数字图像处理

课程设计报告

题目：基于HOG特征的嘴部识别算法

学号：3200306048

姓名：刘鑫宇

班级：计算机2006

2020年5月24日

目 录

1. 课程设计任务

1.1 设计任务

1.2 任务分工

1. 课程设计原理及设计方案

2.1 课程设计原理

2.2 课程设计方案

1. 课程设计步骤和结果

3.1 图像预处理

3.2 HOG特征提取

3.3 分类器对HOG特征向量分类

3.4 加载关键定位器

3.5 判断嘴巴是否张开

3.6 基于形态学分割嘴巴

1. 课程设计总结
2. 设计体会

参考文献

一、课程设计任务

1.1 设计任务

通过传统图像处理技术对人脸进行检测并识别嘴巴是否张开，要求不能使用深度学习技术。能处理不同环境下的人脸，且有一定的准确性。

1.2 任务分工

组长刘鑫宇：对图像进行HOG特征提取和分类器分类实现人脸识别，68个关键点定位五官，论文撰写。

组员孙君宇：基于关键点技术判断嘴部是否张开逻辑实现，ppt制作，论文撰写。

组员林晋：基于关键点技术图像分割嘴部并用形态学操作处理图像，ppt制作，论文撰写。

二、课程设计原理及设计方案

2.1课程设计原理

想要实现对图片嘴部的识别以及判断是否张嘴，首先需要对复杂的图片进行预处理，我们最后实现的设计应该具有一定可行性，即不同环境、不同人种、不同性别、不同年龄等人都能通过这个设计锁定到嘴部并判断是否张开。预处理还要统一输入图片尺寸，以便于向后面程序处理。图片的人脸可能在各个位置，以及大小形状都有可能不同，那么首先需要做的就是人脸的检测，只有检测到人脸才能进行后续的操作。我们的设计准确到人的五官，当锁定了人脸后便可以通过某种算法选中我们需要的嘴巴，以便于后续对嘴巴的状态进行判断。锁定到嘴巴后通过某种算法判断嘴巴的状态。通过图像切割的方法分割出我们的嘴巴并最后输出嘴巴的状态。整体设计思路如图1所示：

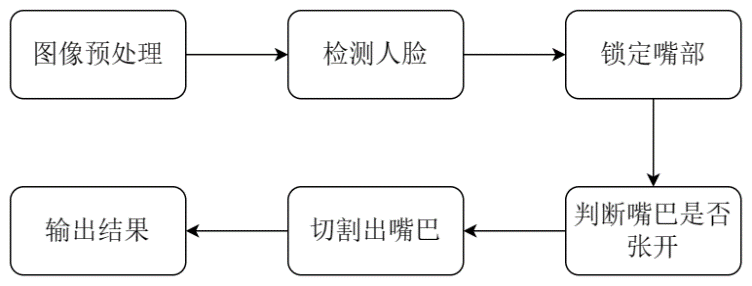


图1 整体设计思路

2.2设计方案

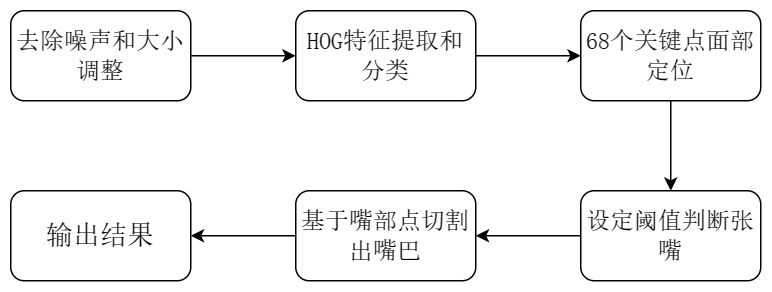


图2 整体设计方案

（1）图像预处理

图像预处理是计算机视觉应用中非常重要的一步，可以提升后续图像处理任务的准确性和效率。以下是图像预处理可能采取的方案：

去除噪声：使用滤波器（例如高斯、中值等）去除噪声。

图像大小调整：将图像缩放为特定的大小或纵横比例以便于后续处理。

色彩空间转换：将图像从RGB色彩空间转换至灰度或其他代表颜色信息的色彩空间，以保留更少的数据，提高计算效率。

图像增强：通过调整对比度、亮度等方式强化图像的特征。

边缘检测：使用边缘检测算法（例如Sobel、Canny等）找出图像中的边缘。

直方图均衡化：调整图像的直方图，以使其像素值分布更加均衡，以提高图像质量。

形态学操作：通过腐蚀、膨胀等形态学操作对图像进行处理，以消除干扰、突出轮廓等。

阈值分割：将图像二值化，以提高对对比结构的敏感程度。

本实验选输入图像是有一定要求，所以预处理工作不必太过繁琐，只需要去除噪声以及图像大小调换。

（2）检测人脸

检测人脸是如今一项重要的识别任务，常用的人脸识别算法包括：

Haar cascades：基于AdaBoost分类器和Haar-like特征的算法，通过级联检测器从图像中找出人脸区域。

HOG (Histogram of Oriented Gradients)：利用图像中的梯度方向信息作为特征，采用SVM（Support Vector Machine）进行分类，以识别人脸。

DNN (Deep Neural Networks)：基于深度学习的神经网络算法，通过多层卷积和池化操作，训练网络完成对人脸的鉴别任务。

Faster R-CNN：基于区域提取卷积神经网络(RPN)和Fast R-CNN算法，实现了单阶段的端到端对象检测算法，可以准确地检测人脸。

SSD (Single Shot Multibox Detector)：单次检测多框架检测器(SSD)采用卷积和后处理技术，顺序性地预测若干边界框和所含目标的置信概率，在保证高召回率的前提下，大幅提升了准确率。

排除掉深度学习算法，我们选择HOG算法。基于HOG特征和分类器的人脸检测算法是一种经典的计算机视觉技术，其原理是通过“人脸+非人脸”的训练数据集来训练分类器，再利用训练好的分类器对待检测图像HOG特征进行分类，判别可能存在的人脸区域，找到真实的人脸。

该算法具有可拓展性好、计算速度快等优势，在物体检测和定位任务中有很广泛的应用。缺点是无法区分人脸的朝向、表情和光照等因素的影响，对人脸方向的检测较为敏感。

（3）锁定嘴部

对于定位五官的任务我们选择使用别人已经训练好的模型68 landmark localization-agnostic model，它通常也被称为全局人脸特征（global facial feature）模型。旨在通过对人脸的深度学习训练来定位面部的各个关键位置。

这种算法提取了人脸图像中的关键特征，并使用神经网络以端到端的方式对这些特征进行处理和学习。具体地说，模型将人脸图片输入到CNN（卷积神经网络）中以提取特征。然后，池化层、全连接层等操作被应用于从这些特征中提出的数据来预测每个关键点的坐标。

相比于传统的基于回归模型的关键点检测算法，无角色定位模型主要在以下两个方面取得了进展：首先，与直接对坐标进行回归不同，该算法从原始图片中提取特征后再进行坐标预测，因此可以更好地适应人脸的变形和旋转；其次，由于该算法是无角色定位的，因此即使面部有不同程度的遮挡或改变表情带来的变形，算法也可以稳健地进行预测。

68的关键点是指模型可以同时预测68个面部关键点的坐标，包括眼睛、鼻子、嘴巴、下巴等特征。与更少数量的关键点（如5或12个）相比，68个关键点可以提供更丰富的面部细节信息，从而使应用程序能够进行更精确的人脸分析和识别。

（4）判断是否张嘴以及切割嘴巴

基于68个面部关键点判断嘴巴是否张开，设定一个阈值，通过判断上嘴皮下部分与下嘴皮上部分两个点的距离大小来判断嘴巴是否张开。切割嘴巴以便于我们实验观察也是直接通过嘴部的关键点进行操作。

（5）输出结果

三、课程设计步骤和结果

3.1 图像预处理

人脸图像可能是任意大小的，通常在多个尺度上分析多个图像位置，唯一的限制是被分析的滑动窗口（win）具有固定的长宽比（纵横比）。例如滑动窗口需要有一个长宽比为1:2。然后再将滑动窗口的图像调整为128x64像素的大小，该设置在行人检测任务中可以获得更好的结果。

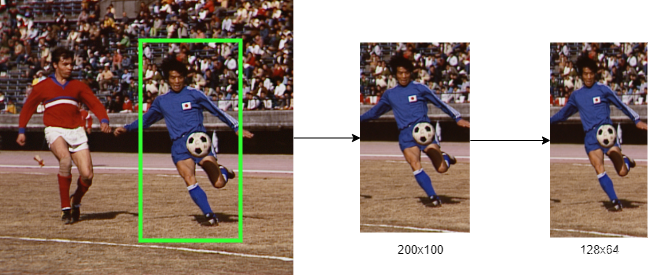


图3 图像预处理

3.2 HOG特征提取

HOG特征提取主要步骤如下：

（1）计算图像梯度

  梯度是通过结合图像的大小和角度来获得的。考虑3x3像素的块，首先计算每个像素的和。对于每个像素值，首先使用以下公式计算和:

其中r为像素点的横坐标，c为像素点的纵坐标，为图像像素点的像素值，在计算和后，使用下面提到的公式计算每个像素的梯度幅值和角度：

使用opencv中内核大小为1的Sobel算子实现计算图像的水平和垂直梯度，并计算每个像素的梯度和角度如图4所示：

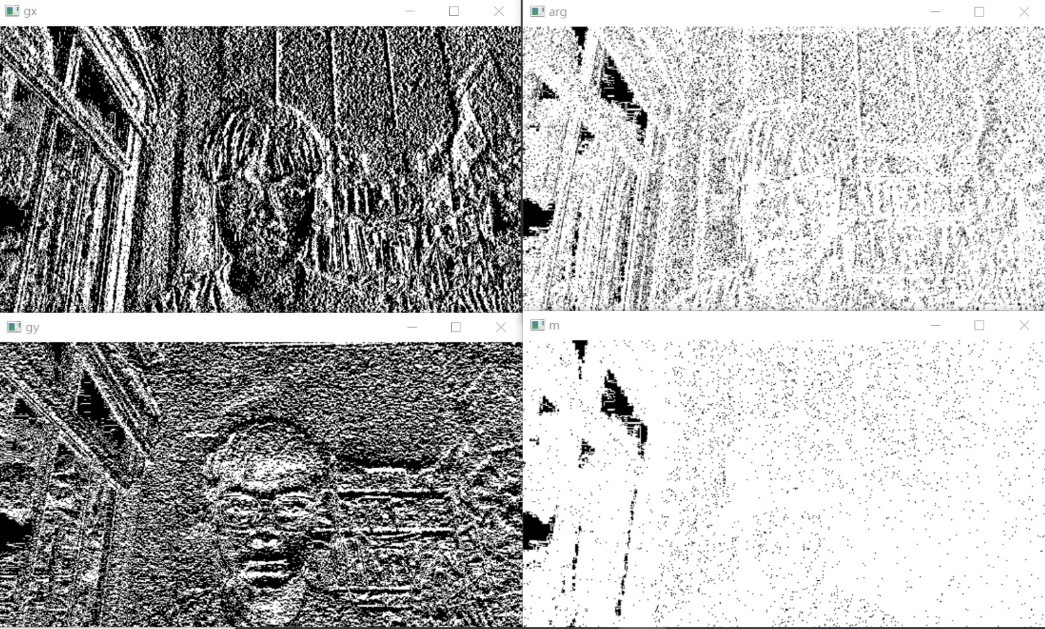


图4 人脸每个像素梯度和角度

其中左上为水平梯度图像，左下为垂直梯度图像，右上为角度图像，右下为图像梯度图像。水平梯度图像是计算像素点所有像素值的变化趋势，由于柜子的缝隙是竖直的，而缝隙两边图像相似，计算得到缝隙位置的梯度值基本相同，于是可见一条竖线，而垂直梯度则可以看见横着的书架平台。

（2）计算细胞梯度直方图

在获得每个像素的梯度后，梯度矩阵（梯度和角度矩阵）被划分为8x16个大小为8x8的单元细胞（cell），如下图5所示：

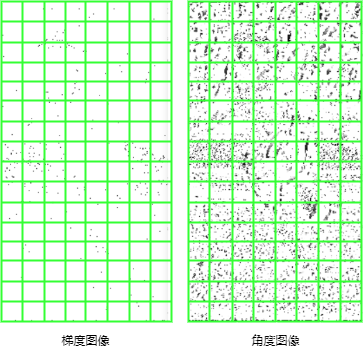


图5 8×8细胞单元

我们之所以把图像分为8×8细胞是因为它提供了一个紧凑的特征表示。8×8单元细胞包含8x8 = 64像素值。这个单元细胞梯度每像素包含2个值（大小和方向），那就变成8x8x2 = 128个数。通过后面我们会看到这128个数字是用9-bin直方图(可存储为长度9的数组，通俗地说是分别装到9个箱子里)表示。不仅是表示更简洁，计算在细胞梯度直方图具有更强的鲁棒性。个别的颜色梯度可能有噪音，但用一个8x8单元细胞梯度来表示梯度直方图对噪声不敏感，换句话说，就是受噪音影响不大。如图6所示，让我们看一个8×8单元细胞的图像梯度的面貌。

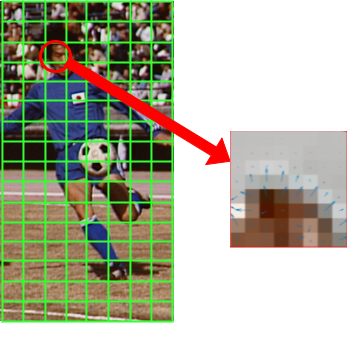


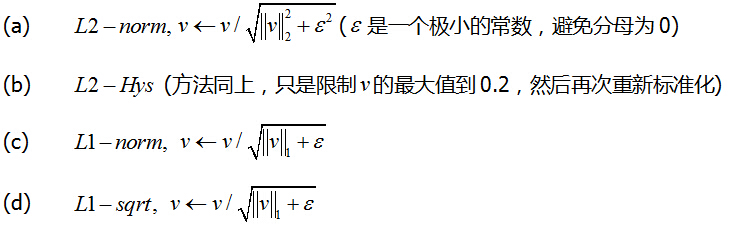
图6 细胞梯度与角度矩阵

上图显示了用箭头表示梯度的图像的贴片，箭头显示梯度的方向，其长度显示大小。注意箭头的方向指向强度的变化方向，其大小表示差异的大小。在右边，我们看到原始的数字表示在8×8细胞有一个小的差异，那就是角度是0度和180度的梯度，而不是0到360度之间。这些被称为“无符号”渐变，因为梯度正值和它的负值用相同的数字表示。换句话说，梯度箭头和与之相对的180度箭头被认为是相同的。这样我们就得到了一个9×1的向量。

（3）直方图归一化

在前面的步骤中，我们根据图像的梯度创建了一个直方图。图像的梯度对整体亮度是敏感的。如果通过将所有像素值除以2来使图像变暗，则梯度矩阵元素将改变一半，因此直方图值将改变一半。理想情况下，我们希望特征描述符不受光照变化的影响。换句话说，我们希望使直方图归一化，这样它们不会受到光照变化的影响

归一化的方法大致有四种：



通过查看文献对比发现L2-Hys归一化方法最好，于是采用该方法进行归一化。

（4）计算HOG特征向量

通过滑动窗口，每次移动一个cell，计算每个窗口的向量，直到整个图像覆盖完，最后得到该图像的HOG特征向量

HOG 是根据细胞单元格创建的，这些细胞单元格组合成区块，并且区块之间可以重叠，因此很难对它进行直观显示。不过可以通过显示每个单元格的直方图来表示HOG。如图14所示，显示方向直方图时，不使用柱状图，而是采用更加直观的星形图，每个线条的方向与 bin 对应，长度与 bin 的数量成正比。可以用这种方法在图像上绘制 HOG。画出图像如图7所示。

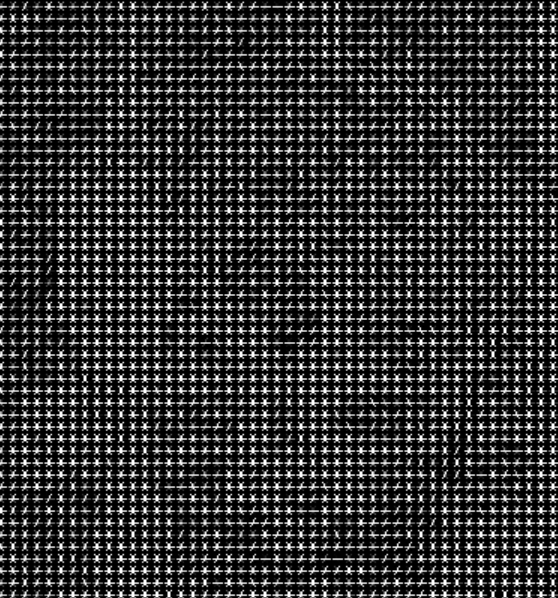


图7 人脑袋的HOG特征向量星形图

3.3 分类器对HOG特征向量分类

该分类器使用了Dlib库提供的support vector machine（SVM）算法，在训练过程中采用hard negative mining策略，即在更多的背景图片上进行不断的负样本分类器训练，进一步提高了分类器在真实场景下的鲁棒性。同时，该分类器还使用了cascade structure、pyramid representation等技术，将计算复杂度降至较低水平，从而在保证检测准确率的同时大大提高了检测速度。

3.4 加载关键点定位器



使用了 Dlib 库中的 shape\_predictor() 函数，并加载了一个名为 'shape\_predictor\_68\_face\_landmarks.dat' 的数据文件。这个数据文件是训练好的一个面部关键点检测模型，可以用于在图像或视频中自动检测面部关键点。

在具体应用中，我们可以将一张包含人脸的图像作为输入，然后利用上述加载的模型，调用它的 predict() 函数来获取该张图像中所有人脸的 68 个预定义关键点位置，从而实现一些功能，例如：面部特征分析、面部表情识别、人脸识别、头部姿态估计以及虚拟化妆等视觉应用。

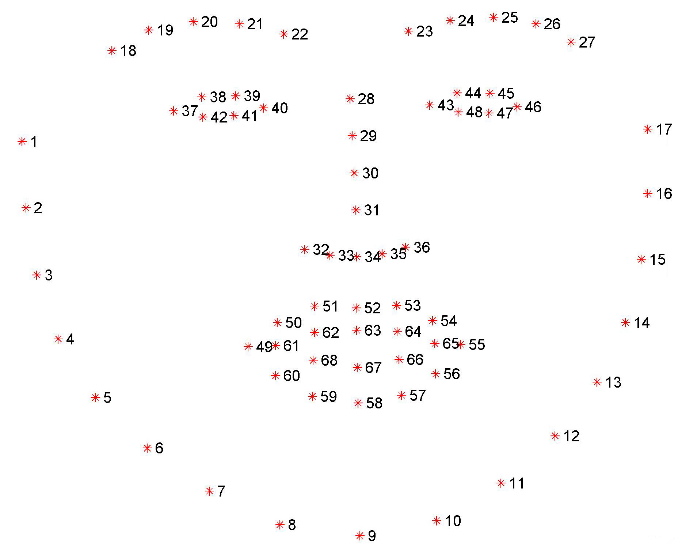
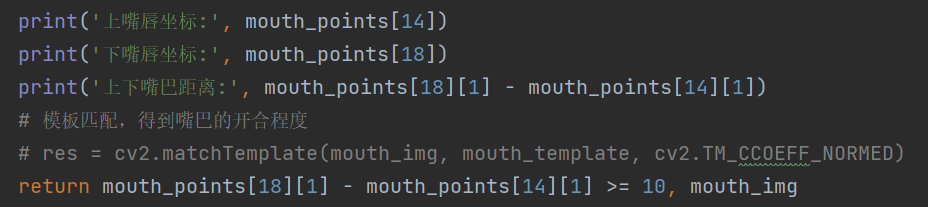


图8 68个关键点人脸定位器

3.5 判断嘴巴是否张开

判断嘴巴是否张开的操作基于68个关键点。首先获取嘴部的关键点，以便于后续绘制显示器，以及分割嘴部。然后使用63点和67点，即上嘴唇下部和下嘴唇上部，计算两点的距离，再设定一个阈值，若两点距离大于阈值则判断嘴巴张开，否则嘴巴是闭上的状态。



如上所示上嘴唇下部点坐标为mouth\_point[14]，下嘴唇上部点坐标为mouth\_point[18]，直接返回一个bool类型判断是否嘴巴是否张开。如下图9所示，张开嘴巴程序不断输出两点坐标并计算距离，和阈值10作比较。距离为28，大于10，则嘴巴是张开的。



图9 嘴巴张开测试

如下图10所示，张开嘴巴程序不断输出两点坐标并计算距离，和阈值10作比较。距离基本上为0和1，小于10，则嘴巴是张开的。若实验效果不佳可返回调整参数到合适位置。

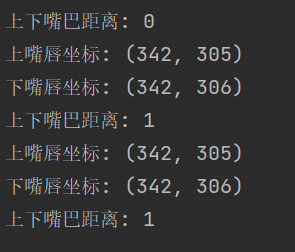
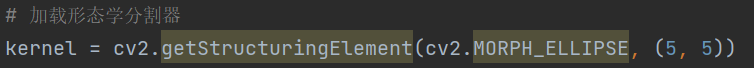


图10 嘴巴闭上测试

3.6 基于形态学分割嘴巴

为了便于我们实验观察，我们打算将嘴部图像分割出来，同时使用形态学开操作处理嘴部图像，得到效果更佳的实验结果。



创建了一个结构元素 (kernel) ，用于进行形态学上的图像处理。 getStructuringElement() 函数用于创建指定尺寸和形状的结构元素，它的两个参数分别是：形态学操作的类型和内核的大小。

其中，第一个参数 cv2.MORPH\_ELLIPSE 表示使用椭圆型的内核，第二个参数 (5, 5) 表示该内核的大小为 5 行 5 列。

椭圆型的内核比较适合用来处理像眼睛和嘴巴等区域比较复杂且形状不规则的部位。此处我们使用开操作对口部区域进行处理，从而能够更好地分割出嘴巴区域，为后续分析、特征提取和识别等任务提供必要支撑。

最后实现效果如下图11所示：

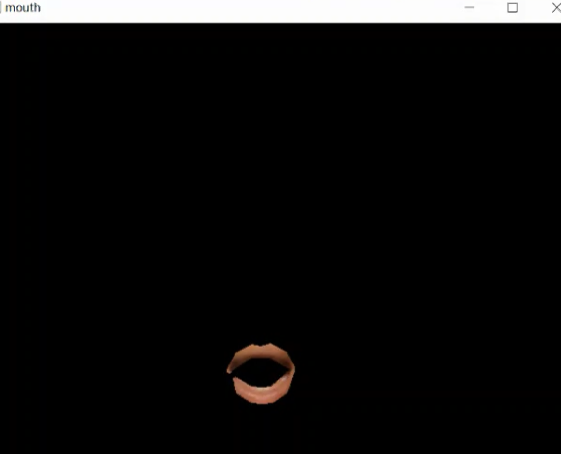
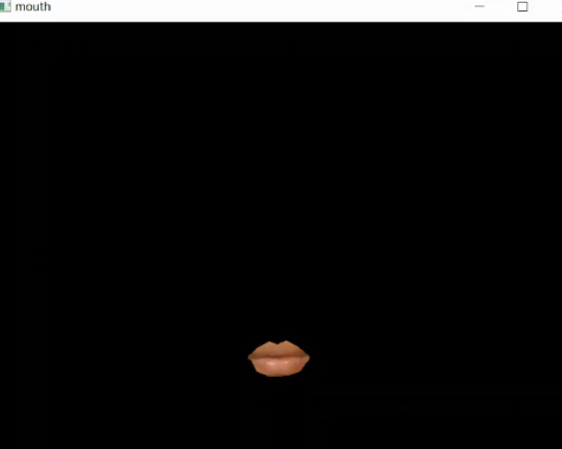


图11 分割嘴巴实现效果

四、课程设计总结

本课程设计通过HOG特征提取技术提取图像特征，然后使用SVM分类器对特征进行分类，判断是否为人脸的一部分，实现对人脸的检测。然后通过68个关键点定位模型实现对人脸五官的定位，为后面判断嘴部状态做准备。然后基于上嘴唇下部点和下嘴唇上部点距离判断嘴巴是否张开并进行输出。最后通过图像分割技术以及图像学开操作处理得到嘴部分割图像，便于实验观察效果。实现效果较好，具有一定可行性，但使用时人需要主要光照强度，不用环境下可能会有不同的效果。

同时本课程设计的分类器和人脸五官定位器基于别人已经训练好的模型，才有较好的实验效果。我们可以自己训练相关模型，或者多尝试其他非深度学习方法来更加完善我们的设计。

五、设计体会

通过小组团队合作成果完成了本次课程设计，受益匪浅。课程设计中，我们集思广益，从众多方法中不断筛选合适的实验方法，查找相关文献，期间也多次失败，但我们小组还是坚持不懈最后成果实现了。这让我意识到课堂上学的知识是永远不够的，特别是对于我们计算机专业的学生，当做起项目来更多的还是现查找文献资料，所有我们平时更应该加强对自己感兴趣领域的研究，多看文章不断学习先进技术，这样才不会被时代淘汰。

参考文献

[1]李国玄,马凯凯,王文博.基于HOG特征提取和SVM的手势识别方法研究[J].传感器世界,2022,28(12):30-36.DOI:10.16204/j.cnki.sw.2022.12.011.

[2]任梦茹. 基于图像多特征提取的行人检测算法研究[D].西安工业大学,2021.DOI:10.27391/d.cnki.gxagu.2021.000069.

[3]刘力,龚勇,赵国强.三维人脸识别研究进展[J/OL].计算机工程与应用:1-24[2023-06-01].http://kns.cnki.net/kcms/detail/11.2127.TP.20230531.0936.002.html

[4]张家彬,孟建军.基于SVM的人脸识别研究[J].信息与电脑(理论版),2023,35(04):205-207.

[5]代一凡. 基于嘴部识别的智能助餐机器人设计研究[D].东华大学,2021.DOI:10.27012/d.cnki.gdhuu.2021.001002.