



AffectNet Dataset Description and Training Result on fer2013
AffectNet数据集的具体说明和几种深层网络模型训练结果

Jiaming Nie

2018 年 5 月 18 日

1 AffectNet 基本信息

AffectNet的标签共有8种表情，以及3种非表情标签(不需要)，具体信息可见于表 1。

表 1: AffectNet数据集标签

Category	分类	数量
Neutral	中立	80,276
Happy	高兴	146,198
Sad	悲伤	29,487
Surprise	惊讶	16,288
Fear	害怕	8,191
Disgust	厌恶	5,264
Anger	愤怒	28,130
Contempt	轻蔑/蔑视	5,135
None	非表情	35,322
Uncertain	不确定	13,163
Non-Face	非人脸(真实人脸)	88,895

11种分类共计45,6349个样本，所需7种表情(去除三种不明确的表情标签以及轻蔑(contempt)的表情标签)剩余样本总量为31,3834。

2 EmotionNet 算法

EmotionNet是一种基于人脸面部肌肉动作单位(Action Units)的数量和强度来对人脸表情进行标记的一种算法。

EmotionNet算法对特定表情的人脸图片进行特征提取并进行分类，输出表情的分类。其分类标签包括特定的表情标记(如高兴，悲伤等)，也包括以面部表情的动作单位(AU,Action Units)进行分类。EmotionNet算法实现了实时处理图片，相对于人工标注提高了运算效率。

2.1 特征空间 Feature Space

EmotionNet对人脸图片进行特征空间的提取，并进行分类。Feature Space的提取通过对人脸面部图片进行德劳内三角化处理(Delaunay triangulation),得到两点之间的距离以及形成三角形的角度，再组成 Feature Space。两点间的距离经过标准化处理(Normalization)。

德劳内三角化处理可见于下图 1。

图 1 所示的图片德劳内三角化处理，标记点的个数为66个，所形成的三角形个数为107个，共有321个角。

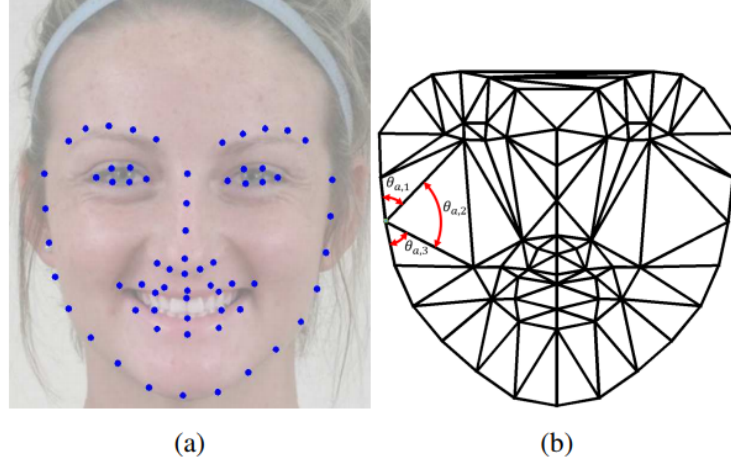


图 1: 德劳内三角化处理

在图 1 所示的66个点中，可分为两类:

1. 解剖学意义上的标注点(眼睛，嘴巴，嘴唇，鼻子和下巴)(共有15个)
2. 非解剖学意义标注点，这些点构成了人面部基础部分的边界，是常数，决定了人面部特征的相对位置。(原文: defining the contour of each facial component (e.g., brows) is constant.)

2.2 Formation of Feature Space

对于特定的动作单位(Action Unit) AU_i ，定义每个二维图像上的点的坐标组成的向量 s_{ij} ，有 $s_{ij} = (s_{ij1}^T, \dots, s_{ijp}^T)^T$ 。其中 s_{ijk} 是 AU_i 上第 k 个标记点的二维坐标。

对于每个特定的 AU_i ，则有132个坐标对，即 $s_{ij} \in 132$ 。

对于每张用于训练算法的图片，距离都被标准化处理为 τ 个像素点(pixels)。标准化处理后的 $\hat{s}_{ij} = cs_{ij}$ ，其中 $c = \tau / \|l - r\|_2$ ， l 与 r 分别是左眼与右眼中心点的坐标。

由此可得出特征空间的表达式:

$$x_{ij} = (d_{ij12}, \dots, d_{ijp-1p}, \theta_1^T, \dots, \theta_p^T)^T \quad (1)$$

在公式 1 中， $d_{ij12} = \|\hat{s}_{ija} - \hat{s}_{ijb}\|_2$ 是两点间标准化处理后的欧式距离(Euclidean Distance)。 $\theta_a = (\theta_{a1}, \dots, \theta_{aqa})^T$ 是德劳内三角化后的角度向量。

特征向量的大小为 $x_{ij} \in \mathbb{R}^{p(p-1)/2+3t}$ 。当 $p = 66, t = 107$ 可得到

$$x_{ij} \in \mathbb{R}^{2466} \quad (2)$$

2.3 Gabor Filter

Gabor滤波器用于提取空间局部频度特征，是一种有效的纹理检测工具。

2.3.1 Gabor Filter Equation

$$g(\hat{s}_{ijk}; \lambda, \alpha, \phi, \gamma) = \exp\left(-\frac{s_1^2 + \gamma^2 s_2^2}{2\sigma^2}\right) \cos\left(2\pi \frac{s_1}{\lambda} + \phi\right) \quad (3)$$

各种符号的标记如下:

- λ - 波长
- α - 方向
- ϕ - 角度
- γ - 空间角度比例
- σ - 滤波器的范围

对于一个特定的 AU_i , Gabor 滤波器变换后的结果为:

$$g_{ij} = (g_{ij1}^T, \dots, g_{ijp}^T)^T \quad (4)$$

2.4 Final Feature Vector

那么Feature Vector的表达式为:

$$z_{ij} = (x_{ij}^T g_{ij}^T)^T \quad (5)$$

2.5 分类(Classification in Free Space)

2.5.1 Kernel Subclass Discriminant Analysis (KSDA)

对于训练集的分类是基于Kernel Subclass Discriminant Analysis, KSDA 是一种基于贝叶斯网络的线性分类器，基于两条标准对已经有的特征向量进行分类。

$$Q_{i1}(\phi_i, h_{i1}, h_{i2}) = \frac{1}{h_{i1}h_{i2}} \sum_{c=1}^{h_{i1}} \sum_{d=h_{i1}}^{h_{i1}+h_{i2}} \frac{\text{tr}(\int_{iC}^{\phi_i} \int_{id}^{\phi_i})}{\text{tr}(\int_{iC}^{\phi_i^2}) \text{tr}(\int_{id}^{\phi_i^2})} \quad (6)$$

第二条标准:

$$Q_{i1}(\phi_i, h_{i1}, h_{i2}) = \sum_{c=1}^{h_{i1}} \sum_{d=h_{i1}+1}^{h_{i1}+h_{i2}} p_{ic} p_{id} \left\| \mu_{ic}^{\phi_i} - \mu_{id}^{\phi_i} \right\|_2^2 \quad (7)$$

3 Fer2013 Dataset

3.1 基础信息

Fer2013数据集中的图片均为 48×48 的灰度图片，表情的分类为7种，标记信息可见表 2:

表 2: Fer2013

Category	分类	标记
Angry	愤怒	0
Disgust	厌恶	1
Fear	恐惧	2
Happy	高兴	3
Sad	伤心	4
Surprise	惊讶	5
Neutral	中立	6

3.2 训练结果

这里的训练结果是用一个简单的三层网络的CNN，基于keras实现。

训练中每个epoch的loss和accuracy如下(蓝线是training 红线是test):

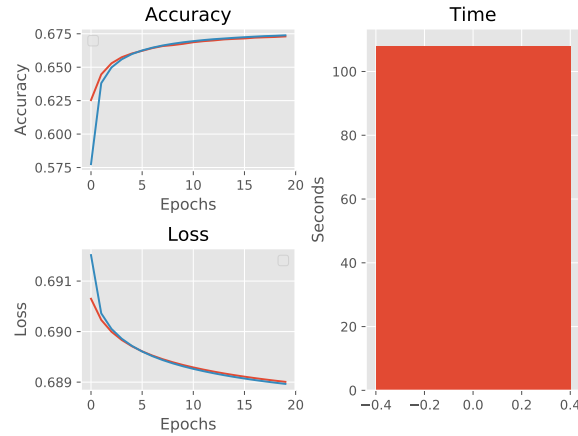


图 2: 简单CNN训练结果