

# 可靠的众包和深度保位学习的表达方式 野生动物中的认可

李山, 邓伟红, 杜俊平 北京邮电大学

fls1995, whDeng, junpingdg@bupt.edu.cn

## 摘要

过去对面部表情的研究使用了相对有限的数据集，这使得目前的方法是否能在现实世界中使用还不清楚。在这篇论文中，我们提出了一个新的数据库RAF-DB，它包含了来自数千人的大约30000张面部图像。每张图片都被单独标注了40次，然后用EM算法过滤掉不可靠的标注。

s. 众包显示，现实世界中的人脸经常表现出复合情绪，甚至是混合情绪。就我们所知，RAF-DB是第一个包含野外复合表情数据库。我们的跨数据库研究表明，RAF-DB中的基本情绪的动作单位比实验室控制的情绪的动作单位更加多样，甚至偏离。为了解决这个问题，我们提出了一种新的DLP-CNN（深度位置保护CNN）方法，该方法旨在通过保护位置的接近性，同时最大限度地提高类间散射来增强深度特征的判别能力。在7类基本表情和11类复合表情上的基准实验，以及在SFEW和CK+数据库上的额外实验表明，所提出的DLP-CNN在野外的表情识别中优于最先进的手工特征和基于深度学习的方法。

## 1. 简介

每天都有数以百万计的图片被用户从不同的活动和社交聚会中上传。人们对设计能够了解人类情感属性和情感表现的系统越来越感兴趣。为了从互联网上自动学习人脸图像的情感状态，需要大型的注释数据库

。然而，情感类别注释的复杂性阻碍了大型注释数据库的收集。在另一方面，流行的AU编码<sup>12</sup>需要特定的专业知识，需要花几个月的时间来学习和掌握。完善，因此，需要替代的解决方案。而且由于

由于对面部表情的感知方式存在文化差异，心理学家很难为每个面部表情定义明确的原型单位。因此，除了少数专家的专业知识外，从大量普通人的判断中研究社会形象的情感也是值得的。

在本文中，我们提议通过一种可靠的众包方法来研究普通的外貌感知。具体来说，我们训练有素的注释者被要求给人脸图像贴上七个基本类别之一的标签<sup>11</sup>，并且每张脸都被独立注释足够多次，即在我们的实验中大约40次。然后，嘈杂的标签被一个基于EM的可靠性评估算法过滤，通过这个算法，每张图片都可以被一个7维的情感概率向量可靠地表示。通过分析从互联网上下载的29672张种类繁多的面部图像的120万个标签，这些真实世界情感脸谱（RAF）<sup>1</sup>

，自然地分为两种类型：单模态分布的基本情感和双模态分布的复合情感，这一观察支持最近在实验室对照条件下的突破性发现<sup>10</sup>。据我们所知，真实世界的表达数据库RAF-DB是第一个大规模的数据库，它提供了在无约束环境下常见的表达感知和复合情绪的标签。

RAF-DB上的跨数据库实验和AU分析表明，现实世界表情的AU比心理学家指导的实验室控制的表情要多样化得多，甚至偏离了。为了解决这种不受约束的情感的模糊性，我们进一步提出了一种新的深度位置保护CNN（DLP-CNN）。受<sup>17</sup>的启发，我们开发了一个实用的反向传播算法，该算法创建了一个位置保护损失（LP损失），旨在将同一类别的本地相邻的脸拉到一起。与经典的softmax损失联合训练，迫使不同的类保持分离，位置性保留损失驱动每个类内的本地集群

<sup>1</sup> [http://whdeng.cn/RAF\\_model1.html](http://whdeng.cn/RAF_model1.html)



表情信息要么被分类为一组面部动作，要么被分类为一种特定的基本情绪[34]。大多数人关注的是后者，并以Ekman的六种基本情绪理论为基础[12]。事实上，如果不对如何确定哪些动作单位构成一个表情进行额外的假设，就不可能对表情类别进行准确的定义。因此，基本情绪表达并不具有普遍性，不足以概括人类面部显示的表情[37]。

### 2.3.用于表情识别的深度学习

最近，深度学习算法已被应用于视觉物体识别、人脸识别和检测、图像分类和许多其他问题，并取得了最先进的结果。到目前为止，由于缺乏足够的训练样本，只有少数深度神经网络被用于面部表情识别。在ICML 2013比赛中[16]，获胜者[41]是基于深度卷积神经网络（DCNN）加SVM。在2013年的EmotiW竞赛[6]中，获胜者[19]结合了模式规格的深度神经网络模型。在EmotiW 2015[8]中，更多的竞争者尝试了深度学习方法：[32]中使用转移学习来解决小数据库的问题，[20]中多列DCNN的分层委员会在SFEW数据库中获得了最好的结果，[22]中提出了LBP特征与DCNN结构相结合。在[24]中，提出了AU-aware Deep Networks (AUDN)来学习带有面部AU解释的特征。在[31]中，提出了一个具有初始层的DCNN，以获得类似的结果。

## 3. 真实世界表达数据库。RAF-DB

### 3.1.创建RAF-DB

数据收集。一开始，从Flickr收集的图像的URL被送入一个自动的开源下载器，以批量下载图片。考虑到Flickr的图像搜索API返回的结果是结构良好的XML格式，其中的URL可以很容易地被解析，然后我们使用一组关键词（例如：微笑、傻笑、哭泣、愤怒、害怕、恐惧、恐怖、震惊、惊讶、厌恶、无表情）来挑选出与六种基本情绪和中性情绪有关的图像。最后，一个以

数据库。近30000表达图像进行注释是一项极其困难和耗

时的任务。考虑到现实世界表达，图像的表达状态的多个视图

应该从不同的标注者那里收集。因此，我们雇用了315名注释者（来自大学的学生和工作人员）。在大学里接受过一小时辅导课程的学生，可以通过以下方式进行学习

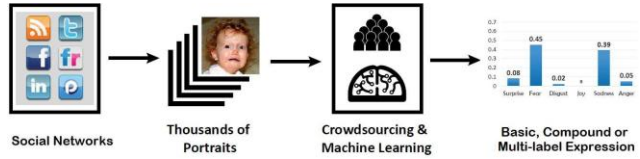


图2.RAF-DB的构建和注释概述。

在一个在线的面部表情注释任务中，他们被要求将图像归入七个类别中最明显的一个。我们开发了一个RAF-DB注释的网站，它显示了每张图像的专属属性选项。图像被随机地、平等地分配给每个标注者，确保由一个人标注的图像之间没有直接关联。每个图像都被保证由大约40个独立的标注者标注。之后，为每张图片得到一个多标签的注释结果，即一个七维向量，每个维度表示相关情感的投票。

元数据。数据提供了脸部区域的精确位置和大小，以及脸部手动定位的静脉标志点（两眼中央、鼻尖和两个嘴角）。此外，还包括一个无需人工标注的自动地标标注模式。从Face++ API[18]提供的注释结果中挑选出37个地标点。我们还手动注释了所有RAF脸的基本属性（性别、年龄（5个范围）和种族）。总之，我们数据库中的受试者年龄从0到70岁不等。他们中52%是女性，43%是男性，5%是不确定的。对于种族分布，有77%的高加索人，8%的非裔美国人和15%的亚洲人。每张图片的姿势，包括俯仰、偏航和滚动参数，都是由人工标注的LO-L-L计算出来的。

脸部地标的变化。

可靠性评估。由于标注者的主观性和不同的专业技能，以及图像难度的广泛性，标注者之间存在着一些不同的意见。为了摆脱嘈杂的标签，在[45]的激励下，使用了期望最大化（EM）框架来评估每个标注者的可靠性。

让  $D = \{(x, y, t^1, t^2, \dots, t^7)\}_{j=1}^n$  是  $t$  的复合属性  $t_j \in \{1, 2, 3, 4, 5, 6, 7\}$  是第一页  $nn$   $t$   $th$

基金会的网站  $a_{-1}$  的复合属性  $emulatedas a si m id uncti n$

$p(t_j = y_j | a_i, \beta_j) = (1 + \exp(-a_i \beta_j))^{-1/\beta_j}$  是

困难  $jtht$   $g$   $a_i$   $is t e elia ilitith$

$nn t t$   $gi$   $tpti$   $ithg ik$   $ihth gi n$



图3.来自RAF-DB的六类基本情感和十二类复合情感的例子。RAF-DB的详细数据分布已经附在每个表达类别上。

标签

$$\begin{aligned}
 \text{最大}_{\beta > 0} l(\alpha, \beta) &= \ln p(t|\alpha, \beta) = \ln p(t, y|\alpha, \beta) \\
 &= \ln Q(y) \frac{p(t, y|\alpha, \beta)}{Q_j(y)} \\
 &\geq Q(y) \ln \frac{p(t, y|\alpha, \beta)}{Q_j(y)} \\
 Q_j(y) &= \frac{p(t_j, y_j|\alpha, \beta)}{p(t_j|\alpha, \beta)} = p(y|t_j, \alpha, \beta)
 \end{aligned}$$

AAAA

nann

alalanman

$$anG_j = \{g_1, g_2, \dots, g_7\}$$

j的

高

$\alpha_i$

x

j yn

$$\frac{t_{ij}\alpha_i\beta_j}{j_{ij}}$$

$g_k = \alpha_i \mathbf{1}_{j \neq k} \alpha_i$  指示性指标。

W

ti  $\beta_{ij}$

n

ia

在

t

t

的作用是  
 $k \in \{1, 2, 3, 4, 5, 6, 7\}$

瘦肉精  
什么

t i antd la el

$\beta_j n \alpha_j$

itii

AAA

AA

基金会

{{

×

$$gmean = \frac{1}{7} \sum_{k=1}^7 g_k$$

罐头

tnt

tt

iyyi

担任过的职务有：副局长、副市长、副市长、副局长、副局长、副局长、副局长、副局长、副局长、副局长、副局长、副局长、副局长、副局长、副局长等。

分级，并

在此基础

启用

y

int

上

一个"U"字形，即"U"字形。

AA

AA

tininitin

残疾人的生活和工作是相似的

e nl 价值

经验是在

et t

Rk n x

ut ima es 它的中立性

i

e pa titi n

i u en ext

i

t

t

其特殊性

样品类和非特殊性样品

类和

非

特殊性

课

堂上，学生们都很认真地

听讲

和RAssataasetudt iitntti n

我们已经对RA的数据进行了分析，并进行了评估

在此，我想说的。

"我们要做的是，把我们的工作做好，把我们的





们的工作是：

$i k = n_c - 1$  ( $n_c$  is the number of elements in the class  $C_k$ )  
在C类T 基金会上，有一项是"不合格，即"不合格的"。  
抽取的样本 中，有一项是"不合格，即"不合格的"。  
在此， 我是："是"  
。终端的距离是多米诺骨牌  
我们将在此基础上，为您的工作提供支持。是  $s_{tmax}$ 。  
国际刑警组织和国际刑警组织 都有一个共同的目标，那就是  
esein l ssic c a acte i es t e l cal scatte s }it in  
级别 ttain te NNs 辨识度：在此过程中，我们可以看到有很多人都在使用"NN"这个词，而不是"NN"。  
}



表3.基于LBP、HOG和Gabor描述符以及SVM、LDA+kNN分类，CK+、JAFfE和RAF的基本表达类性能比较，以及RAF的复合表达性能。该指标是混淆矩阵的平均对角线值。

		基本 的				复式
		CK+	AAA	皇家 空军	皇家 空军	
姆萨维 姆	LBP	88.92	78.81	55.98		28.84
	HOG	90.50	84.76	58.45		33.65
	伽博	91.98	88.95	65.12		35.76
LDA	LBP	85.84	77.74	50.97		22.89
	HOG	91.77	80.12	51.36		24.01
	伽博	92.33	83.45	56.93		23.81

40个Gabor lters在不同的空间尺度和8个方向----。  
s.降样图像的大小被设定为10\*10，产生了4000个维度的特征。

为了客观地衡量追随者条目的性能，我们用ve-fold交叉验证的思想将数据集分成训练集和测试集，这意味着训练集的大小是测试集的ve倍，而且两集中的表达式具有近乎相同的分布。考虑到野外的表达式具有不平衡的分布，在RAF中不再使用对偏差特别敏感、对不平衡数据不再有效的准确性指标[15]。相反，我们使用混淆矩阵的平均对角线值作为最后的衡量标准。

基本情绪 .在这个实验  
中，使用单标签子集的全部1533  
9张图像检测了七个基本情绪类别。LBP的最佳分类精度（由SVM输出）为72.71%，HOG为74.35%，而Gabor为77.28%。当使用混淆矩阵的平均对角线值作为衡量标准时，结果分别下降到55.98%、58.45%和65.12%。为了评估基本情感标签的可靠性，我们还为每个样本分配了一个统一的随机标签，我们称之为天真情感检测器。而在使用Gabor特征时，天真分类器的最佳结果是16.07%，远远低于前者的数值。

为了进行比较，我们在CK+和JAFfE上采用了相同的方法，其中CK+采用了与人无关的5倍交叉验证法，JAFfE采用了离开一个对象的策略。表3中显示的结果证明，现实世界中的表情更难识别，目前在现有数据库中表现良好的普通方法不能解决现实世界中的表情识别问题。

为了评估不同分类器的有效性，我们还将LDA与最近的邻居（NN）分类进行了训练。我们发现，LDA+NN明显不如mSVM。

当在RAF这个极其庞大的数据库上进行训练时，它的表现要好得多。然而，在小规模的数据集（CK+和JAFfE）上训练时，它的表现更好，甚至在某些情况下超过了mSVM。具体的结果可参见表3。

复合情绪。  
对于复合情绪的分类，我们去掉了恐惧厌恶的情绪，因为它太少了，剩下11个复合情绪的类别，总共3954个。LBP的最佳分类精度（由SVM输出）是45.51%，HOG是51.89%，Gabor是53.54%。当使用混淆矩阵的平均对角线值作为衡量标准时，结果分别下降到28.84%、33.65%和35.76%。同样，为了证明复合情感标签的可靠性，我们计算了天真的情感检测器的基线，当使用Gabor特征时，结果下降到5.79%。

正如预期的那样，当涉及到更多的表达方式进行分类时，整体性能明显下降。与基本情感相比，明显较低的结果表明，复合情感更难检测，应该发明新的方法来解决这个问题。除了多模态之外，缺乏来自真实世界的复合表情的训练样本是另一个巨大的技术挑战。

6. 深度学习系统

如今，深度学习已经被应用于大规模的视觉识别任务，并在大量的训练数据下表现得异常出色。然而，由于模型学习的训练样本不足，完全监督的深度模型在面部表情识别任务中很容易被过度使用。因此，大多数用于面部表情识别的深度学习框架[22, 32, 36]是基于预训练的模型。这些预训练的模型，如VGG网络[40]和AlexNet[21]，最初是为人脸识别设计的，缺乏对表情特征的消除能力。因此，在本文中，我们直接在足够大的自我收集的数据库RAF上训练我们的深度学习系统，而不使用其他数据库。

在进行实验时，我们遵循与基线系统相同的数据集划分标准、图像处理方法和分类方法。相关研究[9, 39]已经证明，训练有素的深度卷积网络可以作为分类任务的特征提取工具，具有泛化能力。按照这个思路，我们首先对每个DCNN进行了基本情感识别任务的训练，然后直接使用已经训练好的DCNN模型来提取基本表情和复合表情的深度特征。从原始数据中学习到的2000维深度特征被从DCNN的全连接层提取出来，然后进行分类。

解释为SVM。

表4.不同DCNN在RAF上的表情识别性能。衡量标准是混淆矩阵的平均对角线值。

		基本化合								
		愤怒	厌恶	恐惧	幸福	悲伤	惊喜	中性	平均值	平均值
姆萨维姆	VGG	68.52	27.50	35.13	85.32	64.85	66.32	59.88	58.22	31.63
	淘宝网	58.64	21.87	39.19	86.16	60.88	62.31	60.15	55.60	28.22
	基础DCNN	70.99	52.50	50.00	92.91	77.82	79.64	83.09	72.42	40.17
	中心损失	68.52	53.13	54.05	93.08	78.45	79.63	83.24	72.87	39.97
	DLP-CNN	71.60	52.15	62.16	92.83	80.13	81.16	80.29	74.20	44.55
LDA	VGG	66.05	25.00	37.84	73.08	51.46	53.49	47.21	50.59	16.27
	淘宝网	43.83	27.50	37.84	75.78	39.33	61.70	48.53	47.79	15.56
	基础DCNN	66.05	47.50	51.35	89.45	74.27	76.90	77.50	69.00	28.23
	中心损失	64.81	49.38	54.05	92.41	74.90	76.29	77.21	69.86	27.33
	DLP-CNN	77.51	55.41	52.50	90.21	73.64	74.07	73.53	70.98	32.29

从表4的结果来看，我们有以下观察。首先，在大规模图像识别环境中取得相当合理结果的DCNs，如VGG

网络和AlexNet，对面部表情识别并不有效。其次，在RAF-DB上学习的所有深度特征都比基线系统中使用的未学习的特征要好得多，这表明深度学习架构更加稳健，适用于基本和复合表情。最后，我们的新的本地化保值损失模型比基于中心损失的模型取得了更好的性能。需要注意的是，中心损失可以有效地收敛单模态类，有助于提高网络在基本情感上的性能，但在复合情感上则失败。这表明，在多模态的面部表情识别中，包括基本表情和复合表情，保全位置的损失是有优势的。

为了了解我们训练有素的DLP-CNN模型在其他数据库中的泛化能力，我们随后采用它来对CK+和SFEW 2.0进行无调谐固定长度特征提取。对于实验室控制的数据库CK+，我们遵循基线系统的实验原则。对于现实世界的数据库SFEW 2.0，我们遵循EmotiW 2015[8]中的规则，SFEW best是EmotiW 2015的冠军[20]中使用的单一最佳模型的结果。请注意，在[20]中，作者用SFEW的额外数据训练他们的模型。从表5的比较结果中，我们可以看到，我们的网络也可以达到与其他最先进的方法相当甚至更好的性能，不仅是RAF，而且还有其它数据库。这表明，我们提出的网络可以作为一个高效的特征提取工具，用于面部表情数据库，而不需要在传统的DCNNs中花费大量时间来执行。

7. 结论和未来工作

本文的主要贡献是提出了一个新的众包优化算法和一个新的本地-----。

表5.DLP-CNN和其他最先进的深度学习方法在CK+和SFEW 2.0

上的比较结果。				
CK+	AUDN FP+SAE	SFEW	best DLP-CNN	
93.1	91.1	93.2	95.78	
[27]	[27]	[20]	(不含 netune)	
SFEW 2.0	30.14	47.7	52.5	51.05

基于真实世界公开的面部表情数据库RAF-DB，为深度学习保留了损失层。优化的算法有助于保持来自标签者的最佳注释结果。新的DCNN可以为表情识别任务学习更多的鉴别性特征。RAF-DB包含：1) 29672张真实世界的图像，标注了不同的表情、年龄范围、性别和姿势；2) 每张图像的7维表情分布向量；3) 两个不同的子集：单标签子集，包括7类基本情绪；双标签子集，包括12类复合情绪；4) 人工准确检测的地标点的位置；5) 基本情绪和复合情绪的基线分类输出。我们希望这个数据库的发布能使更多的研究者对真实世界的表情分布或检测的影响有勇气，并成为研究者在挑战条件下比较其面部表情分析算法有效性的有用的基准再来源。

8. 鸣谢

这项工作得到了国家自然科学基金的部分支持（项目6157306861471048，61375031和61532006），北京。资助项目：Z161100004916088；中央高校基本科研业务费资助项目：2014ZD03-01；大学新世纪优秀人才计划（NCET-13-0683）。

## 参考文献

- [1] C.F. Benitez-Quiroz, R. Srinivasan, and A. M. Martinez. Emotionet: 一种准确的、实时的算法, 用于自动注释野外的一百万个面部表情. In *Proceedings of IEEE International Conference on Computer Vision & Pattern Recognition (CVPR16)*, Las Vegas, NV, USA, 2016.
- 2 V.Bettadapura. Face expression recognition and analysis: the state of the art. arXiv preprint arXiv: 1203.6722, 2012.
- 3 J.C. Borod. *情感的神经心理学*. Oxford University Press New York, 2000.
- 4 C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1-27:27, 2011. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- 5 N.N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition*, 2005. CVPR 2005. IEEE 计算机协会会议, 第一卷, 第 693 页. IEEE, 2005.
- 6 A.Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon. 2013 年野外挑战中的情感识别. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 509-516. ACM, 2013.
- 7 A.Dhall, R. Goecke, S. Lucey, and T. Gedeon. 艰难条件下的静态动物表达分析: 数据、评估协议和基准. In *Computer Vision Workshops (ICCV Workshops)*, 2011 IEEE International Conference on, pages 2106-2112. IEEE, 2011.
- A.Dhall, O. Ramana Murthy, R. Goecke, J. Joshi, and T. Gedeon. 基于视频和图像的情感识别在野外的挑战: Emoti w 2015. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 423-426. ACM, 2015.
- 9 J. Donahue, J. Jia, O. Vinyals, J. Hoffman, N. Hang, E. Tzeng, and T. Darrell. Decaf: 一种用于通用视觉识别的深度卷积行为特征. In *ICML*, pages 647-655, 2014.
- 10 S. Du, J. Tao, and A. M. Martinez. 情感的复合面部表达. *美国国家科学院院刊*, 111 (15): E1454-E1462, 2014.
- 11 P. Ekman. 面部表情和情绪. *American psychologist*, 4 (4): 3-4, 1993.
- 12 P. Ekman and W. V. Friesen. 面部动作编码系统. 1977.
- 13 P. Ekman, W. V. Friesen, M. O. Sullivan, A. Chan, I. Diacoyanni-Tarlatzis, K. Eider, R. Krause, W. A. LeCompte, T. Pitcairn, P. E. Ricci-Bitti, et al. 面部情绪表达判断的普遍性和文化差异. *人格与社会心理学杂志*, 53 (4): 712, 1977.
- 14 B. Fasel and J. Luettin. 自动面部表情分析: 一项调查. *模式识别*, 36 (1): 259-275, 2003.
- 15 C. Ferri, J. Hernandez-Orallo, and R. Modroiu. 分类的性能措施的经验心理比较. *Pattern Recognition Letters*, 30(1): 27-3, 2009.
- 16 I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Amner, W. Cukierski, J. Tang, D. Thaler, D.-Lee, et al. Challenges in representation learning: 三个机器学习竞赛的报告. In *Neural information processing*, pages 117-124. Springer, 2013.
- 17 X. e and P. Niyogi. Locality preserving projections. 在 *NIPS*, 第16卷, 2003.
- 1 M. Inc. Face++ 研究工具包. [www.faceplusplus.com](http://www.faceplusplus.com), Dec. 2013.
- 19 S.E. Kahou, C. Pal, X. Bouthillier, P. Froumenty, C. Gulchere, R. Memisevic, P. Vincent, A. Courville, J. Ben-gio, R. C. Ferrari, et al. 结合特定模式的深度神经网络进行视频中的情感识别. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 543-550. ACM, 2013.
- 20 B.-K. Kim, J. Roh, S.-Dong, and S.-Lee. Hierarchical committee of deep convolutional neural networks for robust facial expression recognition. *多模态用户杂志 Interfaces*, 第17页, 2016.
- 21 A. Krizhevsky, I. Sutskever, and G. E.inton. 用深度卷积神经网络进行图像分类. In *Advances in neural information processing systems*, pages 1097-1105, 2012.
- 22 G. Levi and T. assner. 通过卷积神经网络和映射的二进制模式在野外进行情感识别. *s. 在 2015 年 ACM 国际多模态交互大会上的论文集*, 第503-510页. ACM, 2015.
- 23 C. 刘和. Wechsler. 基于 Gabor 特征的分类, 使用增强的 Fisher 线性判别模型进行人脸识别. *Image processing, IEEE Transactions on*, 11(4): 467-476, 2002.
- 24 M. Liu, S. Li, S. Shan, and X. Chen. 用于面部表情识别的 Au-aware 深度网络工程. In *Automatic Face and Gesture Recognition (FG)*, 2013 10th IEEE International Conference and Workshops on, pages 1-6. IEEE, 2013.
- 25 M. Liu, S. Li, S. Shan, and X. Chen. Au-inspired deep networks for facial expression feature learning. *Neurocomputing*, 159: 126-136, 2015.
- 26 P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, J. Ambadar, and I. 马修斯. 扩展的 cohn-kanade 数据集 (ck+). 一个完整的数据集, 用于动作单元和情感指定的预演. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010 IEEE Computer Society Conference on, pages 94-101. IEEE, 2010.
- 27 J. Lv, J. Feng, and C. Xu. 通过深度学习的面部表情识别. 在 *智能计算 (SMARTCOMP)*, 2014 国际会议上, 第303-310页. IEEE, 2014.
- 2 M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek. 菲律宾女性面部表情 (afpe) 数据库. 199.
- 29 M. J. Lyons, J. Budynek, and S. Akamatsu. 单一面部图像的自动分类. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, (12): 1357-1362, 1999.
- 30 D. McDuff, R. El Kaliouby, T. Senechal, M. Amr, J. F. Cohn,

and R. Picard. *Affectiva-*  
*mit* 面部表情数据集 (am-

- (Feed)。在野外收集的自然而自发的面部表情。在计算机视觉和模式识别研讨会 (CVPRW) 上, 2013 年 IEEE 会议, 页 1。IEEE, 2013.
- 31 A.Mollahosseini, D. Chan, and M. Mahoor. 使用深度神经网络进行更深入的面部表情识别。In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pages 10. IEEE, 2016 年。
  - 32 .-W.Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler. 在小数据集上使用转移学习进行情感识别的深度学习。在 2015 年 ACM 国际多模态交互会议论文集上, 第 443 449 页。ACM, 2015。
  - 33 T.Ojala, M. Pietikainen, and T. Maenpää. 多分辨率灰度和旋转不变的纹理分类与局部二进制模式。Pattern Analysis and Machine Intelligence, IEEE Transactions on, 24(7):971-977, 2002.
  - 34 M.Pantic and L. J. M. Rothkrantz. 面部表情的自动分析: 技术的现状。IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12):1424-1445, Dec 2000.
  - 35 M.Pardas and A. Bonafonte. 使用隐藏马尔科夫模型的面部动画参数提取和表情识别。Signal Processing: Image Communication, 17(9):675-686, 2002.
  - 36 X.Peng, .Xia, L. Li, and X. Feng. 争取在野外进行面部表情识别。一个新的数据库和深度识别系统。在 IEEE 计算机视觉和模式识别研讨会论文集上, 第 93-99, 2016。
  - 37 J.A. Russell. 是否存在从面部表情中识别情绪的普遍性--跨文化研究的回顾。心理公告, 115 (1) : 102, 1994。
  - 38 C. Shan, S. Gong, and P. W. McOwan. 基于局部二进制模式的面部表情识别。一个全面的研究。图像与视觉计算, 27(6):03-16, 2009.
  - 39 A.Sharif Razavian, .Azizpour, J. Sullivan, and S. Carlsson. 现成的 Cnn 特征: 一个令人震惊的识别基线。In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 06-13, 2014.
  - 40 K.Simonyan and A. Isserman. 用于大规模图像识别的极深卷积网络。arXiv 预印本 arXiv:1409.1556, 2014。
  - 41 .Tang. 使用线性支持向量机的深度学习。arXiv preprint arXiv:1306.0239, 2013.
  - 42 .I. Tian, T. Kanade, and J. F. Cohn. 识别面部表情分析的动作单元。IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2):97-115, Feb 2001.
  - 43 .Wen, K. Tang, .Li, and .Liao. 一种用于深度人脸识别的特征学习方法。在欧洲计算机视觉会议上, 第 499-515 页。Springer, 2016.
  - 44 J.Whitehill and C. W. Omlin. Linear features for face and gesture recognition. In Automatic Face and Gesture Recognition, 2006. FGR 2006. 第七届国际会议, 第 5 页。IEEE, 2006.
  - 45 J.Whitehill, T.-f.Wu, J. Bergsma, J. R. Movellan, and P. L. Ruvolo. Whose vote should count more: 优化整合来自未知专长的标签者的标签。In Advances in neural information processing systems, pages 2035-2043, 2009.
  - 46 Eng, M. Pantic, G. I. Roisman, and T. S. Huang. 情感识别方法的调查。音频、视觉、和同步表达。Pattern Analysis and Machine Intelligence, IEEE Transactions on, 31(1):39-50, 2009.
  - 47 X. Tang, L. Yin, J. F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, and J. M. Girard. Bp4d-spontaneous: 一个高分辨率的自发三维动态面部表情数据库。图像和视觉计算, 32 (10) : 692-706, 2014。