

学校代码 10125

专业代码

山西财经大学

Shan xi University of Finance and Economics

# 本科毕业论文（设计）

论文题目：基于唇语识别的身份认证系统的设计  
与实现

姓 名：卫 鑫

学 号：201608010216

班 级：16 级信息管理与信息系统 2 班

专 业：信息管理与信息系统

学 院：信息学院

指导教师：郑芳讲师

完成时间：2020 年 5 月 5 日

# 说 明

一、毕业论文（设计）须用 70 克 A4 纸计算机双面打印，具体打印装订格式参见《本科毕业论文（设计）内容与格式规范》。

二、封面中“专业代码”应填写教育部统一的普通本科专业代码（具体参见教务处主页“下载专区”《山西财经大学现设专业一览表》）。

三、毕业论文（设计）装订顺序

- 1、封面
- 2、扉页
- 3、学术承诺、使用授权的说明
- 4、中文摘要、关键词
- 5、外（英）文摘要、关键词
- 6、目录
- 7、正文
- 8、参考文献
- 9、附录
- 10、致谢
- 11、本科毕业论文（设计）指导教师评定表
- 12、本科毕业论文（设计）答辩成绩与总成绩评定表

学校代码 10125

专业代码

山西财经大学

Shan xi University of Finance and Economics

# 本科毕业论文（设计）

中文题目	基于唇语识别的身份认证系统的 设计与实现
英文题目	<i>Design and Implementation of Identity Authentication System Based on Lip Recognition</i>
姓 名	卫鑫
学 号	201608010216
班 级	16 级信息管理与信息系统 2 班
专 业	信息管理与信息系统
学 院	信息学院
指导教师	郑芳 讲师
完成时间	2020 年 5 月 5 日

## 毕业论文（设计）学术承诺

本人郑重承诺：所呈交的毕业论文是我个人在导师指导下进行的研究工作及取得的研究成果。除了文中特别加以标注和致谢的地方外，论文中不存在抄袭情况，论文中不包含其他人已经发表的研究成果，也不包含他人或其他教学机构取得研究成果。

作者签名：\_\_\_\_\_ 日 期：\_\_\_\_\_

## 毕业论文（设计）使用授权的说明

本人了解并遵守山西财经大学有关保留、使用毕业论文的规定。即：学校有权保留、向国家有关部门送交毕业论文的复印件，允许论文被查阅和借阅；学校可以公布论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存论文。（保密的论文在解密后应遵守此规定）

作者签名：\_\_\_\_\_ 指导教师签名：\_\_\_\_\_

日 期：\_\_\_\_\_ 日 期：\_\_\_\_\_

## 摘 要

随着科技时代的到来，个人身份的虚拟化以及人们对于个人身份安全的需求越来越高。怎么确认登陆系统的人员身份，且不给登陆的用户带来太过繁琐的流程，是现代系统需要考虑的核心问题。面部识别技术由于具有唯一性且方便快捷，使得其在金融、通行、网络安全等领域，得到了飞速的技术进步。但是，当这种识别技术得到普及的时候，一些图谋不轨的不法分子也逐渐开始寻找和利用面部识别的系统漏洞，来破解登陆系统，这使得信息认证受到了挑战。而系统的开发这只能通过登陆系统的时候，对用户提出随机动作，例如点头和眨眼等，以此来确认登陆系统的不是图像和机器，但是这样的作法，不能很好的解决系统的缺陷。

因而人们开始将研究重心进行了转移，在研究如何检测真人的方法中，加入唇语识别技术。每个人的嘴唇形状和纹理都是不一样的，哪怕有人的嘴唇形状相似，但是说话方式是也肯定不一样。由于唇语识别在身份认证上有着得天独厚的优势，所以我们接下来讨论唇语识别的原理和使用到的技术。本文主要先将唇部区域进行分割，然后用 AdaBoost 训练器训练，最后通过卷积神经网络的 Tensorflow 框架进行处理。

**关键词：**唇语识别；卷积神经网络；身份认证

## Abstract

Because of the development of the technological age, the virtualization of personal identities and people's demand for personal identity security are increasing. How to confirm the identity of the person who logs in the system, and does not bring too complicated processes to the logged-in user, is the core issue that modern systems need to consider. Facial recognition technology is unique, convenient and fast, which has made rapid technological progress in the fields of finance, traffic and network security. However, when this kind of recognition technology became popular, some criminals with bad intentions gradually began to find and use the system vulnerability of facial recognition to crack the login system, which made the information authentication challenged. The development of the system can only be done by presenting random actions to the user when logging into the system, such as nodding and blinking, etc., to confirm that it is not the image and the machine that log into the system, but such a method cannot solve the system well. defect.

Therefore, people began to shift the focus of the research, adding lip recognition technology to the research on how to detect real people. Everyone's lips are different in shape and texture. Even if someone's lips are similar in shape, the way they speak is definitely different. Because lip recognition has a unique advantage in identity authentication, we next discuss the principles and techniques of lip recognition. This article mainly divides the lip region first, then trains with the AdaBoost trainer, and finally processes it through the Tensorflow framework of the convolutional neural network.

**Key words:** Lip recognition; Convolutional Neural Network; Authentication

# 目 录

1 导 论 .....	2
1.1 选题背景与意义 .....	2
1.2 国内外文献综述 .....	3
1.3 论文的结构及主要内容 .....	3
1.4 论文的研究方法 .....	4
2 嘴唇的分割和特征值的提取 .....	4
2.1 嘴唇的分割 .....	4
2.1.1 嘴唇区域的研究 .....	4
2.1.2 嘴唇的分割 .....	5
2.2 特征值的提取 .....	6
3 基于卷积神经网络的唇语识别处理 .....	7
3.1 卷积神经网络 .....	7
3.1.1 卷积神经网络的优点 .....	7
3.1.2 卷积神经网络的原理 .....	7
3.1.3 AlexNet 的基本结构 .....	8
3.2 对数据集进行的训练和测试 .....	9
3.2.1 处理过程 .....	9
3.3 唇语认证的结果和分析 .....	10
3.3.1 实验的工具和配置 .....	10
3.3.2 实验平台的配置 .....	11
3.3.3 实验的结果和分析 .....	11
4 总结和展望 .....	13
4.1 工作总结 .....	13
4.2 未来展望 .....	13
参考文献 .....	15
致 谢 .....	16

# 1 导 论

## 1.1 选题背景与意义

二十世纪以来，计算机和互联网在不断进步和普及，变得越来越贴近我们的生活，人们对于个人的信息安全有了较大的重视，而个人身份认证也有着各种各样的方式，例如密码认证，口令认证，面部识别，虹膜识别等。传统的密码认证和口令认证有着较大的风险，而虹膜识别设备要花费的金钱，对于普通用户来说价格不菲，所以面部识别是一个比较折中的方式，既保证了安全性也有较低的成本。

但是面部识别可能造成认证身份不够安全的问题，而一些不法分子也利用这种方式来迷惑身份认证系统。在现实生活中，面部识别所遇到的主要攻击有下面几种：1.采用照片攻击的手段，例如有些不法分子会通过照片，用其攻击识别系统的漏洞，达到混淆和欺骗系统识别的目的。2. 采用视频攻击的手段，例如用其他不当手段获得的面部视频来混淆和欺骗识别系统，且这种方法更容易混淆和瞒过认证系统，所以很难防备[4]。3.采用 3D 攻击手段攻击，这种方法的攻击性更强，一般系统更加难以防备。

尽管攻击手段有着很多种，但是目前大多数的系统都是按照上面三种攻击方法，来进行针对性的防备的。目前大部分的系统，基本采用的是活体检测方法。活体检测，简单来说就是系统判断登陆系统的是否是活的生物，也就是真人。

根据人们对于生物识别技术的探索，我们发现，每个人的嘴唇的运动方式都是不一样的，每个人的嘴唇部位肌肉的发达强度和运动的力度以及方式都有着细微的差别。同一个词，不同的人来读的话，都是不一样的。而且使用唇语而不是不包括语音的好处是，在其他背景声音太大的情况下例如机场，工地等场所，该认证系统也能很快的、不受干扰的进行验证。由于每个人的说话方式不同，所以就算被别人知道密码，也无法通过验证系统。

虽然随着技术的进步，如今社会上已经涌现出了很多非常优秀的人脸识别算法，比如 DeepID3[15]，VGG[16]和 GoogLeNet[17]算法，但是随着对生活隐私和信息安全的要求越来越高、越来越强烈，所以衍生出了一种基于唇语识别的技术，这样使得系统在判定是否是真人的同时，也能够比较快速方便地构建安全系数更高的身份识别系统。因此这里主要探讨的是唇语识别中的提取的方法主要有下面几种：

- 1.对唇部区域进行分割，提取唇语特征，然后通过训练得到相关的分类模型，然后系统可以用训练建立的分类模型对于输入识别的数据进行分析，然后得出结果。

- 2.使用 Fourier 模型建立唇形分类方法[6]，实现唇形分类系统。首先要做的就是处理原始图像，要先将输入的图像进行初步的识别或者处理，由于光线和亮度，或者遮挡物会造成误判的情况，所以我们需要先对图像中的其他干扰因素进行处理。

- 3.提取用户的唇部特征。唇语识别中两个最重要的特征就是几何形状和灰度级。这里我们会着重讲解以下如何提取唇部的形状特征。



## 1.2 国内外文献综述

如今人们关于唇语识别，主要的研究方向有下面两个，其一是，在唇语识别方面，基于句子以及语言数据库实现对于唇语的识别；其二是，将唇语作为面部识别中的活体检测的手段，以达到安全的身份认证。

唇语视频我们可以看成很多个图片帧，而且我们只需要关注说话人唇部肌肉的运动特征就可以了。但是由于视频很容易收到外界条件的影响，例如光线和唇色的就会有有很大的影响。这样就会导致说话者说两个一样的词，但是得到的结果差别很大。所以我们需要提取出说话者的关键特征信息，因为不管环境怎么改变，这些比较重要的核心特征信息，是不会随着环境的不同而发生较大的变化，所以我们可以从嘴唇的形状和运动方向等信息来进行研究。

唇语检测最重要的就是在选择特征信息上，以及特征信息的提取，用不同的方法，对唇部区域的提取和分割也会不同，导致结果的不同，而在特征信息的提取，大部分使用的是提取形状特征，包括嘴唇的形状、大小和丰满程度的相关特征信息，除了上述方法外，还有一种是基于纹理特征的提取[10]。

因为人们研究水平的提高和对唇部区域研究的不断深入，进而提出了其他更好的算法，主要是为了对嘴唇区域进行切割，主要是原来的算法已经不能够满足人们的需要了，改善后的算法将形状信息和颜色信息结合在了一起，该函数可以增强嘴唇区域与其他区域之间的差异[13]。

然后再使用关键点对于嘴唇进行分割，这些关键点要能够很好的描述唇部的信息，做到保留核心的特征信息，用这些关键点来拟合出嘴唇信息，且忽略那些影响不是很大的特征信息，在保证系统准确性的同时做到减少系统的一些复杂计算。

因为每个人的嘴唇的大小和形状都不一样，也使得系统不能很好的定位嘴唇位置，传统的措施主要是，系统先辨识出人脸，然后再通过已经识别的人脸来得到眼间距，最后找出嘴唇位置，但是随着这方面的需求不断地增长，我们可以通过 `Opencv` 库来对唇部进行实时的追踪。唇部的特征提取主要指的是提取下面两个两方面的信息，形状特征，也就是从唇部区域范围中提取这个目前使用范围较为广泛；后来人们发现不同人的唇部的灰度值也是不同的，所以灰度特征也成为唇部特征的一个主要信息。

嘴唇经过一系列的处理后，可以切割得到一些曲线，然后通过规范化，处理这些得到的曲线。经过特征提取之后，得到大量的特征参数，然后我们可使用不同的分类器来处理和分析数据，拟合出来函数的基本形状。在分类器的选择上，可以采用 `HMM` 模型来进行研究。`HMM` 包含两部分，第一部分，也是最重要的部分就是训练，另一部分是测试，这两部分分别对应不同的算法——`Baum-Welch` 算法和 `Viterbi` 算法[2]。

## 1.3 论文的结构及主要内容

本文主要是使用唇语识别，研究如何安全、方便和有效地进行身份认证，主要结构如下：

第一章为导论，主要是阐述研究目标和研究流程，说明唇语识别的发展前景，对于一些国内外的研究现状做一些基本描述，以及现在研究唇语识别的必要性。

第二章为对于嘴唇的分割和特征值的提取做大概描述，使用相关模型对于收集到的数据集做相关处理，主要是将视频以每一帧分散储存为图片，然后再针对图片进行处理。

第三章为对于得到的数据先进行预处理，需要用到 Dlib 库等，然后运用卷积神经网络来对模型进行训练，处理输入的数据，然后使用 Alexnet 算法进行研究，最后得到相关的数据，并且进行分析。

第四章为总结和研究思想，主要总结这次的研究工作得到的灵感和思考。

## 1.4 论文的研究方法

因为现实生活中，活跃着的人脸识别算法比较多，而且也很好用与方便，比如 DeepID3, VGG 和 GoogLeNet 等算法[5]，所以在这里主要，本文主要针对探讨的是，唇语识别技术以及相关的识别算法。有以下主要方法：

- 1.提取唇语特征，然后获得相应的数据集对系统模型进行训练，得到基本的训练结果[9]。

- 2.嘴唇的分割，实现唇形分类系统。从视频中得到唇部特征的时候，对唇部的形状进行描述，提取唇部特征值，然后描绘出唇部曲线，这样就完成了预处理。

- 3.使用卷积神经网络，在训练好了系统以后，将视频输入进去，然后系统会对输入的内容进行分析，使用 K 均值分布的方法，与训练后的数据进行匹配，最终得到结果。

## 2 嘴唇的分割和特征值的提取

### 2.1 嘴唇的分割

#### 2.1.1 嘴唇区域的研究

嘴唇的定位是唇语识别的前提，早期的嘴唇定位是直接拍摄嘴部视频，或者由人工后期标识，而且由于光线和、肤色等其他条件影响，使得嘴唇的定位变得困难。随着各种检测方法的发展，现在主流方法是用 OpenCV 这个计算机视觉库进行的实时唇部定位与跟踪算法[8]。OpenCV 是现在比较流行的项目，其主要包括如光线追踪和 3D 显示。

而嘴唇的特征主要分为形状特征、像素特征，随后由于人们发现可以同时提取两者的特征，就衍生出了混合特征，具体的对比如图 2-1。其中形状特征是最简单的特征，主要是用来建立嘴唇形状模型，用来表达嘴唇的几何形状参数，其优点是特征值数量少和便于统计与计算。但是其缺点也是十分的明显，就是在光线不充足的情况下，视频或者图像就会成像不好，而且其产生的噪点会严重影响系统的判断，会使得模型的识别率低；而像素特征是获得嘴部区域的像素特征信息，运算的效率和识别率要高于基于形状特征，但是其数据特征值的维度较高，计算代价大，需要专门的降维处理。而混合模式则就是图像特征和像素特征的混合，取了各自的优点，在降低特征值的维度时，提高了辨识度，但是其算法的复杂程度也是最高的。

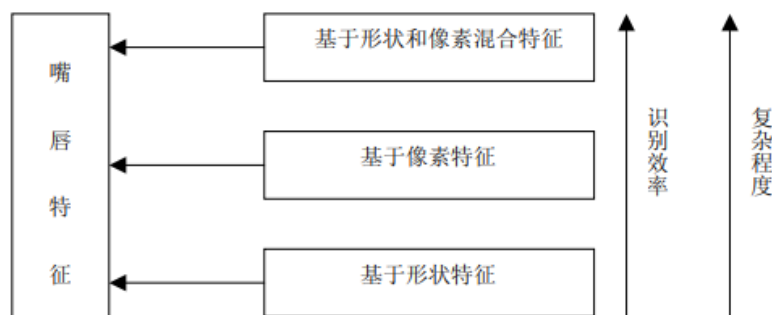


图 2-1

### 2.1.2 嘴唇的分割

在检测到人脸后，就可以很方便的对于嘴唇进行定位了。这样使得定位嘴唇位置更加准确和快捷，减少了需要处理的图像区域，减少了算法的复杂度。

使用 AdaBoost 级联分类器进行训练，AdaBoost 算法的一个非常大的优点就是在于：当每一个基本分类器训练完成后的时候，不是直接结束训练，而是会用训练完的数据，继续训练下一个分类器，而且会在每一个分类训练器训练完成后，添加一个新的分类器 [11]，这样的好处是可以不断的完成迭代而且不用担心错差太大，而且使得系统的准确性得到了一部分提高。

在 AdaBoost 算法没有出现前，当时主流的是 Boosting 算法，其基本流程如图 2-2：

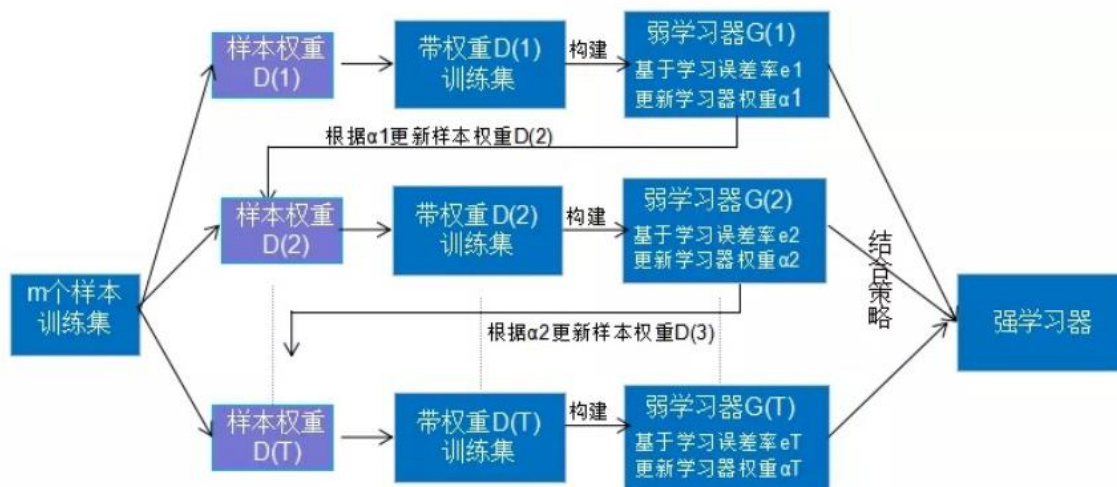


图 2-2

但是在这样的过程下，很容易出现下面的问题：集成效果差和分类效果不好。为了解决这两个问题，故而研究出来了 Adaboost 算法。其解决了数据处理的过程中的过度拟合问题，使得数据得到更好的训练和处理。

而算法强化的过程如图 2-3 所示，就是逐步改变样本的权重，也会在每个弱分类器进行分类完成后，不是直接结束，而是使用上一个分类完成后得到的数据，继续进行分类，进而得到结果，这样防止了过度拟合，也使得结果会更加准确。

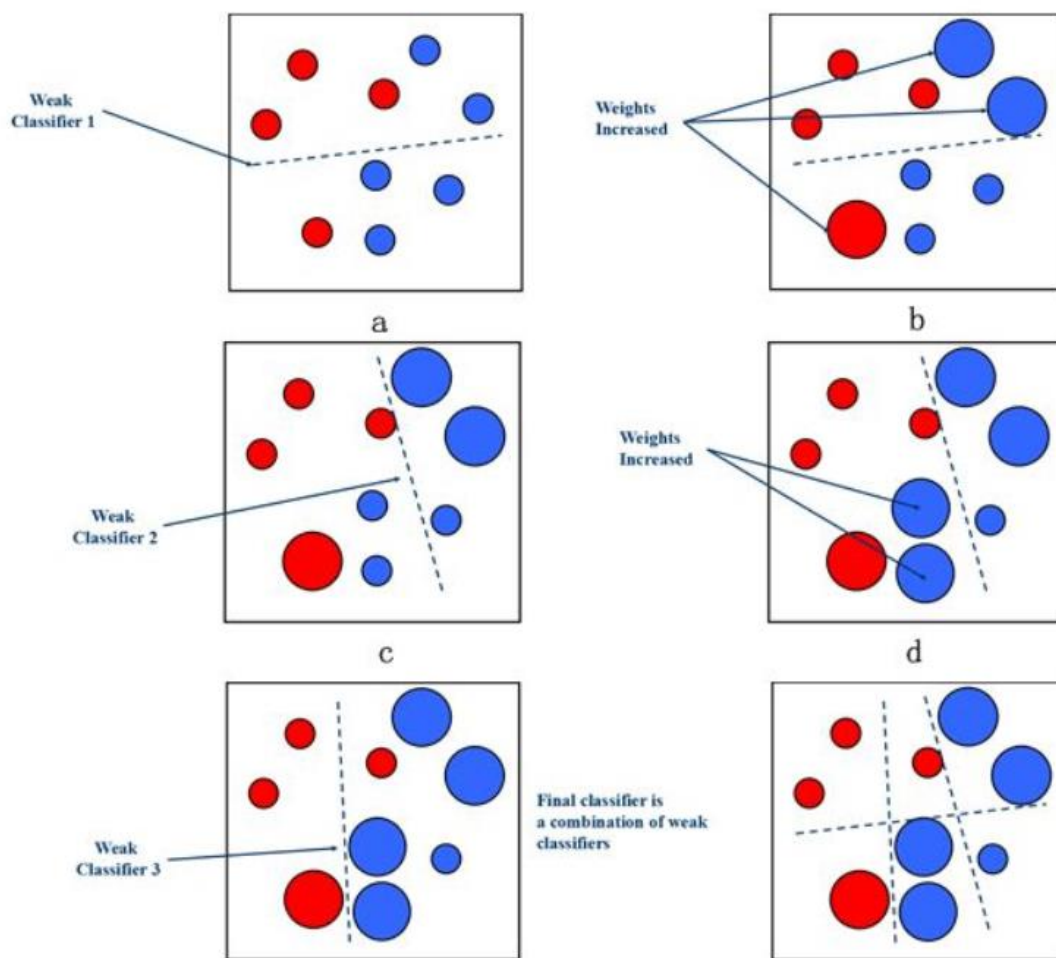


图 2-3

## 2.2 特征值的提取

对于建立嘴唇轮廓模型中，我们还需要在嘴唇轮廓上选取必要的一些用来描述嘴唇信息的关键点，这些关键点要能够描述完整的嘴唇上下曲线，还要再唇部轮廓变化的时候完整表达轮廓曲线[14]。如果选取的关键点过少，则不能很好的反应嘴唇的轮廓，如果选取的关键点过多，则会造成算法的复杂度上升。通过实验，我们能发现，在选取 14 个关键点[14]的时候，能够较好的描述嘴唇轮廓，其中的选取的关键点主要分布如图 2-4 所示。

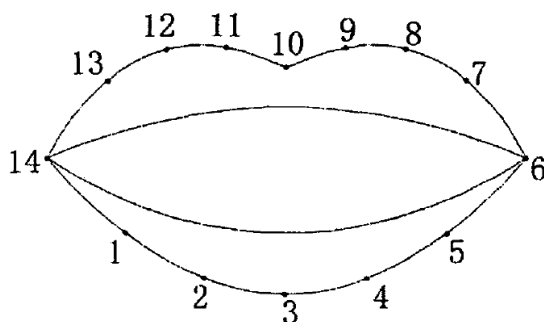


图 2-4

在获取可以描述嘴唇形状的 14 个关键点后，再对关键点的模型拟合，再用类似抛物线的三条曲线对嘴唇的形状进行拟合。分别是上部左轮廓、上部右轮廓和下部轮廓[1]。然后算出曲线所围面积。并且从日常生活中的发声我们能知道，一个完整的发音包括多个口型，所以我们还需要对于得到的特征信息进行降维处理，这里的降维处理不是本文论述的重点，所以不多做篇幅描述。

## 3 卷积神经网络在唇语识别的作用

### 3.1 卷积神经网络

#### 3.1.1 卷积神经网络的优点

在卷积神经网络没有出现的时候，用人工智能来处理图像是个很大的问题，主要有以下的问题：首先，由于处理图像非常占用计算机的算力，即占用计算机的资源；其次就是在处理图像的过程中，对于图像进行压缩和转化的过程中，就会使得图像的原有特征很难在处理过程中完整的保存下来，导致处理图像的误差较大。

做所周知，计算机中保存的图像，主要是由大量的像素组成的，而像素又可以区分为红、绿、蓝三种 RGB 颜色，例如一张 2000\*2000 像素的图片，在计算机处理过程中，就需要一两千万个参数来描述这个图像。这样的话就非常的消耗计算机的资源，而随着人们对于数码技术的需求越来越高，照片的像素越来越高，如果直接处理这些图片的话，会对计算机资源有着很大的压力。

针对上述的问题，卷积神经网络，首先能够处理的问题是，可以将复杂的问题简单化，即简化复杂的问题，由于提取特征值的时候会有很多的参数，造成计算的困难，于是卷积神经网络的另一个优点，就是可以在大量的参数中，采用降维的方法，然后得到一些比较重要的核心参数，然后再对这些重要的核心参数进行处理。其最大的优点在于，这种降维虽然减少了特征值的数量，但是只是减少非核心要素的影响，不但没有使结果不准确，反而从某种意义上来说是增加了系统的准确度。卷积神经网络还利用类似于视觉的方式，保留了更多的图像的特征，即使图像的相对位置发生了变化，它能快速准确的识别出来，这样就使得我们不用再去在意图像的一些其他问题，这说明卷积神经网络在保留图像特征的情况下有着得天独厚的优点了。

#### 3.1.2 卷积神经网络的原理

卷积神经网络有很多构成说法，众说纷纭，但其中最核心的和结构最清晰的还是“卷积层，池化层以及全连接层”如此分类的组成标准，将其分为这三类是普遍被人接受的。

首先卷积层的作用，是提取图像核心特征值的，是必不可少的一部分；但是大量的数据使得系统的负载较重，降维是系统需要考虑的问题，为了解决这问题，卷积神经网络引入了池化层的概念。有时当数据不是很大的时候，也可以没有这个部分，但是大部分时候，图像都是比较大的，如果我们保留了所有的参数，并且将所有的这些参数

进行训练，其需要的计算量是一个天文数字，所以减少参数是一个很好的方法。于是池化层的作用就是在卷积层之后，用来帮助系统减少图像的参数，以便于减少每次的计算量。全连接层，又叫做输出层。因为卷积层和池化层在处理过程中，并不能直接输出数据，而是仅仅是提取特征，所以输出层的出现是帮助系统减少需要处理打的图像的相关参数，从而形成最终有效的输出。我们通过全连接层的帮助，为来输出我们训练的到的相关结果。

但是，我们生活中的卷积神经网络，并非仅仅只是上面提到的一种结构，而是有着多种结构，例如：使用多个卷积层和池化层组成的神经网络，如图 3-1 所示。

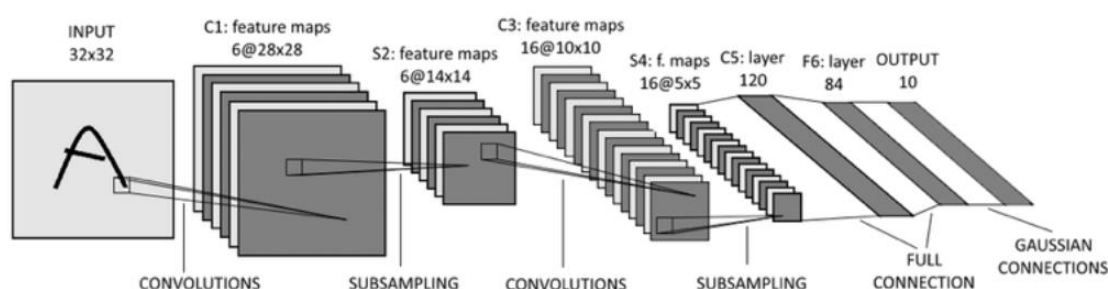


图 3-1

我们在实际应用中，由于数据的复杂性，简单的一层卷积结构不能很好的处理数据，就衍生出了多层卷积结构，这样的多层结构能够更好的处理相对复杂的数据，由于一层卷积学习到的信息有时不能很好的拟合出数据的真实情况，所以我们要使用多层卷积，帮助系统学习全局的信息，而层数越多，学习到的信息就越完整。

在日常生活中，也有很多比较常见的其他卷积神经网络，例如：（1）LeNet-5：这个模型共有 7 层，其设计出来的初始目的就是为了解决数字识别问题。（2）AlexNet：其增加了网络的性能，使得该算法得到广泛的应用，而且掀起来更多人学习研究卷积神经网络的浪潮。（3）ResNet：有共 152 层，在图像分类上，这个神经网络已经做的比人类的误差还小，但是它靠的不是算法的深度而获得如此优秀的运算能力，而是通过残差学习法，用这种方法来训练模型，减少参数量。

### 3.1.3 AlexNet 的基本结构

在结构设计上，Alexnet 为了增加网络的性能，其使用了多种技术，例如：为了扩充数据集而采用数据增强的方法，将 relu 函数放到每一个卷积层的结尾处，这样的话能够解决当网络较深的时候，造成梯度消失的问题。并且其加入了局部相应归一化提高网络泛化能力和采用最大值避免出现图像模糊。在卷积神经网络的唇部图像结构简单和很强的特征提取能力的结合下，使得 Alexnet 对于提取唇部特征有着得天独厚的优点。

AlexNet 的基本模型如图 3-2 所示，以及内核的参数设置如图 3-3 所示，此神经网络有着 6000 万个参数以及 65 万的神经元，其中输出层有 1000 个神经元。AlexNet 网络模型共有八层组成，为了更好的处理数据，其采用了五个卷积层来帮助进行数据的处理，以及三个全连接层用来对于数据进行输出，用三个全连接层主要的目的是，帮助获得每次



训练的数据，否则每次等到数据训练完才发现结果不合适，十分的耗费时间，这个也是此基本模型如此流行的原因之一。

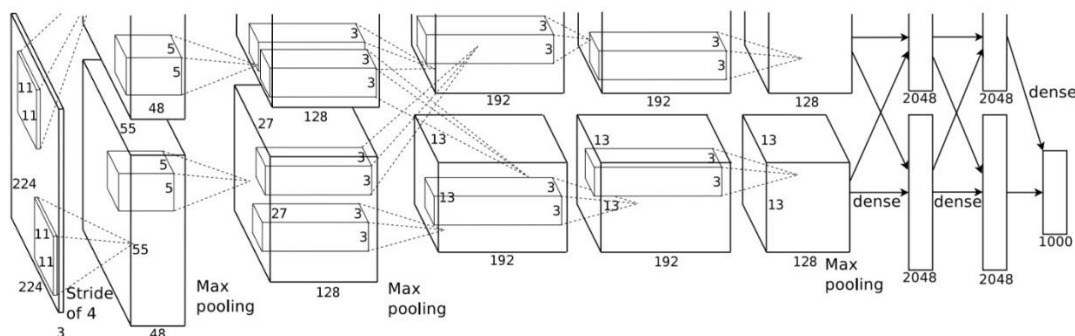


图 3-2

网络层名称	输入大小	卷积核大小	输出大小
<b>Conv1</b>	$224 \times 224$	$11 \times 11 \times 96/4$	$55 \times 55 \times 96$
<b>Pool1</b>	$55 \times 55 \times 96$	$3 \times 3 \times 96/2$	$27 \times 27 \times 96$
<b>Conv2</b>	$27 \times 27 \times 96$	$5 \times 5 \times 256/1$	$27 \times 27 \times 256$
<b>Pool2</b>	$27 \times 27 \times 256$	$3 \times 3 \times 256/2$	$13 \times 13 \times 256$
<b>Conv3</b>	$13 \times 13 \times 256$	$3 \times 3 \times 384/1$	$13 \times 13 \times 384$
<b>Conv4</b>	$13 \times 13 \times 384$	$3 \times 3 \times 384/1$	$13 \times 13 \times 384$
<b>Conv5</b>	$13 \times 13 \times 384$	$3 \times 3 \times 256/1$	$13 \times 13 \times 256$
<b>Pool5</b>	$13 \times 13 \times 256$	$3 \times 3 \times 256/2$	$6 \times 6 \times 256$
<b>Fc6</b>	$6 \times 6 \times 256$	——	4096
<b>Fc7</b>	4096	——	4096
<b>Fc8</b>	4096	——	1000

图 3-3

AlexNet 网络模型有着其他的模型无法比拟的优点：首先使用 ReLU 非线性的激活函数使得梯度下降训练的速度要快很多，这样使得模型的训练可以大幅度降低，而且不需要担心准确度；其次可以在多个 GPU 上面训练，在不同的 GPU 上进行的训练同时进行，减少训练所需要的时间和侧面提升了训练速度；利用多层的卷积层和池化层，防止过度拟合。

## 3.2 对数据集进行的训练和测试

### 3.2.1 处理过程

在对于视频部分的处理中，我们对视频进行处理，使其便为 30f/s 的相等帧速率。然后，我们使用 Dlib 库来进行处理[3]，这个库对于面部跟踪和嘴巴区域的提取有着很好的效果，且不要求其他的条件。然后我们将处理完成后的输入图像，按照分割成固定的大小的部分进行处理，然后将这些相同大小的部分放在一起，最后这些固定大小的数据集合组成输入要素的数据集。对视频进行处理的具体步骤如图 3-4 所示。

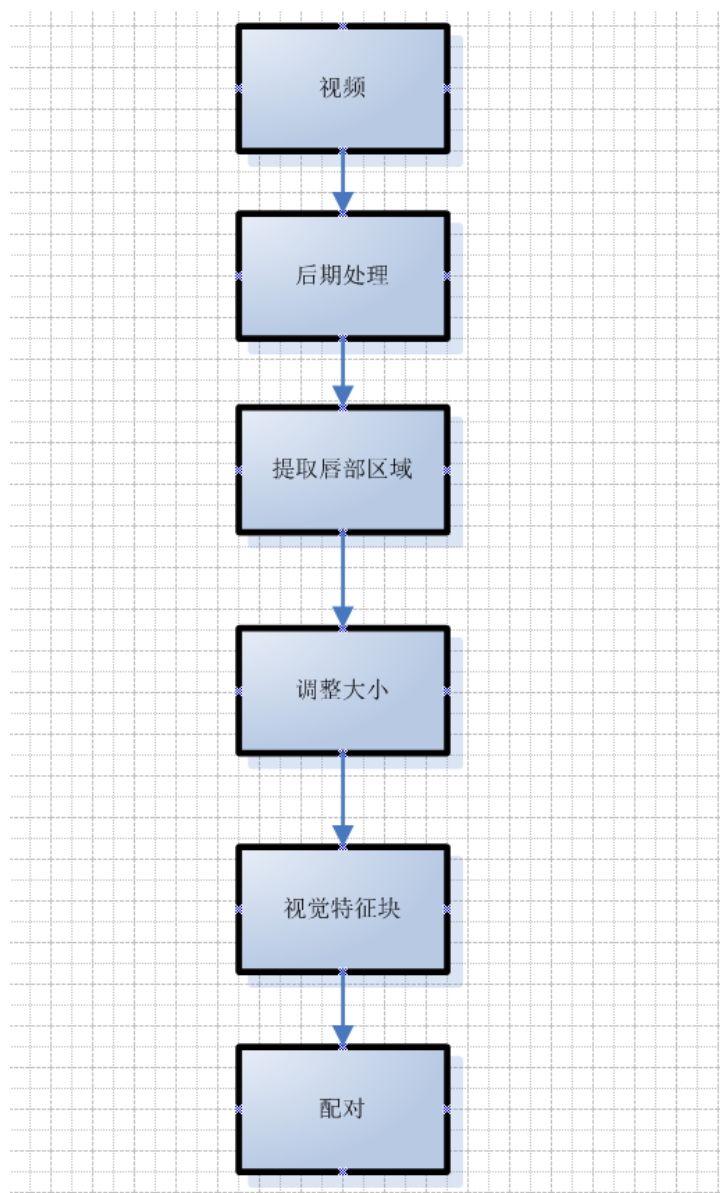


图 3-4

在此过程中视频后期处理，是将每个视频剪辑的帧率进行调整，例如将视频调整为一秒钟 30 张图片，然后在对单个图像进行处理。这样的话就可以对于单个图像提取唇部信息，然后调整合适的大小，最后进入系统，通过和之前训练完成的数据进行对比和配对，就能够获得最终的结果，其中训练的数据集对于结果有着很大的影响，选择合适有效的训练集是重中之重。

### 3.3 唇语认证的结果和分析

#### 3.3.1 实验的工具和配置

由于卷积神经网络优异性能，使得其这几年有着快速的发展，而且各种学习研究的开源化，使得深度学习框架不断得到补充和进步，不同的框架有着各自的优点和缺点，如图 3-5 为了探讨他们的各项相关的信息：



框架	语言	文档资料	CNN 兼容	RNN 兼容	上手难易	速度	并行支持	Keras 兼容	支持团队
Theano	Python/C++	++	++	++	+	++	+	+	蒙特利尔大学
TensorFlow	Python	+++	+++	++	+++	++	++	+	Google
Torch	Lua, Python	+	+++	++	++	+++	++		Facebook
Caffe	C++	+	++		+	+	+		贾扬清 加州伯克利
MXNet	Python, R, Julia	++	++	+	++	++	+++	+	李沐, Amazon
Neon	Python	+	++	+	+	++	+		Intel
CNTK	C++	+	++	+++	+	++	+		Microsoft

图 3-5

结合实验的需求，本次实验采用的框架是 Tensorflow 框架，其主要使用范围是：帮助开发者进行机器学习和神经网络的研究。主要优点有三个：第一，兼容性和可移植性好，既可以部署到普通计算机也可以部署到云端服务器上面，且可以支持多数量的 CPU 和 GPU；第二，Tensorflow 拥有包括卷积神经网络等很多深度学习的算法，拥有全面的深度学习 API；第三，Tensorflow 还有着例如 TensorBoard 等可视化组件，能够对于训练过程进行跟踪，大大增加了实验的直观性[7]。

### 3.3.2 实验平台的配置

实验环境如下：

CPU：

GPU：

内存：12G；

操作系统：Windows 10 64 位

开发环境：Python 3.6；PyCharm 2019.3.3 64 位；OpenCV-python 4.2.0.32；Dlib 19.8.1；

深度学习框架：Tensorflow 1.4.0；

### 3.3.3 实验的结果和分析

本文是对 Code Ocean 的官方项目 lip-reading-deeplearning 和 Github 上面的公开项目 LipNet 在数据集上的准确性进行了验证，根据实验的结果可以得到下图 3-6。在一开始训练的时候，卷积神经网络的准确性，会随着训练的训练次数达到一定程度的时候基本保持不变，这就意味着这个时候的数据能被该模型较为准确的拟合，根据试验的不断尝试，我们发现迭代次数达到 8 次的时候，神经网络的准确性就比较稳定了，而整个神经网络则达到最优。

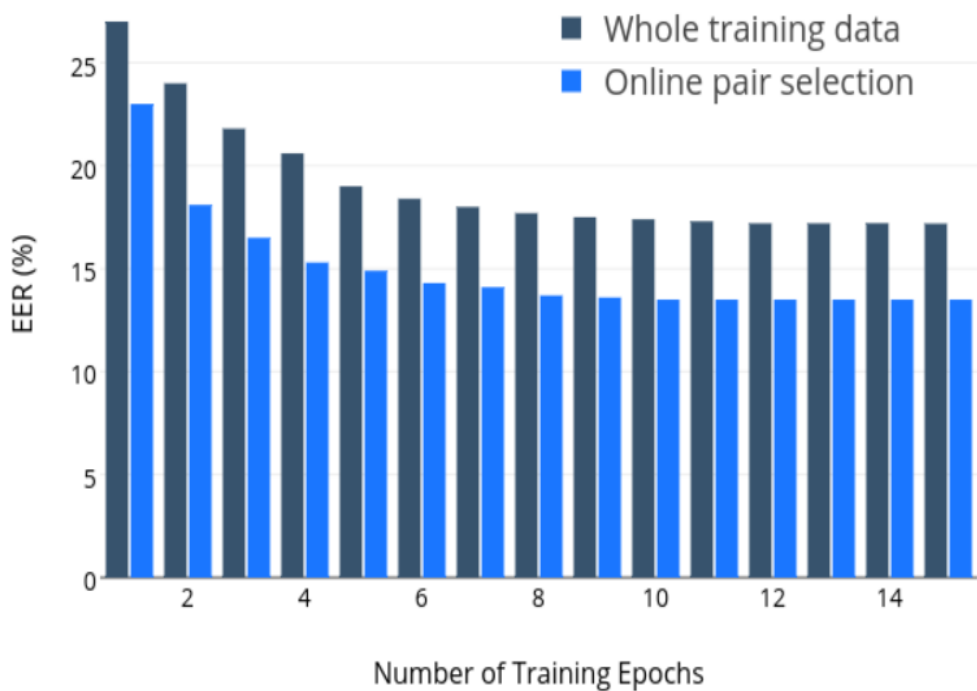


图 3-6

对于识别的环节，本文主要使用 K 均值聚类算法。使用 K 均值聚类算法，一般来说，第一步就是要确定 K 值，合适的 K 值能够使得实验的准确度大幅度的提高。然后对于下面的十个单词进行训练，得到下图 3-7 所示。

识别对象	隐含状态数：4	隐含状态数：5	隐含状态数：6
ZERO	80	90	90
ONE	100	100	90
TWO	100	100	100
THREE	100	100	100
FOUR	90	90	80
FIVE	80	90	100
SIX	100	100	100
SEVEN	100	100	100
EIGHT	90	100	70
NINE	100	100	90

图 3-7

本文实验采用的是官方数据集进行训练，主要内容是数字 0-9 的英文发音的识别。根据实验结果我们可以发现，在进行多次实验以后我们可以发现当观测类别数目为 5 的时候，系统的整体识别程度最好。而且以上的单个数字的辨识度基本都能超过 90%，所以使用上述的组合数字，则能够较为安全的认证用户身份。

## 4 总结和展望

### 4.1 工作总结

唇语识别对于未来的身份认证有着非常深远的意义，这项技术也会随着应用场景的不断增多和复杂化，而不断的趋于成熟与准确。本文在对唇语识别技术进行了研究中，使用了卷积神经网络和深度学习框架。主要工作分为了下面的几个部分：

第一，对于视频进行预处理，将输入的视频变成 30f/s 的相等帧速，然后我们使用 Dlib 库，进行面部有效信息的提取，再将所有唇部区域的大小调整为具有相同的大小用 14 个关键点描述嘴唇特征；

第二，使用卷积神经网络提取嘴唇区域的特征，并且使用 Alexnet 卷积神经网络来对每一个帧的图像进行处理，由于其较强大的特征处理能力，使得提取出嘴唇区域的特征，并将卷积神经网络的全连接层的输出的输出作为特征[6]。

第三，局部感知能力强，这使得其能够更好的感知图像的更多细节地方，还有就是参数共享也是卷积神经网络的优点，这使得我们只注重每个神经元感知的部分区域，不需要关心其他的区域范围之外的区域，这类似与生物的神经网络，使得模型算法的时间复杂度和空间复杂度降低。这样的神经网络能够从大量的样本中学习到相应的特征，且避免了平移，倾斜和旋转等变形带来的不确定性的影响，而且其降维的特点，减少不相关的特征值的数量，能够很好的避免复杂的特征提取，使得算法的运行更有效率。

### 4.2 未来展望

本文对于卷积神经网络的学习和唇语的识别有了一定的研究，而且唇语识别与生活有着紧密的联系，本文并没有开发出新的框架或者提出什么更好的算法，只是对于网上一些公开的项目，进行了反复实验，验证数据，得到相关数据。尽管实验达到了预期设定的目的，但是如果能够改良实验的其他相关的要素或者环境的话，将可以得到更好的实验结果。下一步的研究可以从下面的一些方面去改进：

（1）使用更多的数据，选择更加合适的样本集，样本数据量的大小会在实验中对改进模型或者系统有着直接性的关系，直接影响了模型参数的选择和数据的处理方式。对于数据集的选择可以选择样本数量更多的数据集，或者可以对于每一个样本进行更多的特征值提取和数据处理，使得数据的完整性更好；

（2）使用更好的设备，随着实验模型和算法的复杂度不断的提高，我们发现硬件配置在实验的重要性越来越高，更加强大的硬件能够支持更多的数据计算和储存；

（3）增加视频文件的帧处理，文本对于视频处理为 30 f/S，可以选择对视频采用更

多的帧数处理，但是这伴随着成几何倍数增长的计算量，对于系统和模型的稳定性要求也就越高，也对硬件造成了更高的要求；

（4）使用更加高效的算法，同时可以进一步降低目前算法的时间复杂度和空间复杂度，这个在识别阶段中，能够使得识别速度更快，处理速度更快。

上面的叙述是我们对于未来唇语识别发展的推测，希望未来人们可以在唇语识别的领域，做出突破性的进展和研究。

## 参考文献

- [1] 邓广宏, 曹万华, 夏奕, 蔡斌, 池志强, 程雄. 基于特征向量的基本口型分类[J]. 计算机应用与软件, 2010, 27(10):46-48+64.
- [2] 吕品轩. 基于唇语身份识别的特征鉴别力分析[D]. 上海交通大学, 2007.
- [3] 刘娜. 基于卷积神经网络的行人重识别算法[D]. 华东师范大学, 2017.
- [4] 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]. 中国科学院重庆绿色智能技术研究院, 2016.
- [5] 任玉强, 田国栋, 周祥东, 吕江靖, 周曦. 高安全性人脸识别系统中的唇语识别算法研究[J]. 计算机应用研究, 2017, 34(04):1221-1225+1230.
- [6] 苏小芳, 戴静. 基于 Fourier 级数的谐波计算方法[J]. 科技创业月刊, 2014, 27(01):180-181.
- [7] 吴岳. 社交多媒体数据语义理解和关联表达[D]. 中国科学技术大学, 2017.
- [8] 吴晓阳. 基于 OpenCV 的运动目标检测与跟踪[D]. 浙江大学, 2008.
- [9] 汪飞跃. 基于步态触觉信息的足迹提取与动态识别方法研究[D]. 安徽大学, 2015.
- [10] 徐铭辉, 姚鸿勋. 基于句子级的唇语识别技术[J]. 计算机工程与应用, 2005(08):86-88.
- [11] 杨恒翔. 基于图像的嘴唇特征提取及口型分类研究[D]. 天津大学, 2014.
- [12] 张洋阳. 基于卷积神经网络的场景分类的研究与应用[D]. 兰州理工大学, 2017.
- [13] 张亚飞, 耿梦伟, 尹玉倩, 齐立萍. 基于人工智能的唇语识别技术[J]. 信息与电脑(理版), 2018(10):121-122.
- [14] 蔡瀛. 基于隐马尔可夫模型的唇语识别方法研究[D]. 北方工业大学, 2018.
- [15] Multimodal person recognition using unconstrained audio and video. Choudhury, T., Clarkson, B., Jebara, T., Pentland, A. Proceedings, International Conference on Audio- and Video-Based Person Authentication . 1999
- [16] Face liveness detection using 3D structure recovered from a single camera. Wang Tao, Jian wei Yang. Biometrics (ICB) 2013 International Conference on . 2013
- [17] Going deeper with convolutions. Szegedy C, Liu W, Jia Y, et al. Computer Vision and Pattern Recognition . 2015

## 致 谢

时光飞逝，随着毕业时间的到来，大一新入学的那一天就像昨天一样，却永远也回不去了，在这里我要感谢那些在大学四年里陪伴我的那些人致谢。

首先就是我的导师，郑芳老师。郑老师在我刚入学的时候，在学习上培养了我独立学习的能力，这使得我的大学期间以至于未来都受益匪浅，在生活上，郑老师对我来说亦师亦友，在我迷茫和困惑的时候给我答疑解惑。

然后就是我的那些可爱的室友和同学们，以及五湖四海的朋友们，是他们陪伴我走过了一场独一无二的大学生涯。他们在学习和生活上给了我灵感和帮助，让我的大学生活才如此的绚烂多彩。

最后，感谢我的家人，多年来的精神和思想上给予我的关心、鼓励和信任，让我在前进的道路上，有了坚强的后盾，是家人的支持，才让我有了充实的大学生活。



修德立信 博学求真