

# 深度学习与计算机视觉作业

## ——表情识别

### 一、数据集分析

本实验使用的 CK+视频表情数据集，包含了 Surprise、Sad、Happy、Fear、Disgust、Contempt、Angry 七类表情数据，各类别的数据分布如下图 1。

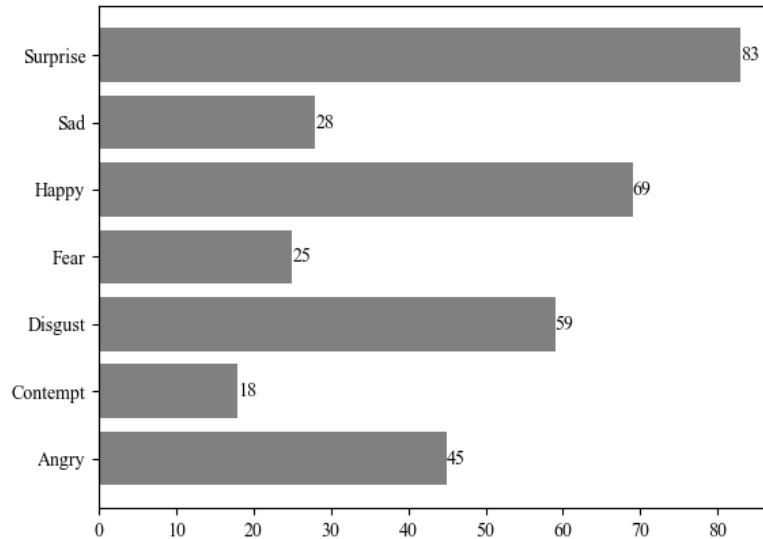


图 1 七类表情数据

该数据集共有 327 个视频片段，采样自 118 个实验对象，每个视频片段含义 5 张图片，每一帧图像的尺寸为  $720 \times 480$  或者  $640 \times 490$ ，通道数目为 1 或者 3，值域为  $[0, 255]$ ，同一视频片段中的 5 帧图像具有相同的尺寸和通道数目。先从数据集中随机选取 78 个片段作为验证集，余下的作为训练集。

### 二、数据预处理

2.1 统一将原始图像的扩展为 3 通道的 RGB 图像

2.2 将去掉验证集的数据集作为训练集，首先将训练集的是所有图像扩展为 RGB 的 3 通道图像，将原始图像整张输入，但是在输入后再原始图像上随机剪切  $256 \times 256$  大小的子图输入网络。同时也将验证集对图像调整到  $256 \times 256$ 。如图 2



a

b

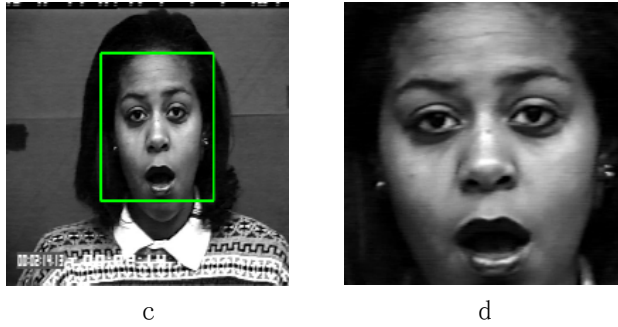


图 2 a、c 为截取前，b、d 为截取后

### 三、模型框架

#### 3.1 模型设计

- 1) 分别采用了 VGG19 来完成表情的识别与分类;
- 2) VGG19 的每一个小块是有一个卷积层, 一个 BatchNorm 层, 一个 ReLU 层和一个平均池化层来构成的;
- 3) 在全连接层之前加入了 dropout 策略, 增加了模型鲁棒性;
- 4) 去掉了传统 VGG19 中的多个全连接层, 直接在一个全连接层后直接分为 7 类去识别。

#### 3.2 损失函数

我们选择了交叉熵损失函数, 公式如下:

$$J(\theta) = -\frac{1}{m} \sum_{i=1}^m [y^i \log(h_{\theta}(x^i)) + (1 - y^i) \log(1 - h_{\theta}(x^i))]$$

#### 3.3 参数设置

Batchsize 为 32, learning\_rate 为 0.01, epoch 为 60。

### 四、实验结果及分析

利用验证集对训练好的模型进行测试, 结果如下表。

#### 4.1 混淆矩阵

测试的到的 image-level 混淆矩阵如表 1 所示。

	Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
Angry	18	1	18	0	7	5	0
Contempt	0	23	0	0	0	0	2
Disgust	10	0	60	0	0	2	5
Fear	2	0	0	10	15	0	4
Happy	0	0	0	6	68	0	0
Sad	9	8	2	3	2	19	2
Surprise	0	0	7	0	0	8	74

表 1 Image-level confusion matrix

video-level 的混淆矩阵如表 2 所示：

	Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
Angry	4	0	4	0	1	1	0
Contempt	0	4	0	0	0	0	0
Disgust	2	0	13	0	1	0	0
Fear	0	0	0	1	2	0	0
Happy	0	2	0	0	15	1	0
Sad	1	0	2	0	0	4	2
Surprise	0	0	0	0	0	2	16

表 2 Video-level confusion matrix

## 4.2 准确率

根据 image-level 的混淆矩阵，计算得到的 AP 和 CAP 如表 3、表 4：

Image AP	Mean CAP	Image cAP						
		Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
0.6667	0.7571	0.5000	0.8000	0.7222	1.0000	0.7778	0.5714	0.9285

表 3 Image\_AP

Video AP	Mean CAP	Video cAP						
		Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise
0.6974	0.6592	0.4615	0.7888	0.6896	0.5263	0.7391	0.5588	0.8505

表 4 Video\_AP

## 4.3 结果分析

1) Dropout 能有效得降低过拟合，提高准确率。Dropout 的方法相当于训练的时候随机失活掉一些连接，而在测试的时候把这些连接补充回来，这就相当于集成多个不错的模型来做综合的预测。

2) 可能由于某些参数不是很适合或着训练轮数太少，准确率没有达到用网上下载到的已经训好的 83%以上的准确率。