**机器学习实验报告**

**题目 基于Atlas开发板的人脸表情识别实验**

**目录**

[基于Atlas开发板的人脸表情识别实验 3](#_Toc29220849)

[1. 实验介绍 3](#_Toc29220850)

[2. 实验背景 3](#_Toc29220851)

[3. 实验原理 3](#_Toc29220852)

[4. 实验任务 3](#_Toc29220853)

[4.1 数据采集 3](#_Toc29220854)

[4.2 预处理 4](#_Toc29220855)

[4.3 模型训练 5](#_Toc29220856)

[4.4 模型部署 9](#_Toc29220857)

[5. 结果分析及总结 13](#_Toc29220858)

# 基于Atlas开发板的人脸表情识别实验

## 实验介绍

本实验主要介绍改进的AlexNet分类网络在caffe框架下训练得到表情识别的模型并部署在Atlas 200 DK开发板上执行。该实验采集表情图片作为输入，用该网络对表情进行分类，判断表情的标签，并将处理后的分类结果标注在图片上。

## 实验背景

人脸表情识别技术的发展和应用可以更好的实现智能技术与人们的交流，通过机器识别给定的静态图像或动态视频序列中分离出特定的表情状态，从而确定被识别对象的心理情绪，实现计算机对人脸表情的理解与识别，可从根本上改变人与计算机的关系，使计算机可以更好地为人类服务，从而达到更好的人机交互。

## 实验原理

基于改进的AlexNet网络结构进行图像特征提取及分类，整体的实验原理如下：

**图像采集**

**图像预处理**

**特征提取**

**表情分类**

**输出类别**

## 实验任务

### 数据采集

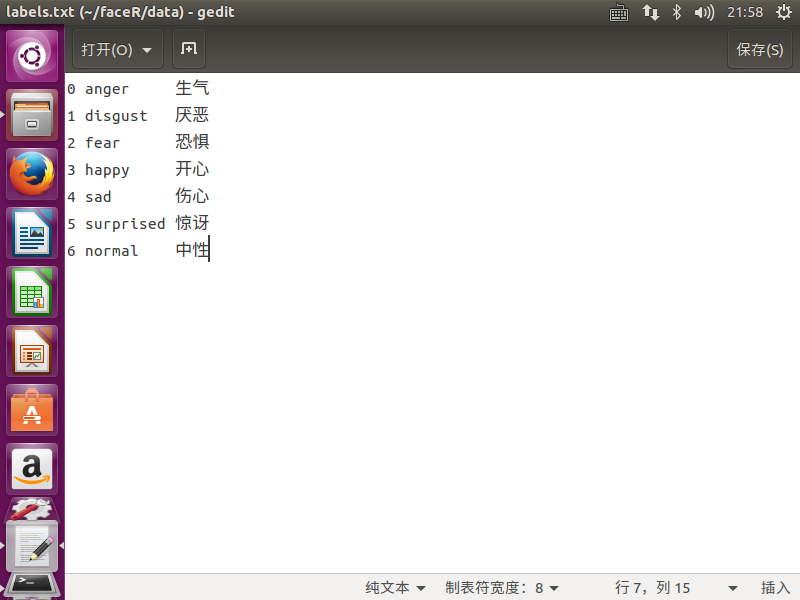
实验采用的数据集是fer2013人脸表情数据集，可以从kaggle官网 <https://www.kaggle.com/kenho211/fer2013-img>下载。该数据集共包含35887张人脸图片， 其中训练集28709张、验证集3589张、测试集3589张。数据库中的图片均为灰度图片，大小为48\*48像素，样本被分为0=anger(生气)、1=disgust(厌恶)、2=fear(恐惧)、3=happy(开心)、4=sad(伤心)、5=surprised(惊讶)、6=normal(中性)七类。训练集、验证集和测试集图片数如下表所示：

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | anger | disgust | fear | happy | sad | surprised | normal |
| Train | 3995 | 436 | 4097 | 7215 | 4830 | 3171 | 4965 |
| Val | 467 | 56 | 496 | 895 | 653 | 415 | 607 |
| Test | 491 | 55 | 528 | 879 | 594 | 416 | 626 |

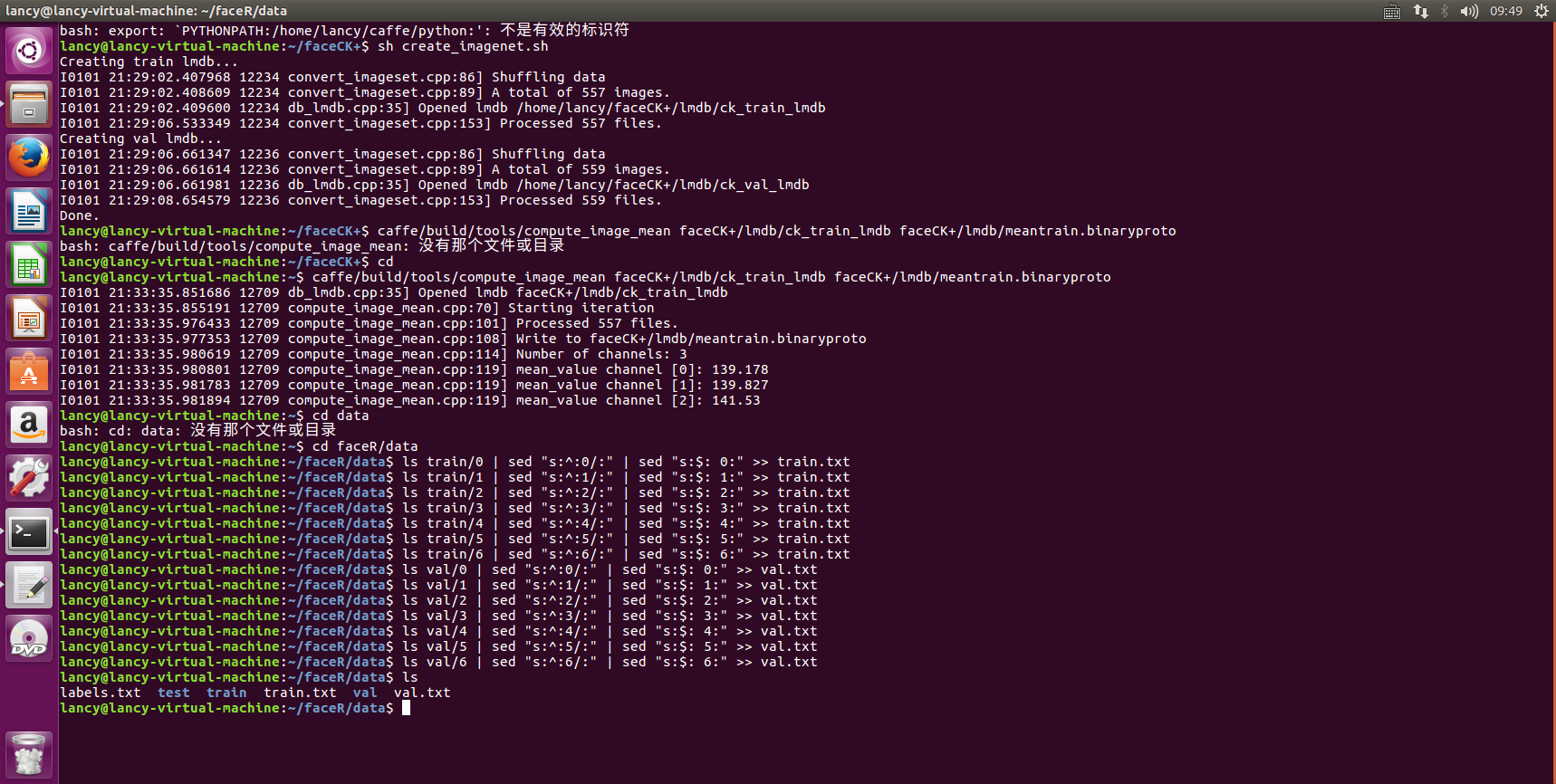
### 预处理

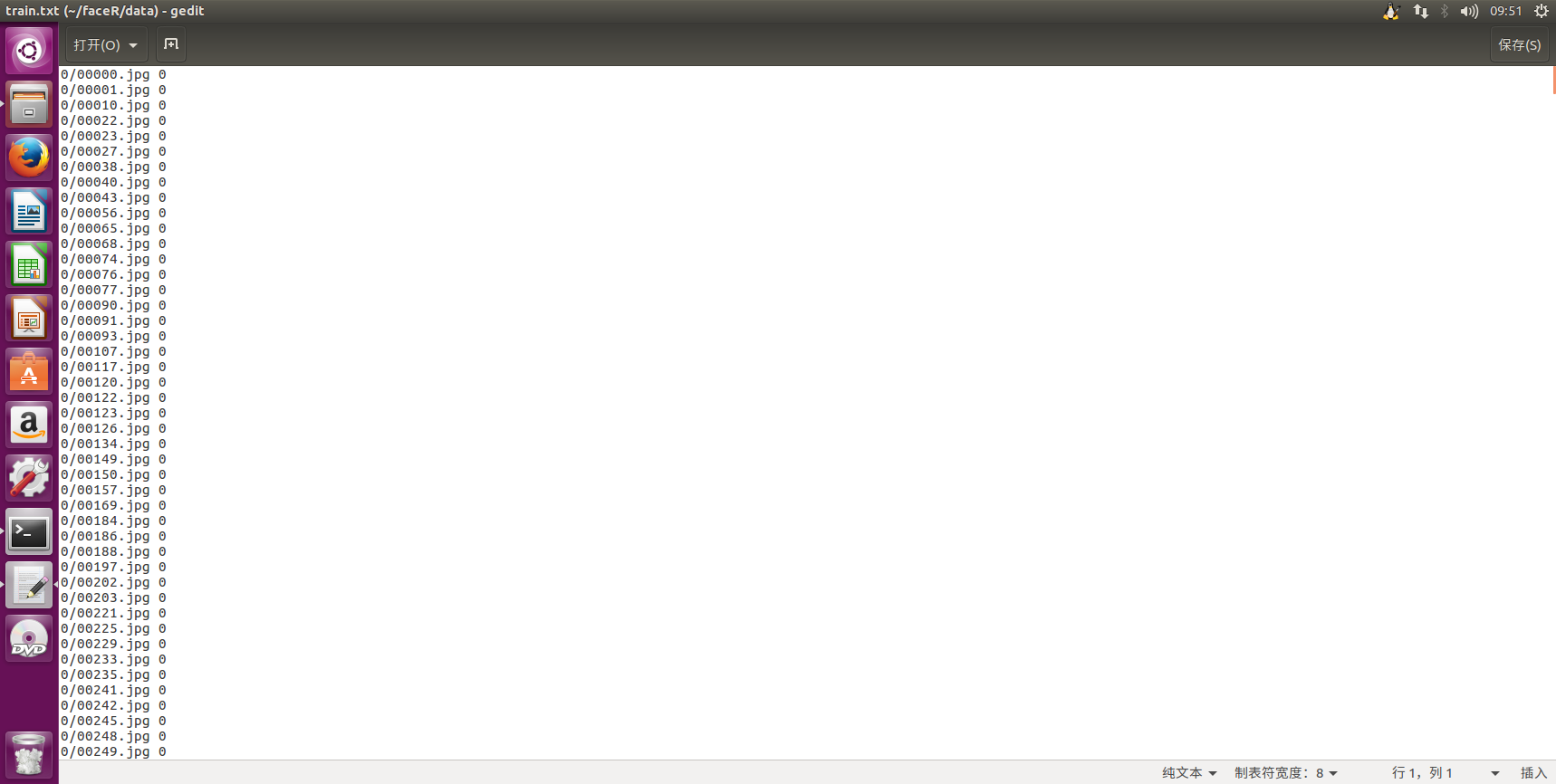
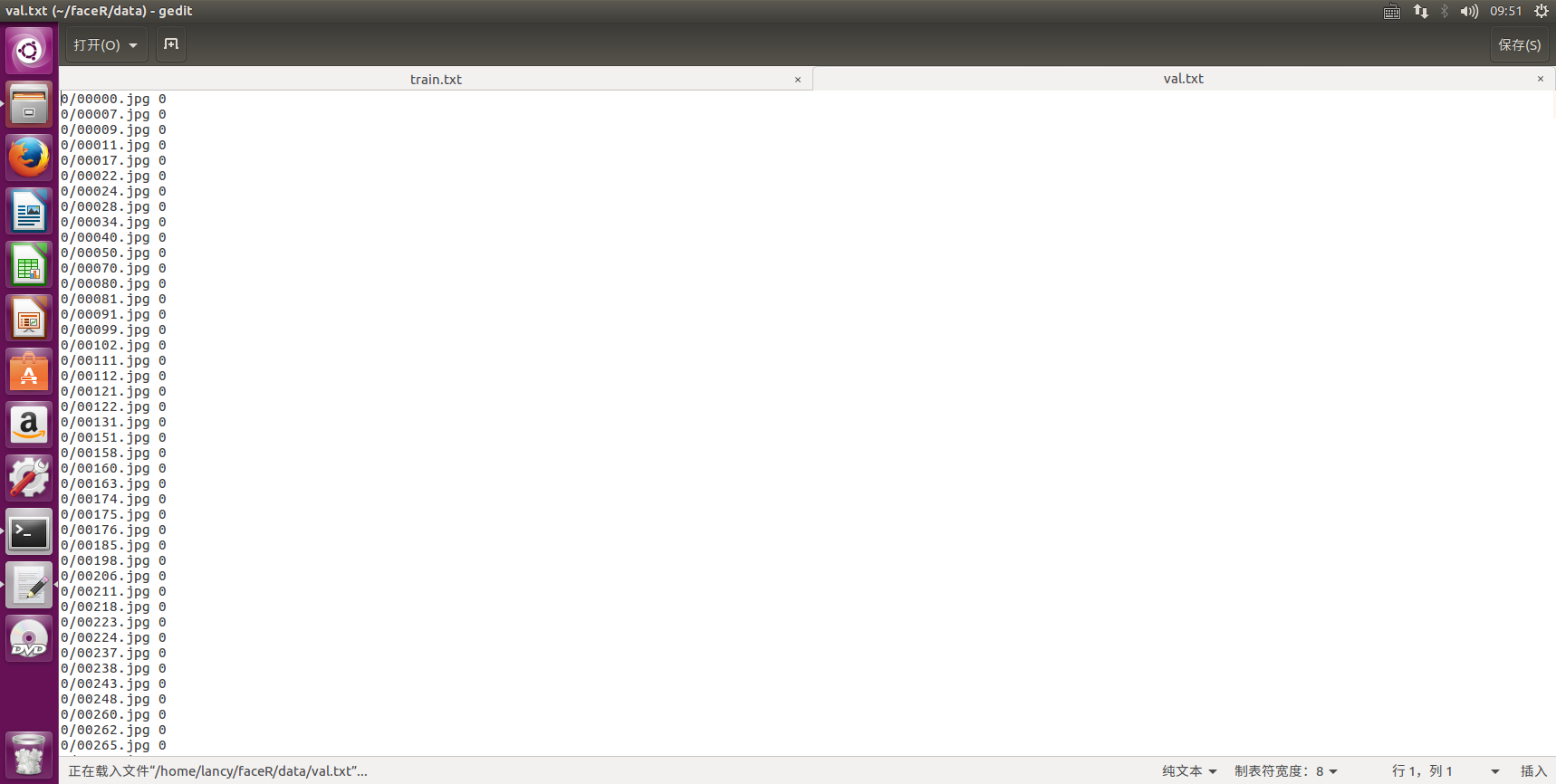
将该数据集存放在~/faceR/data的目录下进行数据预处理。

1. 创建label.txt文件，表示分类序号和表情类别的对应关系，如下图所示：

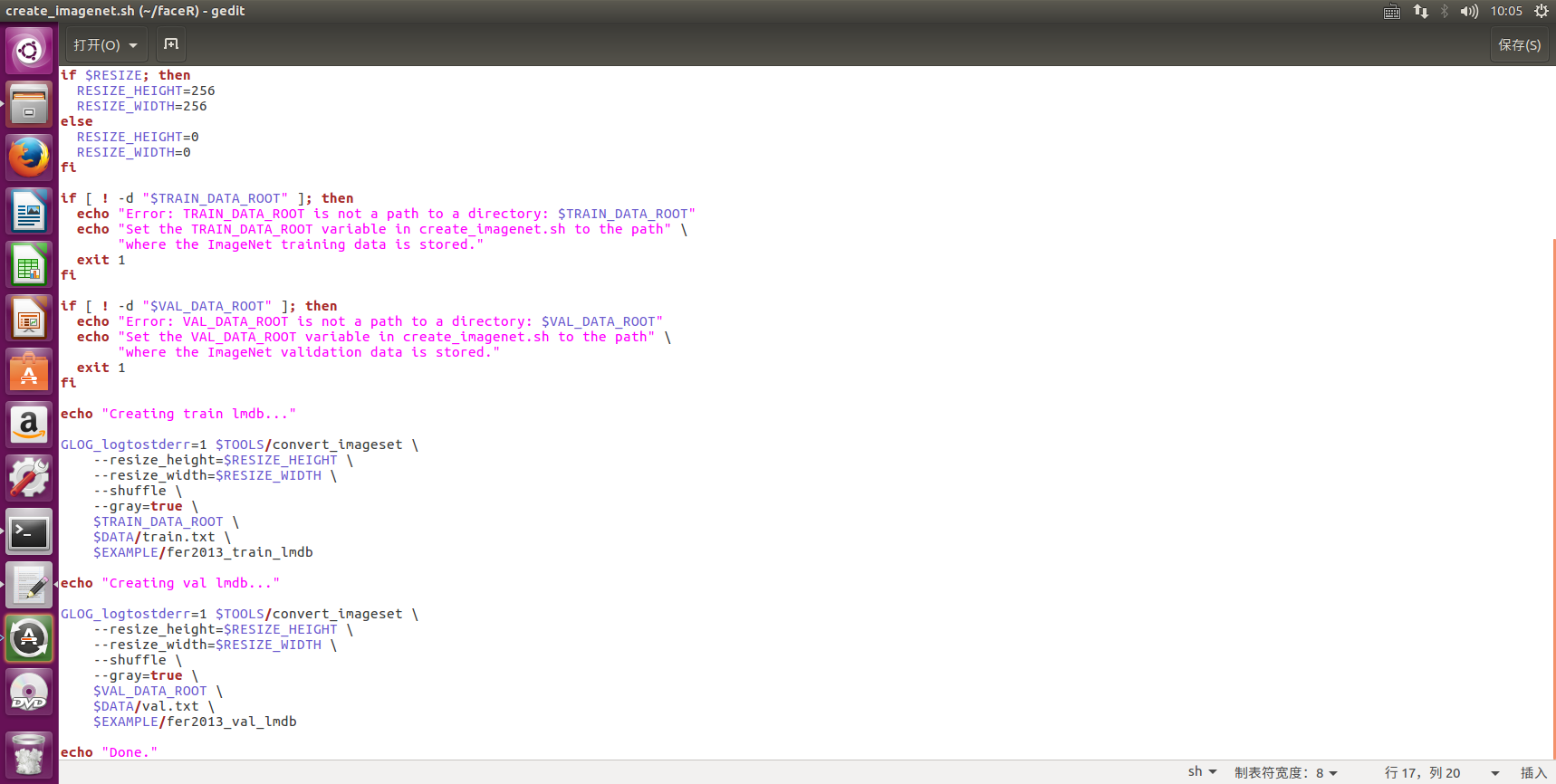


1. 创建train.txt文件，表示训练图片路径和分类序号的对应关系， 创建val.txt文件，表示验证图片路径和分类序号的对应关系。在data目录下使用命令行生成以上文件，如下图所示：



1. 将数据集转化成lmdb格式并生成数据集对应的均值文件。在caffe提供的creat\_imagenet.sh脚本上做出修改，部分代码如下图所示：



### 模型训练

1. 编写模型的训练脚本和相关的配置文件

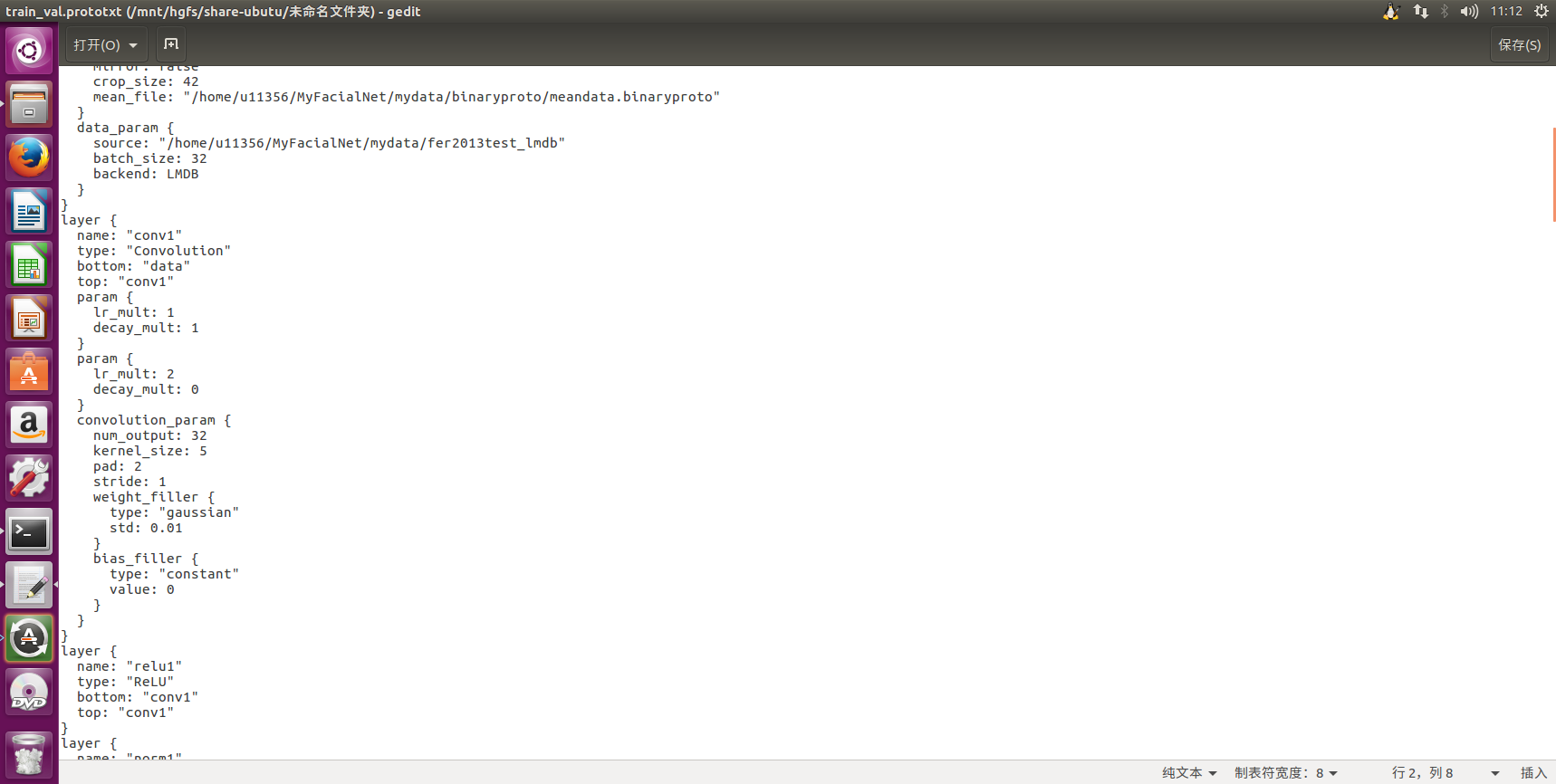
* train\_val.prototxt为定义网络结构的prototxt文件

该实验使用的网络结构是在Alexnet上做适当修改得到的，在Alexnet的基础上减少了3个卷积层，包含3个卷积层、3个池化层和3个全连接层。

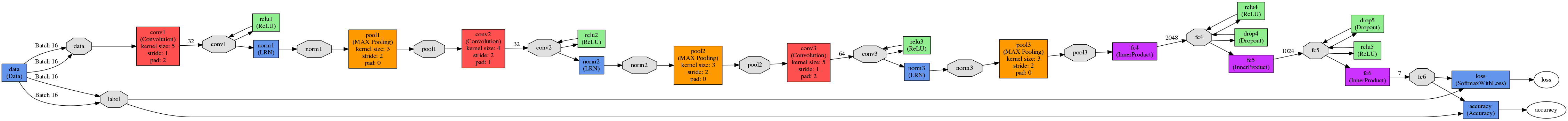
每个层设置的参数如下表所示：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type | Kernel\_size | Kernel\_num | stride | pad | output | dropout |
| data |  |  |  |  | 42\*42\*1 |  |
| convolution | 5\*5 | 32 | 1 | 2 | 42\*42\*32 |  |
| pooling | 3\*3 |  | 2 |  | 21\*21\*32 |  |
| convolution | 4\*4 | 32 | 1 | 1 | 20\*20\*32 |  |
| pooling | 3\*3 |  | 2 |  | 10\*10\*32 |  |
| convolution | 5\*5 | 64 | 1 | 2 | 10\*10\*64 |  |
| pooling | 3\*3 |  | 2 |  | 5\*5\*64 |  |
| innerproduct |  |  |  |  | 1\*1\*2048 | 0.5 |
| innerproduct |  |  |  |  | 1\*1\*1024 | 0.5 |
| innerproduct |  |  |  |  | 1\*1\*7 |  |

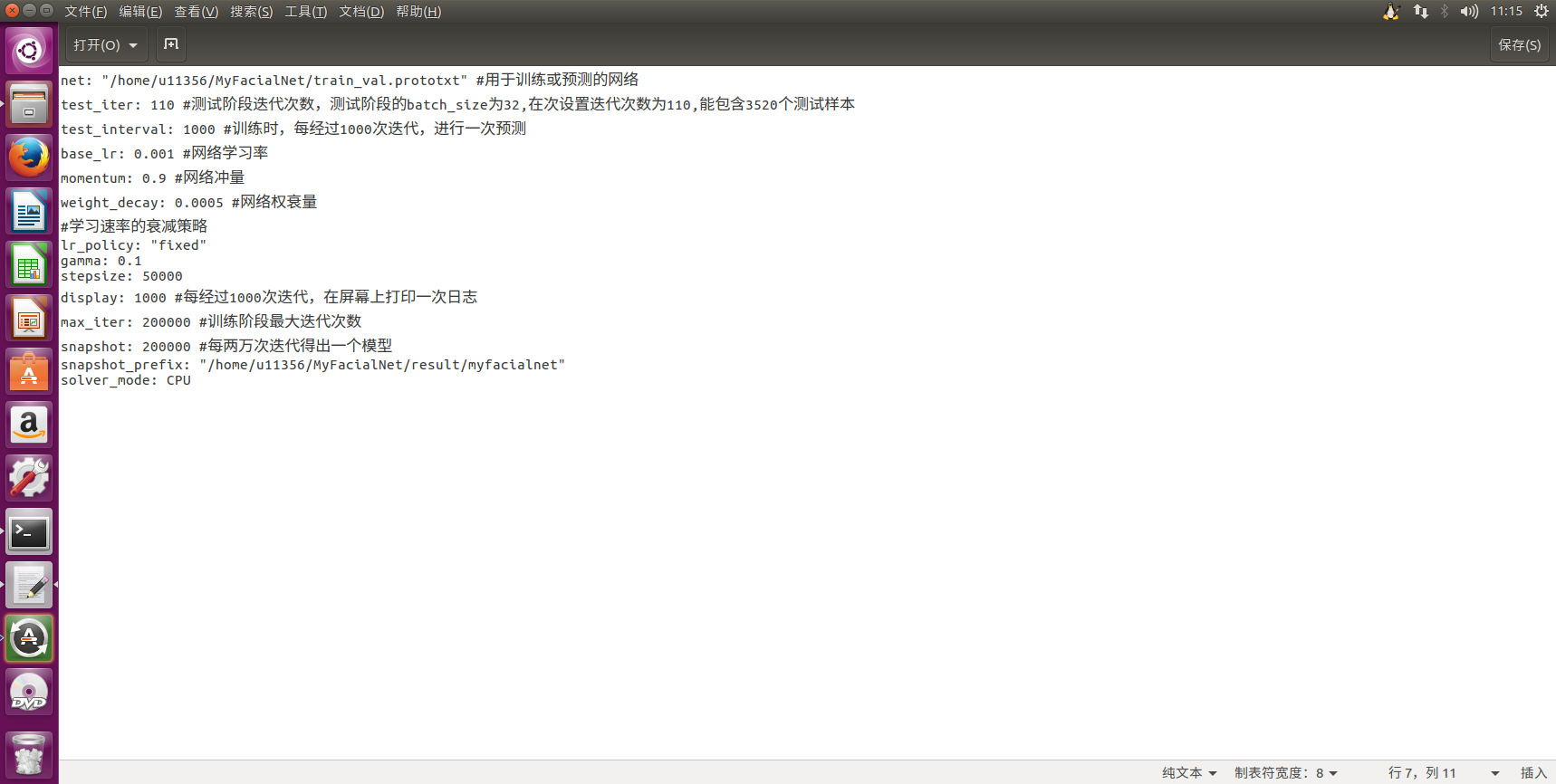
Train\_val.prototxt文件的部分代码如下图所示：



改进后的Alexnet网络结构图如下图所示：

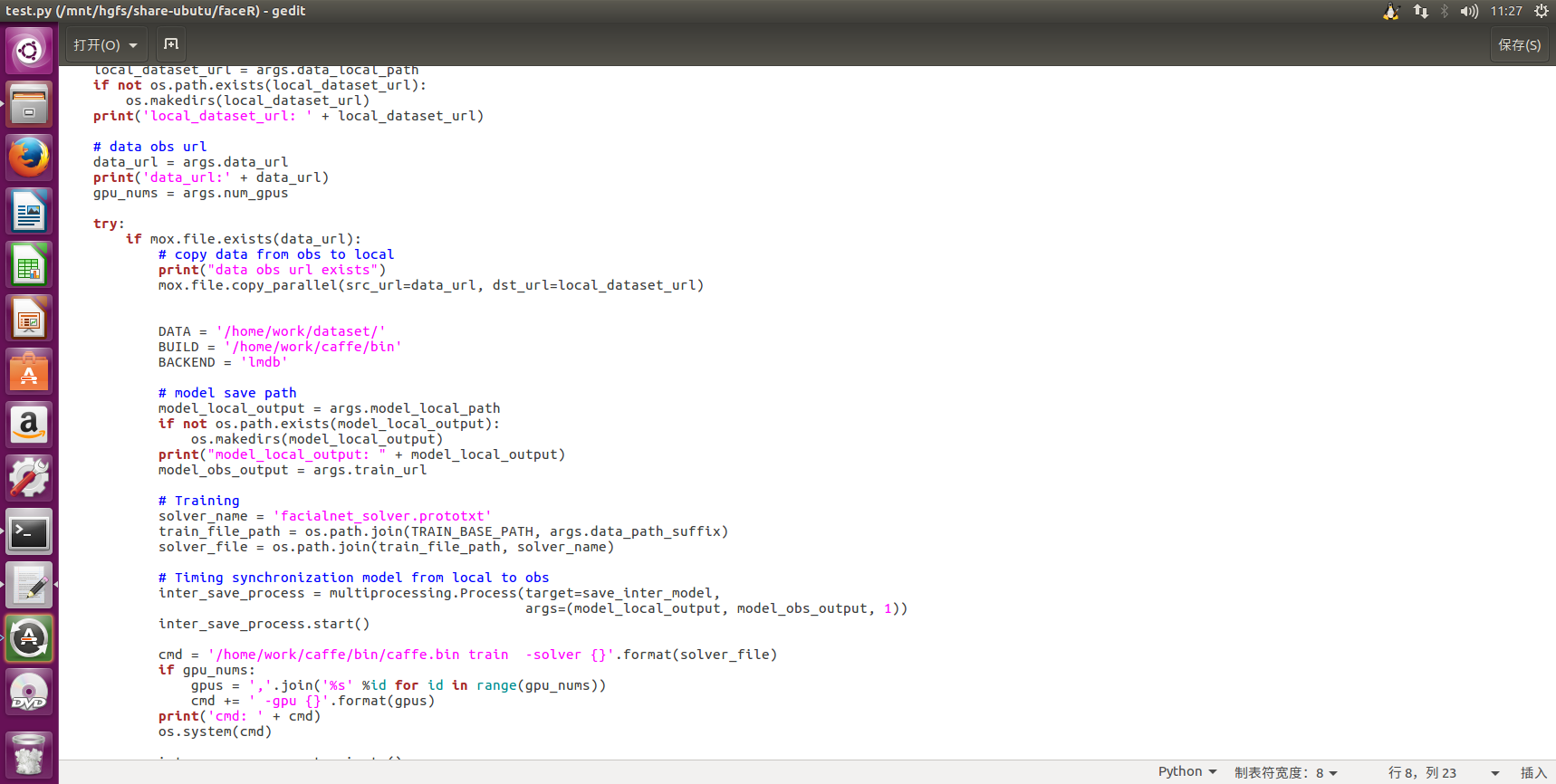


* facialnet\_solver.prototxt为配置训练参数的prototxt文件，文件内容如下图所示：



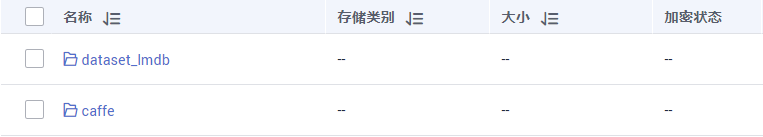
本实验设置训练参数为每迭代1000次进行一次测试，最大迭代次数为200,000。学习率随着迭代次数的增加而逐渐减小，每迭代50,000次，学习率的衰减值减小。

* train.py为训练脚本，脚本内容如下图所示：



注意data和相关配置文件的路径

1. 在obs桶上创建目录并上传相应所需文件，如下图所示：



dataset\_lmdb文件夹中包含文件如下：



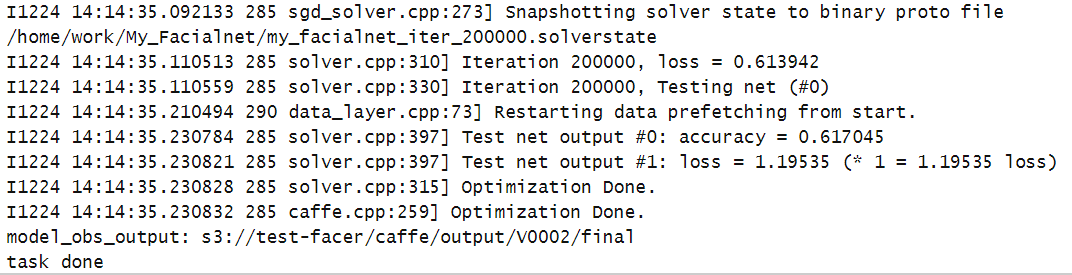
caffe文件夹中包含文件如下：



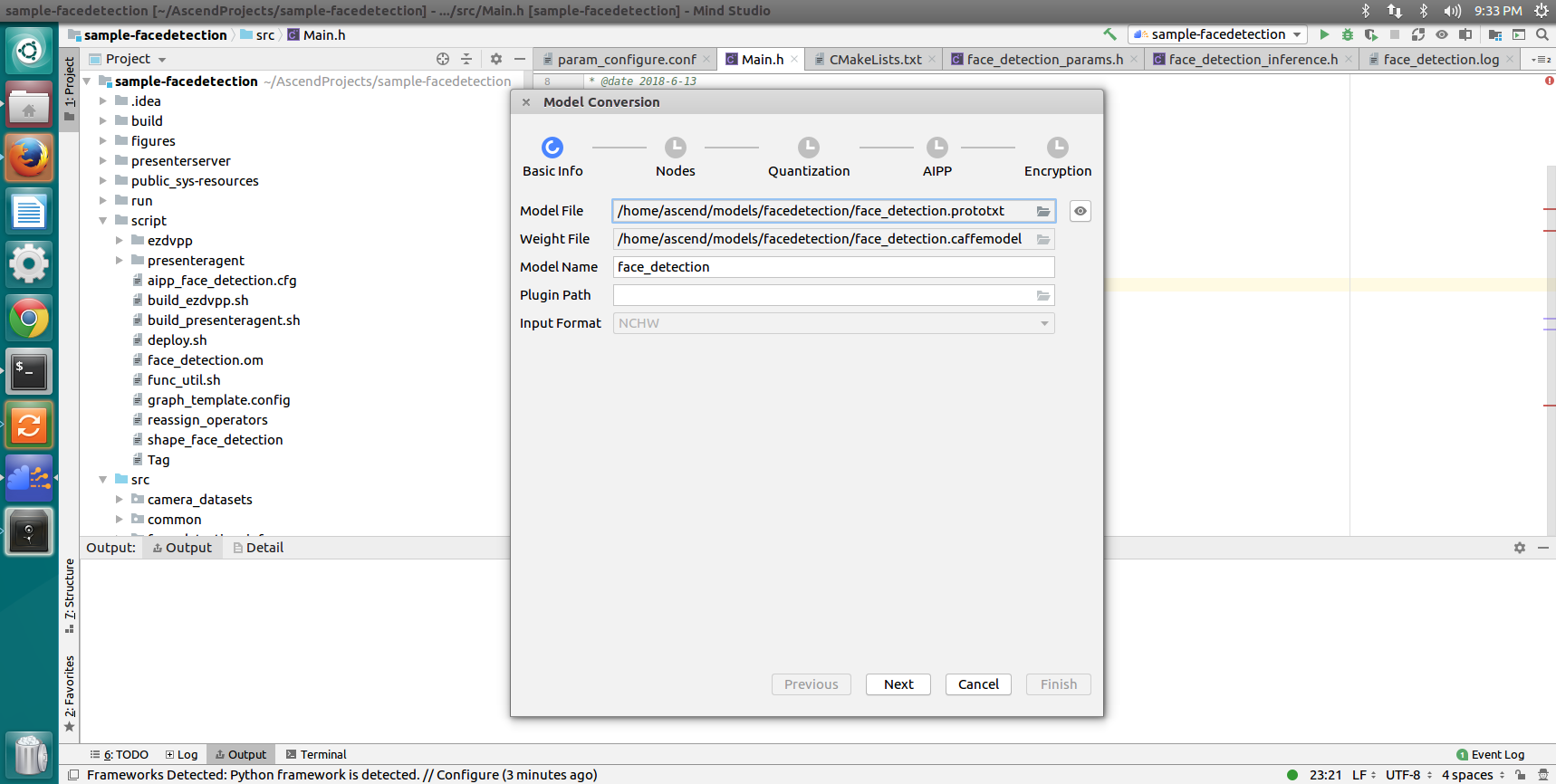
codes文件夹中包含文件如下：

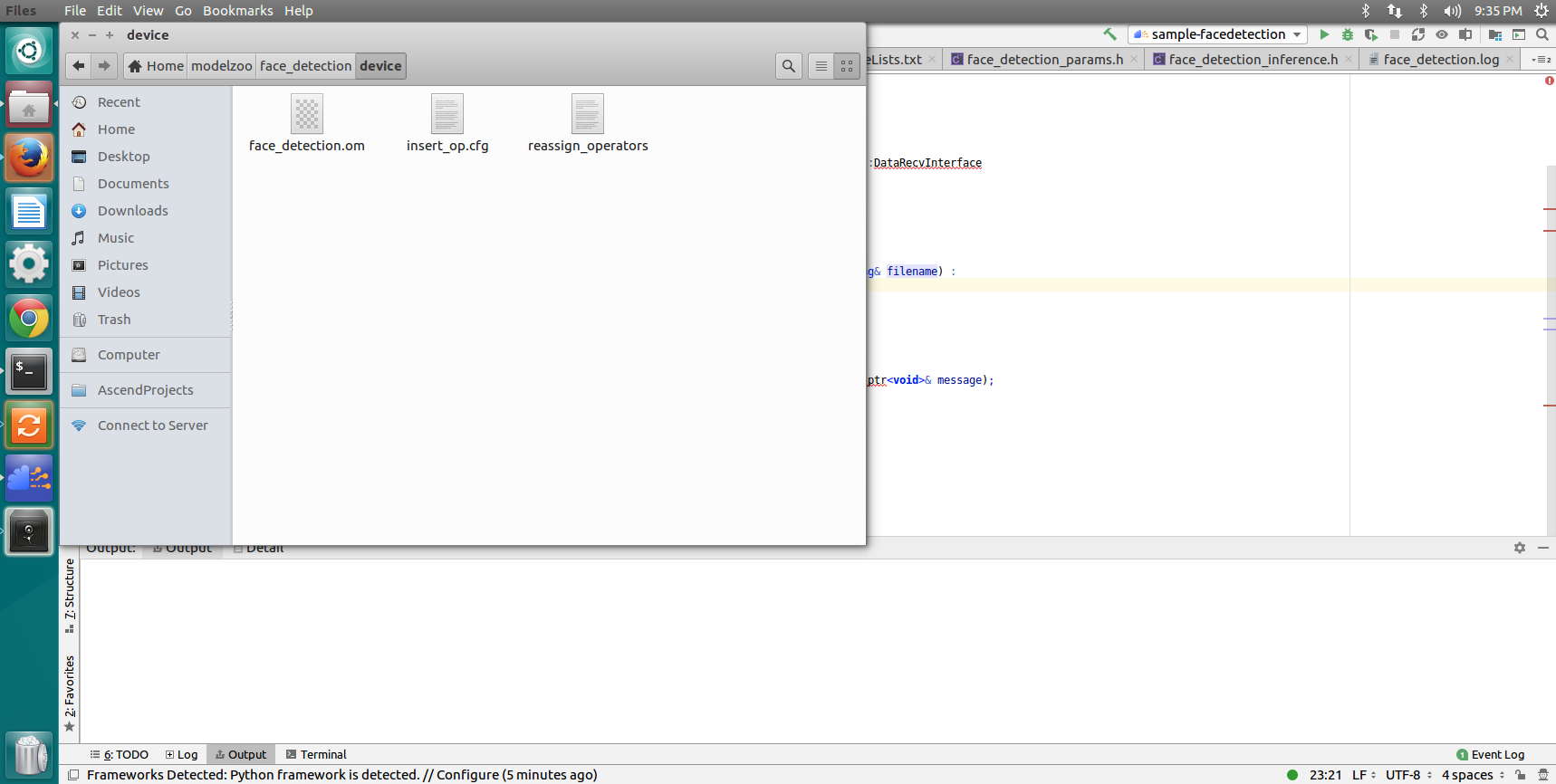


训练结果如下图所示：



得到对应的.caffemodel文件，利用Mind Studio工具将原始网络模型转换成适配昇腾AI处理器的.om模型，模型转换过程如下图所示：





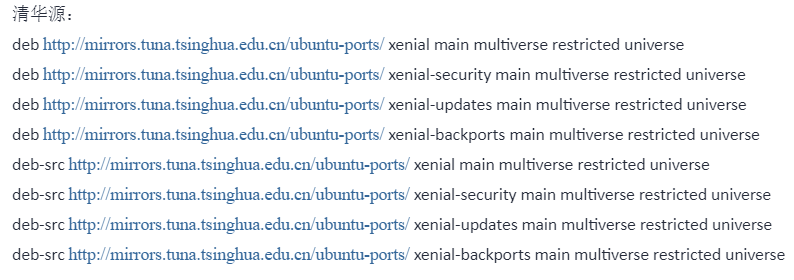
### 模型部署

1. 配置开发板联网：



（2）更新软件源及安装依赖

a. 编辑/etc/apt/sources.list内容替换为如下中科大的源或者清华的源，这里换的是清华源，之后更新：





1. 安装依赖和python库

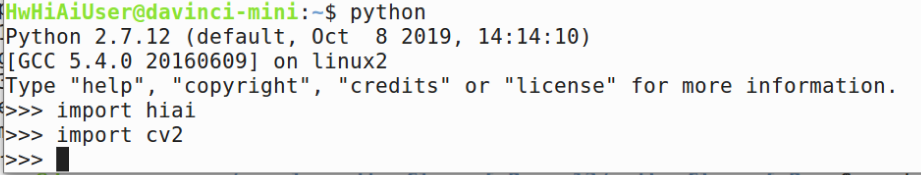




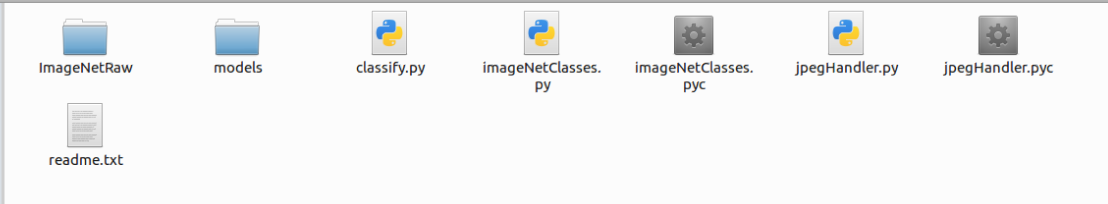
1. 在开发板上安装hiai库和opencv库

将python2\_hiai\_install.sh脚本拷贝到开发板上并执行，同时在开发板上执行

安装opencv，以下是安装成功的结果：



（3）编写表情识别的Python代码



其中ImageNetRaw： 存放输入图片

classify.py: 主程序

imageNetClasses.py: 7种表情分类标签

jpegHandler.py ： jpeg图片处理，如resize、文字标注等

models： 存放模型网络feature\_classify.om（由caffe模型转换得到）

feature\_classifyResult： 存放标注后的图片

源码见附件

1. 将python代码拷贝到开发板：

进入工程所在目录，执行命令：



1. 登陆开发板，进入该工程主程序classify.py所在目录，执行

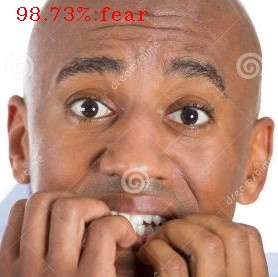
python classify.py运行



上图为运行成功截图

1. 执行以下命令将标注好的结果拷贝到UI Host端查看：



上图为输入图片经模型推理打标签生成的图像。

## 结果分析及总结

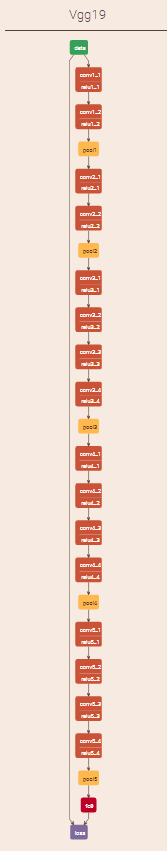
本次实验对传统的Alexnet网络结构进行了修改以更加适应对fer2013数据集的训练，通过对28709张图片进行20万次的迭代训练，得到的模型在验证集的精度在60-70%。并最终在开发板上部署运行，通过对随机的表情图像进行模型推理，得到推理后打上标签的图像。

针对整个实验过程中模型精度不高的问题，我们分析其原因有以下几点：

1. fer2013数据集主要是通过爬虫的方式从网上各大网站得到，其特征欠佳，并且很多图片的label存在错误，限制了模型精度的提高。
2. 我们改进后的 Alexnet网络结构的层数仍略少，对图像深层的特征提取不够。

关于以上两点，我们小组采用CK数据集也进行了模型的训练，但同样出现模型精度不高的问题，主要跟CK数据集的数据量过少有关（几百张左右）

关于改进后的Alexnet结构提取特征不够的问题，我们采用图像分类性能更加强悍的Vgg19网络并对其进行改进（删除两个全连接层）在fer2013上同样进行了训练，但仍达不到好的效果，其网络结构如下图所示：



分析原因，是因为fer2015输入的图像为48\*48，我们做了裁剪操作后得到42\*42，输入大小过小，Vgg的卷积层数较多，经过多层处理后特征图大小变成了1\*1，导致特征丢失。

综上所述，本次实验还有许多不足之处，针对表情识别的训练，应当寻找更加精确的数据集进行训练。我们组认为，可以把人脸表情和一些动画人物表情结合起来以扩充数据集，同时合理运用Vgg网络和Reswnet网络结构进行调整训练，以达到更好的结果。