文章编号：2095-2163(2021)01-0123-03 中图分类号：TP309 文献标志码：A

融入唇语识别技术提升人脸识别安全性的研究

钟逸晟，尹 芳，李嘉乾，李传师，侯耀辉
（哈尔滨理工大学 计算机科学与技术学院，哈尔滨 150080）

摘 要： 为了解决人脸识别的安全性问题，提高对恶意攻击人脸识别系统的安全防护，使人脸识别技术能够获得更广泛应用，本文提出了在人脸识别技术上融入一种基于深度神经网络的唇语识别技术的系统。与现有的唇语识别技术不同的是，该系统主要是识别用户的唇动习惯。运用本系统，用户在进行人脸识别的同时可按照检测方的提示，读出相应的内容，并在对用户的人脸进行验证的过程中，对用户通过唇动说出的内容分别实现唇动识别、比对，从而有效地提升人脸识别的安全性水平。实验结果表明，在故意针对人脸识别系统的攻击中，融入本技术的系统有更好的识别准确率。

关键词： 唇动识别；人脸识别安全；深度神经网络；身份认证

Research on integrating lip language recognition technology to improve the security of face recognition

ZHONG Yisheng, YIN Fang, LI Jiaqian, LI Chuanshi, HOU Yaohui
(School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China)

【Abstract】 In order to solve the problem of face recognition security, improve the protection against malicious attacks on face recognition system, and make face recognition technology more widely used, this paper proposes a facial recognition technology based on deep neural network lip recognition system. Different from the existing lip recognition technology, the purpose of lip recognition is to identify the user´s lip habits. The system requires users to read out the corresponding contents according to the prompts of the detection party while carrying out face recognition. While verifying the user´s face, the system will carry out lip recognition and comparison on the contents spoken by the user through lip movement, so as to effectively improve the security level of face recognition. The experimental results show that the system integrated with this technology has better recognition accuracy in the face recognition system attack.

【Key words】 lip movement recognition; face recognition security; Deep Neural Network; the identity authentication

0 引 言

研究可知，未来社会的全面无卡化是必然趋势，由此也可以预知，得益于区块链技术和 5G 时代的到来，关于会员卡、银行卡、护照等卡片证件，都终将成为历史，如此一来，安全、可靠的身份识别方式也将随即成为技术演变的潮流与热点。与其他身份识别的研究相比，人脸识别具有方便快捷、专属性较高，对用户友好等优点，因此人脸识别现已成为最活跃的研究领域之一，同时也已成为未来身份识别方式的首要选择。

值得注意的是，随着人脸识别的大范围使用，人脸识别安全性问题已经不容忽视。目前，不法分子即已开始针对人脸识别技术的安全漏洞，利用照片、视频、三维模型等技术，攻击人脸识别系统。

本次课题即旨在研究解决人脸识别的安全性问题。研究中，受到文献[1]的启发，在人脸识别技术中融入了唇语识别机制，且在对唇语识别技术概念加以改进的基础上，除了能对用户人脸进行验证外，还能对用户的唇型和通过唇语说出的内容分别进行比对、识别，去验证活体，从而最终能够有效提升人脸识别的安全性。因此，本文研发设计了基于中文词级别的唇语识别系统。该设计过程包括了：人脸关键点提取、深度神经网络的搭建、训练和测试等一系列技术内容的系统研究，具体如图 1 所示。

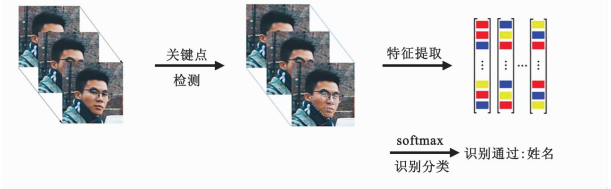

图 1 系统识别的流程

Fig. 1 The flow of system identification

基金项目： 黑龙江省大学生创新创业训练计划项目(201910214112)。

作者简介： 钟逸晟(1999-)，男，本科生，主要研究方向：机器学习、深度学习、区块链技术；尹 芳(1978-)，女，博士，副教授，主要研究方向：机器学习、图像处理。

收稿日期： 2020-11-04

1 人脸关键点提取研究

1.1 建立数据库

作为一项尚未成熟的技术,唇语识别在很多方面都还未见到统一标准。众所周知,语料库的建设即仍亟待完善。国外关于唇语的研究略早于中国,但是国外的语料库都未涉及汉语,因而并不适合本次项目的开发研究。目前,国内已有部分高校和科研机构陆续开启了唇语识别的科研工作,但是相关权威机构却还未能配发有针对性的规范和意见。再者,上述研究主体大多并未将各自使用的语料库予以公开,即使公开的部分也仅限于单个字或者数字的唇语数据集,迄今还未见到句子级别的唇语识别数据库。综上所述,本次项目建立了一个拥有 3 000 个样本的中小型汉语数据库供项目在验证时使用、录制训练样本以及测试样本。

1.2 人脸关键点检测与跟踪

唇语识别的第一步是获得人脸关键特征点在序列图像中的精确定位,这样就有利于后续精准分割出唇部的局部图像。也就是,研究时是源于视频中的每一帧图像,致力于提取出嘴唇局部区域,本次研究即采用了主动外观模型[2]来提取大幅图像中的上述区域,用关键点来定位唇部。

1.3 特征提取

特征提取是一种降维方法,在项目研发中起着重要作用。好的特征可以让识别事半功倍,其具备的共性是:用更少的数据来区分不同的类别,即类内一致性和类间区分性,这样就可使识别任务更加快速且泛化。

目前,学界已经推出了多种提取唇语视频视觉特征的方法,但这些视觉特征提取的方法都不是通用的,究其原因就在于视频视觉信息的多样性,所以传统的唇语视频的特征提取就表现出一定的局限性。针对该问题,本次研究中拟通过神经网络进行特征提取,这样就不仅能够满足研究中对唇语特征所要求的区分性质,同时还可满足对训练性能的要求。

2 深度神经网络构建研究

近年来,深度学习的热度不断攀升,在各种应用中都能看到其身影。在计算机视觉领域的很多任务上,深度学习都取得了良好的应用效果。在深度学习模型中,相比于其他神经网络,AlexNet[3]是经典的卷积神经网络模型,AlexNet 的网络结构在整体上与 LeNet[4]相似,都是先做卷积操作、再进行全连接层。但两者在细节上有很大不同,AlexNet 模型更为复杂。AlexNet 有 5 层卷积,3 层全连接网络,最终的输出层是 1 000 通道的 softmax,还用到 2 块 GPU 进行计算,大大提高了运算效率,并且更适用于视频序列学习任务,近年来,在人脸识别[5]等领域都取得了可观进展。

本项目在识别任务中采用的是 AlexNet 网络模型,这是基于 LeNet-5 网络模型[6]的。分析可知,此种卷积神经网络的特点为:该网络不需要预先获取输入和输出之间的准确映射关系,只需要利用已知模型对神经网络进行训练,就能够学习出相关映射的一种多层的非线性关系,这正是 AlexNet 的独特优势所在,也是其他网络难以比拟的。

AlexNet 网络共有 8 层,主要分为 5 层卷积和 3 层全连接层,如图 2 所示。为了有针对性地强化深层特征的提取,使 AlexNet 网络的信息提取效果更趋完善,AlexNet 网络的末端三层的输出特征将一并输入到最后一层全连接层,这样将有利于在浅层的特征输出,同时也减少了网络在卷积及池化过程中的种种问题困扰,如降维导致的信息丢失等。

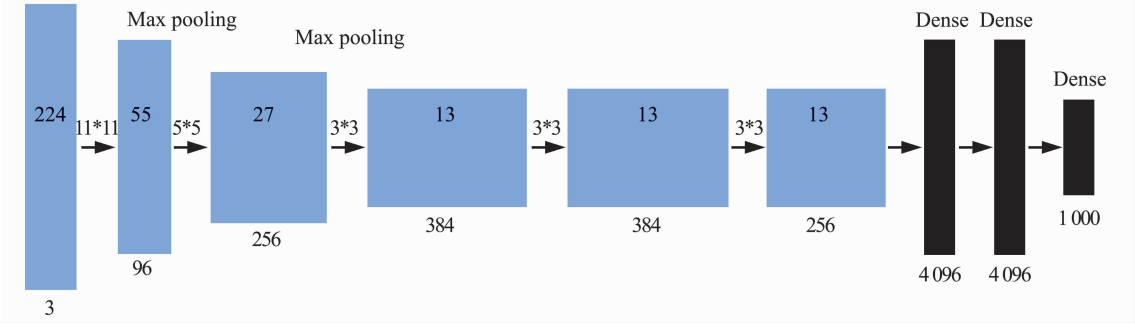

图 2 AlexNet 网络结构

Fig. 2 AlexNet network structure

3　分类

在接下来的分类研究中，文中采取了最广泛应用的 Softmax 分类器，同时为了提升 AlexNet 网络结构的识别准确率，研究中在 Softmax 分类器后附加了中心损失函数。由中心损失（Center Loss）函数配合 Softmax 损失（Softmax Loss）函数来进行分类和识别。进一步来说，本项目中人脸识别技术的主要步骤为：输入人脸视频，将视频中的每一帧进行图片预处理后，输入到 AlexNet 网络中，再将提取到的特征传入含有全连接层的 Softmax 分类器中，对人脸视频进行分类，由此在唇语人脸数据库中实现图像序列的识别。

4　实验

综上研究所述，针对本系统的应用场景，由于没有类似的可对照识别率数据，本文采取了较为理想的方式进行实验。实验中，选择 2 个对象 A、B，B 想要伪造 A 的身份，由此设置了多组多次的对照实验，即：A 本人，B 戴本人照片面具、戴 A 照片的面具但 B 将嘴巴漏出、B。研究按照以上 4 种情况分别进行识别测试，最终得到了理想情况下的实验识别结果，详见表 1。

表 1　理想情况实验识别率

Tab. 1　Recognition rates for ideal case experiments　%

	A 本人	B 戴 A 照片面具	戴 A 照片的面具但 B 将嘴巴漏出	B
正确读出指令	98.7	3.6	5.1	1.7
非正确读出指令	37.5	3.4	3.7	2.1
静止	28.2	2.5	2.2	1.8

分析表 1 结果可知，当 A 本人进行正确操作时，系统的识别率非常高；而当不是本人的脸，并且也不是按照本人的唇动习惯说话时，则有 97% 以上的几率无法识别通过，这就清晰表明了本系统有着良好的鲁棒性和安全性，也标志着该种人脸识别方法的研发获得了成功。

5　结束语

本项目的研究旨在要求唇语识别部分能够识别到每一个人的唇语。考虑到不同人的唇动方式也是不同的，因此，通过唇动序列的比对就可以辨别出是否为待测者本人的嘴唇，这就有效解决了在照片上扣洞或戴上人皮面具来读取内容进行识别的攻击手段带来的弊端，而且也可以有效辨别出正在进行比对的是否为双胞胎兄弟。在唇语识别研究中，特征提取方式采用的是卷积神经网络，同时还结合了长短时记忆网络（LSTM），分析视频并对视频数据中的时间以及语义信息进行挖掘，这也是该项目研发的创新点之一。

若成功地结合了唇语和人脸识别技术，即将唇语识别用到的特征和人脸识别特征相结合，就可以得到：通过唇语和人脸识别的综合比对，最终可证得在摄像头下是实时、并且也是本人的嘴唇。首先，人脸识别排除了通过照片或者是视频回放的攻击手段，然后又排除了立体模型和化妆、用双胞胎代替识别的攻击手段。在此基础上，由实验结果分析得出的结论就是：在摄像头前的就是本人。融合 2 种识别技术来提高人脸识别的安全性，这也体现了本项目 1+1>2 的研发思路。研究中，只是使用普通摄像头、普通的手机或者电脑，并不需要另行添加其他任何辅助设备，就能够达到提高人脸识别技术安全性的目标。综上所述，这些优点使得融入了唇语识别的人脸识别系统的成本较为低廉、易于实施。

进一步分析可知，提升了安全性的人脸识别系统，在使人们享受到人脸识别带来便利的同时，也使其信息、财产等方面获得了更为强大的安全保障。不仅如此，这种人脸识别方式还可应用在更广阔领域中，例如：远程身份认证、刷脸门禁考勤、人脸支付、人脸登录等场合。故而，本次项目研发成果对于当前社会的快速发展有着重要的现实意义。

参考文献

[1] 任玉强. 高安全性人脸识别身份认证系统中的唇语识别算法研究[D]. 重庆：中国科学院重庆绿色智能技术研究院，2016.

[2] 蔡凡. 基于主动外观模型的图像分割研究[J]. 闽江学院学报，2014，35(2)：80-87.

[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] LÉCUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11)：2278-2324.

[5] ALMABDY S, ELREFAEI L. Deep Convolutional Neural Network-based approaches for face recognition [J]. Applied Sciences，2019，9(20)：4397.

[6] MAATTA J, HADID A, PIETIKAINEN M. Face spoofing detection from single images using microtexture analysis [C]// Proceedings of the 2011 International Joint Conference on Biometrics. Washington, DC, USA：IEEE，2011：10-17.

[7] 李丹，沈夏炯，张海香，等. 基于 Lenet-5 的卷积神经网络改进算法[J]. 计算机时代，2016(8)：4-6，12.