



## Facial Expression Summary

Jiaming Nie

June 19, 2018

## 1 选取网络以及相关参数

网络的选取和相关参数可见如下表格

Table 1: 选取的网络和相关参数

Name	Batch Size	Epoch	Learning Rate	Optimizer	L2 Regularizer
ResNet18	16	10	0.1	SGD	0.0001
ResNet34	16	30	0.1	SGD	0.0001
GoogleNet	16	8	0.1	SGD	0.0001

## 2 训练结果与BaseLine

### 2.1 ResNet18

对原有的ResNet 18网络添加了Batch Normalization层，在卷积添加了添加了Dropout层。

#### 2.1.1 ResNet18 准确率与损失曲线

训练集和验证集的准确率(Accuracy)和损失(Loss)曲线可见图~1,

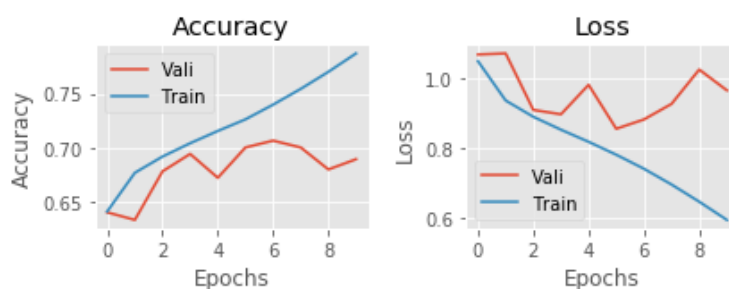


Figure 1: ResNet 18 的准确率与损失曲线

#### 2.1.2 ResNet 18在测试集上的表现

ResNet 18在测试集相关的结果如下:

Table 2: ResNet18 在测试集的表现

Accuracy	69.44%
Recall	0.6944
F1 Score	0.6800

#### 2.1.3 Confusion Matrix

ResNet18在测试集上的Confusion Matrix 如下:

Table 3: ResNet 18在测试集上的Confusion Matrix

	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
Neutral	<b>0.7815</b>	0.1086	0.0298	0.0312	0.0081	0.0068	0.0326	0.0013
Happy	0.0922	<b>0.8791</b>	0.0044	0.0111	0.0015	0.0037	0.0059	0.0014
Sad	0.4387	0.0592	<b>0.3478</b>	0.0276	0.0276	0.0276	0.0711	0
Surprise	0.4326	0.1063	0.0212	<b>0.2978</b>	0.099	0.0014	0.02836	0
Fear	0.3538	0.0307	0.0923	0.1231	<b>0.2769</b>	0.0461	0.0769	0
Disgust	0.2619	0.1190	0	0.0714	0	<b>0.3809</b>	0.1667	0
Anger	0.4521	0.0459	0.0498	0.0496	0.0191	0.0498	<b>0.3371</b>	0
Contempt	0.4339	0.4339	0.0377	0.0188	0	0.0186	0.0377	<b>0.0186</b>

## 2.2 ResNet 34

对原有的ResNet 34网络添加了Batch Normalization层，在卷积添加了添加了Dropout层。

### 2.2.1 ResNet34 准确率和损失曲线

训练集和验证集的准确率(Accuracy)和损失(Loss)曲线可见图~2，

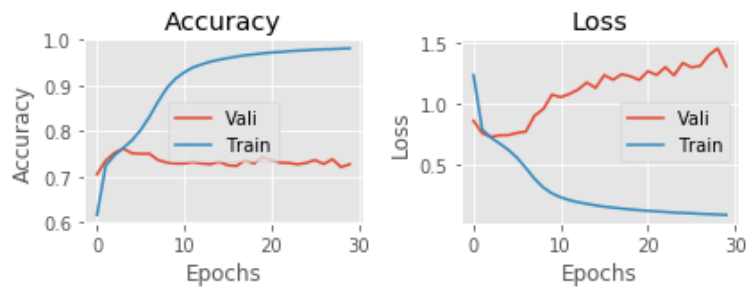


Figure 2: ResNet 34 的准确率与损失曲线

在ResNet34的训练结果中，过拟合的程度超过GoogleNet和ResNet18，后两者在训练集上无法得到超过80%的结果。

在最初几个Epoch以后验证集的准确率即趋于平稳。

### 2.2.2 ResNet34 在测试集的表现结果

ResNet34在测试集相关的数据如下：

Table 4: ResNet34 在测试集的表现

Accuracy	72.27%
Recall	0.7227
F1 Score	0.7177

### 2.2.3 Confusion Matrix

ResNet 34在测试集上Confusion Matrix如下:

Table 5: ResNet 34在测试集上的Confusion Matrix

	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
Neutral	<b>0.6807</b>	0.1426	0.0652	0.0353	0.0054	0.0013	0.0638	0.0054
Happy	0.0733	<b>0.8873</b>	0.0074	0.011	0	0	0.0074	0.0007
Sad	0.225	0.059	<b>0.577</b>	0.0237	0.0079	0.0079	0.0909	0.0079
Surprise	0.2553	0.1276	0.0496	<b>0.3972</b>	0.1134	0	0.0567	0
Fear	0.1846	0.0769	0.0461	0.1538	<b>0.4615</b>	0	0.0769	0
Disgust	0.1428	0.0952	0.0476	0.0476	0.0238	<b>0.3809</b>	0.2619	0
Anger	0.2452	0.0421	0.0689	0.0536	0.0191	0.0153	<b>0.5517</b>	0.0038
Contempt	0.3207	0.4905	0	0	0	0	0.0754	<b>0.1132</b>

## 2.3 GoogleNet

对原有的GoogleNet添加了Batch Normalization, L2正则化的系数设置为0.0001 ( $10^{-4}$ )。

### 2.3.1 GoogleNet 训练集与验证集准确率和损失曲线

训练集和验证集的准确率(Accuracy)和损失(Loss)曲线可见图~3,

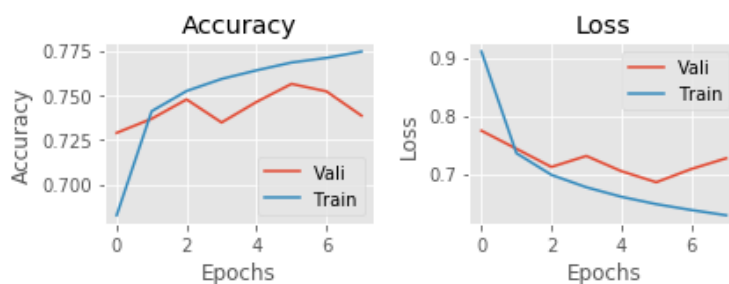


Figure 3: GoogleNet的准确率与损失曲线

在经历8个Epoch后, 验证集的准确率为73.84%, 损失为0.727。训练集的准确率为77.44%, 损失为0.6289。

### 2.3.2 GoogleNet在测试集的表现

在模型训练结束后, GoogleNet在测试集的表现如下:

Table 6: GoogleNet在测试集的表现

Accuracy	75.96%
Recall	0.7596
F1 Score	0.7465

### 2.3.3 测试集结果的Confusion Matrix

GoogleNet在测试集表现结果的Confusion Matrix:

Table 7: GoogleNet在测试集上的Confusion Matrix

	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger	Contempt
Neutral	<b>0.8059</b>	0.121	0.026	0.02	0	0.0013	0.0229	0.0013
Happy	0.059	<b>0.9186</b>	0.00295	0.011	0	0	0.0074	0.0007
Sad	0.346	0.06	<b>0.4489</b>	0.0039	0.0236	0.0157	0.094	0
Surprise	0.2535	0.1338	0.042	<b>0.45</b>	0.0633	0.007	0.0492	0
Fear	0.1384	0.0769	0.0923	0.1077	<b>0.4923</b>	0	0.0923	0
Disgust	0.2142	0.1667	0.0476	0	0.023	<b>0.3095</b>	0.2381	0
Anger	0.3168	0.0343	0.0458	0.0267	0.0038	0.0190	<b>0.5534</b>	0
Contempt	0.2264	0.6603	0	0	0	0	0.0377	<b>0.0754</b>

## 2.4 AffectNet Paper给出的confusion matrix

AffectNet给出了使用AlexNe训练结果给出的BaseLin以使用Microsoft Cognitive Service Emotion API预测得到的结果 [1].

### 2.4.1 使用标准的多分类Loss Function

AffectNet Paper中所给出的confusion matrix可见于下图4.

Prediction \ Ground truth	Neutral	Happiness	Sadness	Surprise	Fear	Disgust	Anger	Contempt
Neutral	<b>0.734</b>	0.15	0.05	0.02	0.012	0	0.034	0
Happiness	0.052	<b>0.944</b>	0	0.004	0	0	0	0
Sadness	0.384	0.094	<b>0.412</b>	0.014	0.028	0.002	0.066	0
Surprise	0.344	0.228	0.044	<b>0.252</b>	0.106	0.002	0.024	0
Fear	0.25	0.078	0.104	0.1	<b>0.362</b>	0.014	0.092	0
Disgust	0.252	0.166	0.106	0.03	0.016	<b>0.136</b>	0.294	0
Anger	0.356	0.068	0.062	0.016	0.024	0.016	<b>0.458</b>	0
Contempt	0.284	0.636	0.02	0.002	0.	0.008	0.048	<b>0.002</b>

Figure 4: AffectNet Confusion Matrix

### 2.4.2 使用加权损失函数

AffectNet Paper中所给出的confusion matrix可见于下图4.

加权损失函数提升了Sadness, Surprise, Fear, Disgust, Anger和Contempt的准确率, 但同时也降低了Neutral和Happy标签的准确率。

**TABLE 8**  
**Confusion Matrix of Weighted-Loss Approach on the Test Set**

		Predicted							
		NE	HA	SA	SU	FE	DI	AN	CO
Actual	NE	53.3	2.8	9.8	8.7	1.7	2.5	10.4	10.9
	HA	4.5	72.8	1.1	6.0	0.6	1.7	1.0	12.2
	SA	13.0	1.3	61.7	3.6	5.8	4.4	9.2	1.2
	SU	3.4	1.2	1.7	69.9	18.9	1.7	2.8	0.5
	FE	1.5	1.5	4.6	13.5	70.4	4.2	4.3	0.2
	DI	2.0	2.2	5.8	3.3	6.2	68.6	10.6	1.3
	AN	6.2	1.2	5.0	3.2	5.8	11.1	65.8	1.9
	CO	16.2	13.1	3.5	3.1	0.5	4.3	5.7	53.8

Figure 5: AffectNet Confusion Matrix 使用加权损失函数

### 2.4.3 AffectNet使用Microsoft Cognitive Service Emotion API

微软表情API在识别Neutral和Happy标签的图片上给出了较高的准确率，但对一些负面情绪的标签识别准确较低在。

具体为：

- Neutral: 94% (准确率过三种模型和BaseLine)
- Happy: 85 %
- Fear: 25 %
- Disgust: 27%
- Contempt: 4 %

## 3 一些API的调查

### 3.1 Microsoft Cognitive Services emotion API

一家名为SightHound的公司对微软的情绪识别API进行了测试 [2]，表情的分为7种，总体准确率为61.3%，Confusion matrix如下：

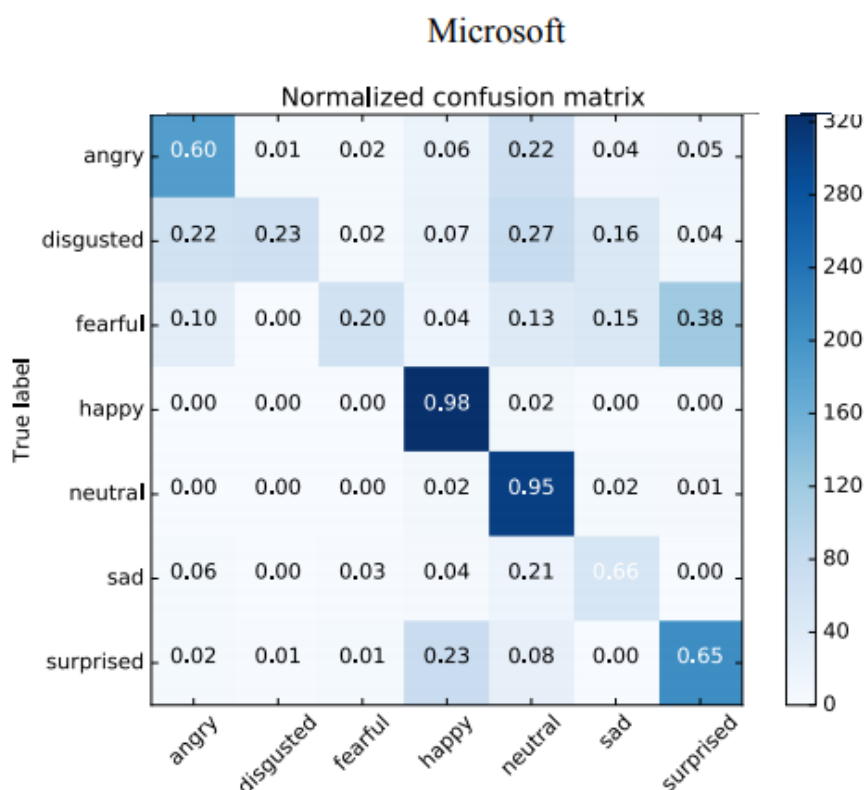


Figure 6: Sighthound使用MS API结果的 Confusion Matrix

### 3.1.1 EmotionNet

EmotionNet Challenge [3]是对一个包含95图片的数据库进行分类识别，2017年第一的队伍准确率为82%。

## 4 总结

### 4.1 网络深度问题

对于AffectNet的表情数据集，使用了三种网络，ResNet18, ResNet34 和 GoogleNet, 在训练一个Epoch之基本验证集的效以及已经达到最好的效果，以上三种网络结构较复杂，可采取较为简单的网络模型。

### 4.2 准确率问题

三种模型共同特点是特点是对Neutral和Happ标签的准确率较高，可达到80%和90%左右，但是对其他标签准确率较低。基础的8类标签中，只有surprise这个标签可以是负面也可以是正面表情，其他表情可以分为Positive, Negative和Neutral标签。一些解决方案可如下：

#### 4.2.1 先分三类，再分8类

可考虑先把标签分为正面(即Happy标签),中立(即Neutral) 以及负面(其余的5类)。训练时先将表情进行三分类，再对负面表情进行5分类。

**混合标签** 仅Surprise表情而言，此标签在描述人表情的时候一般和其它表情一同使用，可以考虑采用混合标签，比如惊讶且高兴，惊讶且难过等标注方式用于网络的训练。

#### 4.2.2 使用Arousal和Valence对表情进行回归

Arousal表示情绪的正负面，Valence表示情绪的激烈程度，区间均为 $[-1,1]$ 。AffectNe已经给出了所有图片相关的数值，可以使用神经网络解决这个回归问题。

#### 4.2.3 Action Unit 面部表情活动活动单元

其他的数据集，如EmotionNet [3]，是标注表情的AU进而对表情进行分类。

EmotionNet [4]本身也给出了一种计算AU的方式。

## References

- [1] A. Mollahosseini, B. Hasani and M. H. Mahoor, “Affectnet: A database for facial expression, valence, and arousal computing in the wild,” <https://arxiv.org/pdf/1708.03985.pdf>, 2007.
- [2] A. Dehghan, E. G. Ortiz, G. Shu, S. Z. Masood, “Dager: Deep age, gender and emotion recognition using convolutional neural networks,” Computer Vision Lab, Sighthound Inc., Winter Park, FL.
- [3] EmotionNet , <http://cbcs1.ece.ohio-state.edu/EmotionNetChallenge/index.html#overview>.
- [4] C. F. Benitez-Quiroz, R. Srinivasan, A. M. Martinez , “Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild,” <http://cbcs1.ece.ohio-state.edu/cvpr16.pdf>, 2016.