semi-paired and semi-supervised scenario [J] Journal of Computer Research and Development, 2012, 49(11): 2328-2333

[15] Andrew G, Arora R, Bilmes J, et al. Deep Canonical Correlation Analysis[J]. JMLR W&CP, 2013, 28(3): 1247-1255

[16] Le Q V, Ngiam Ji-quan, Chen Zheng-hao, et al. Tiled convolutional neural networks [C]// NIPS. 2010

[17] Zhu Zhen-yao, Luo Ping, Wang Xiao-gang, et al. Deep Learning Identity-Preserving Face Space [C]// ICCV 2013. 2013: 113-120

[18] Kan Mei-na, Shan Shi-guang, Chang Hong, et al. Stacked Progressive Auto-Encoders for Face Recognition Across Poses[C]// CVPR 2014. 2014: 1883-1890

[19] 胡石，李光辉，卢文伟，等. 基于神经网络的无线传感器网络异常数据检测方法 [J]. 计算机科学，2014，41(11A)：208-211
Hu Shi, Li Guang-hui, Lu Wen-wei et al. Outlier Detection Methods based on Neural Network in Wireless Sensor Networks [J]. Computer Science, 2014, 41(11A): 208-211

[20] 刘逻，郭力，红肖辉，等. 基于参数动态调整的动态模糊神经网络的软件可靠性增长模型 [J]. 计算机科学，2013，40(2)：186-190
Liu Lu, Guo Li, Hong Xiao-hui, et al. Software reliability growth model based on dynamic fuzzy neural network with parameters dynamic adjustment[J]. Computer Science, 2013, 40(2): 186-190

(上接第 60 页)

Tan Lu. The theory and application of the dimension reduction on the high-dimensional data set [D]. Changsha: National university of defense technology, 2005

[2] 蒋宗礼. 人工神经网络导论[M]. 北京：高等教育出版社，2001
Jiang Zong-li. Introduction to artificial neural networks [M]. Bei Jing: Higher Education Press, 2001

[3] Hinton G, Salakhutdinov R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507

[4] Le Q V, Ranzato M A, Monga R. Building high-level scale unsupervised learning [C]// International Conference on Acoustics, Speech and Signal Processing. Vancouver, Canada, 2013: 8595-8598

[5] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks [C]// Advances in Neural Information Processing Systems. California, America, 2012: 1106-1114

[6] Zeiler M D, Fergus R. Visualizing and Understanding Convolutional Neural Networks [J]. arXiv: 1311.2901, 2013

[7] 朱明，武妍. 基于深度网络的图像处理研究[J]. 电子技术与软件工程，2014(5)：101-102
Zhu Ming, Wu Yan. Image processing research based on deep networks [J]. Electronic Technology and Software Engineering, 2014(5): 101-102

[8] Bengio Y. Learning deep architectures for AI [J]. Foundations and Trends® in Machine Learning, 2009, 2(1): 1-12