# 概要

一张照片胜过千言万语; 因此，一张脸的照片可以展示关于心理规格、性别、年龄和情感状态的大量细节。面部表情在基于社区的互动中起着重要作用，常用于情绪的行为分析。从面部图像识别自动面部表情在计算机视觉领域是一项具有挑战性的任务，并且承认有大量的应用，如驾驶员安全，人机交互，卫生保健，行为科学，视频会议，认知科学等。本文提出了一种基于深度学习的人脸表情识别方法。该方法由两部分组成。前者利用局部引力描述子从人脸图像中发现局部特征，后者将描述子输入到一个新的深度卷积神经网络(dcnn)模型中。提议的 dcnn 有两个分支。第一个分支探索几何特征，比如边、曲线和直线，而整体特征则由第二个分支提取。最后采用分级融合技术计算最终分类得分。该方法与25种最先进的方法一起，在5个可用的基准数据库上实现，即面部表情识别2013，日本女性面部表情，扩展 cohnkanade，卡罗林斯卡定向情感面孔，和真实世界情感面孔。数据库包括七种基本情绪: 中性，快乐，愤怒，悲伤，恐惧，厌恶和惊讶。该方法采用准确度、精确度、查全率和 f1评分四个评价指标与现有方法进行了比较。实验结果表明，该方法在所有数据库上的性能优于所有最先进的方法。索引项ー深层卷积神经网络(dcnns)、面部表情识别(fer)、局部引力(gf)描述子、评分级融合、 softmax 分类。

# 介绍

情感计算是一个研究领域，试图开发工具/设备和系统，可以识别，解释，处理和模拟人类的影响。近年来，由于其引人注目的学术和商业应用，如人机交互、虚拟现实、医疗保健、欺骗性检测、多媒体、扩增实境、驾驶员安全和监控等，使得人工智能和计算机视觉领域的研究团体受到了广泛的关注。人类表达过程情感状态的计算模型一般分为两类: 决策模型和预测模型。前者解释了表达的效果，而后者可以识别情绪的状态。从言语、身体姿势和生理信号中推导出不同形式的面部表情的非语言表达模型，为情感计算提供了有价值的资源。感兴趣的读者参考[1]了解更多关于情感计算的各种方法，模型和应用。在这项研究中，基于计算机的面部表情识别(fer)被认为是由于它能够模仿人类的编码技能。在情感计算中，电子计算是必不可少的。面部表情是表达人际关系中内在行为的非言语交际。此外，它是一种情绪分析技术，使用生物特征自动识别七种基本情绪: 中性(ne)、愤怒(an)、厌恶(di)、恐惧(fe)、快乐(ha)、悲伤(sa)和惊讶(su)。尽管为了开发通达情感的工具已经进行了大量的工作，但是识别人类表情仍然是一项具有挑战性的任务，特别是在野外环境下。其中一些值得注意的困难是: 1)当两种面部表情的差异很小时，很难高精度地区分它们; 2)一般来说，由于人与人之间的差异性和他们的脸部生物特征形状，不同的人对特定面部表情的表达是不同的。近年来的研究主要集中在基于手工特征的方法和基于深度学习特征的方法两个方面。前者又分为基于外观的特征和几何特征。基于外观的方法依赖于人脸图像中像素值的各种统计。例子包括 gobar wavelets [4] ，haar wavelet [5] ，local binary pattern (lbp)[6] ，[7] ，方向梯度直方图(hog)[7] ，[8] ，成束强度值的直方图(hbiv)[9] ，动态贝氏网路(dbn)[10] ，等等。另一方面，通过将图像转换为几何基元(比如角点或细节点[11]、边和曲线[12])来获得几何特征。这是完成的，例如，定位独特的特征，如眼睛，嘴，鼻子，和下巴，并测量他们的相对位置[13] ，[14] ，宽度，或许，其他参数。然而，基于传统方法的区别特征提取仅限于人类经验，难以获得，在大数据量情况下难以获得更好的性能。传统的方法不能满足实际的远程应用需求，而且它们还需要较高的计算代价和空间[15]。近年来，深层卷积神经网络(dcnns)在计算机视觉领域得到了广泛的应用。利用 dcnn，在图像分类问题上取得了许多突破，特别是人脸识别任务[16]-[18]。据观察，dcnn 近年来已经超越了传统的手工制作方法[19]-[21]。Dcnn 能够提取假设的特征，从低水平到高水平的面部图像的帮助下，几个非线性的联系[16]。此外，dcnn 还可以通过解决传统方法所带来的一些问题来提取有用的特征。在文献[22]中，将中心损失函数和验证识别模型相结合，设计了针对 fer 的 dcnn，以提供更好的识别能力。在另一项工作[23]中，提出了一个条件生成对抗性网络，以增加数据的大小，dcnn 用于面部表情分类[24]。Fathallahet al [25]讨论了一种基于几何组模型的识别算法。

1. 动机和贡献

从文献中可以清楚地看到，大多数现有的作品在拥有在受控实验室环境中捕获的图像的数据库上表现相当好。然而，这些工作不能产生令人满意的结果，更具有挑战性和实时数据库组成的图像有更大的变化。因此，有必要改进FER系统的性能。FER系统的性能依赖于特征工程。从现有的特性中设计新的特性可以提高系统的性能。这促使我们朝着这个方向进一步努力。从state的最新方法中也可以清楚地看到，大多数与FER任务相关的工作都是基于边缘信息[12]，[26]-[30]，因为它在个体表达上是不同的。重要的特征，如角，线，曲线，可以提取的边缘图像。边缘是图像中强度的显著局部变化。在过去，各种DCNN模型被用来提取假设的深度特征，用于开发FER系统。然而，功能的数量是相当大的。有时，深度特征可能导致过度拟合。此外，深度特征的提取非常耗时，而且需要强大的资源。此外，与大量的功能相比，这些功能中只有一小部分被使用。另一方面，使用梯度的边缘检测捕捉x和y方向上的微小变化，这被称为梯度。梯度是一个具有一定大小(M)和方向(D)的向量。当强度发生剧烈变化时，比如在边缘周围，M会更高。M提供关于边缘强度的信息。另一方面，D总是指向边的D。D表示图像的几何结构。因此，在该方法的第一步中，在计算M和D图像时，采用了基于重力的边缘描述符(GF)[31]，因为它利用了周围的像素信息，而没有考虑相邻像素在x和y方向上的差异。然而，该系统不仅依赖于局部边缘信息，还依赖于整体特征。因此，在第二步中，将M和D图像输入到新的DCNN中提取有用信息。提出的DCNN由两个分支组成:第一个分支由浅层DCNN组成，提取局部特征;第二个分支由主要DCNN组成，从M和D图像中提取整体特征。最后，对M和D图像得到的分类结果采用评分级融合技术得到最终结果。本文方法概述如图1所示。并与现有的25种方法进行了性能比较。所有方法在五个基准数据库上实现，即FER2013 (FER2013)[32]，日本女性面部表情(JAFFE) [33]， Extended CohnKanade (CK+) [34]， Karolinska Directed Emotional Faces (KDEFs) [35]， real - real Affective Faces (RAFs)[36]，[37]。为了衡量所有方法的效率，包括提出的一个，四个分类指标，即准确性，精度，召回和f1评分，考虑定量评价。经验结果表明，提出的方法击败了最先进的25种方法。其余的工作组织如下。在第二节中，回顾了与FER相关的早期工作。本文的方法见第三节。第四节给出了实验结果和讨论。第五节对本文的工作进行了总结。

# 相关工作

基于特征提取技术，可以将fer任务中的所有方法分为两类，即手工特征和深度学习特征。这一节简要地介绍了它们。该任务主要涉及两个步骤，即特征提取和分类。第一步采用高级域理解方法，利用了gobar小波[4]、曲线[12]、尺度不变特征变换[21]、hog[8]、LBP[6]、小点[11]、haar小波[5]、hbiv[9]、DBN[10]和边缘[38]等常规特征。第二步采用支持向量机[39]、前馈神经网络[40]和极限学习机[41]进行分类。Chenet al.[42]提供了一个叫做hog的特征描述符，它来自三个正交的平面(hog-tops)，用于从视频序列中提取动态纹理来表征面部外观变化。然而，手工制作的基于特性的方法在实际应用中性能有限，特别是对于任务。近年来，从文献中可以看出，对于任务[17]，[18]，[43]，[44]，基于深度学习的方法优于基于手工特征的方法。在fer2013[45]上考虑了浅CNN和深CNN对灰度图像的提取，并使用softmax分类器进行分类。在[46]中，在fer2013数据库[32]上分别采用DCNN框架和softmax进行特征提取和分类。Orozcoet等提出了基于alexnet-、vgg19-和resnet的迁移学习方法。Sunet al.[16]提出了一种用于人脸识别的DCNN模型和深度特征。在另一项工作中，sunet al.[48]考虑了Siamese网络来提高fer任务的效率。Barsoumet al.[49]讨论了vgg13网络与数据库的连接问题。该算法采用加权混合深度神经网络，由两个通道组成:其中一个通道采用部分vgg16框架对灰度图像进行面部表情特征提取。另一方面，在jaffe[33]和ck[34]数据库上，利用浅层DCNN对LBP图像进行特征提取，利用softmax分类器对提取的特征进行分类，然后在[50]中进行加权融合，将两个通道的输出进行合并。在[51]中，从imagenet数据库上预先训练的vgg19体系结构中提取特征用于fer任务，并在jaffe和ck数据库上使用SVM进行表达分类。在[52]中，对三个 dcnn 子网络分别进行了 fer2013和 affectnet 数据库的考虑和训练，并进一步采用加权融合的方法对三个网络进行了集成。同时，对网络赋予较大的权值，提高了识别精度。根据文献[54]提出的层次结构，利用 dcnn 提取基于外观的特征，并与基于几何特征的 dcnn 进行特征融合。然而，在 lbp 图像上提取出了外观特征，同样，在 jaffe 和 ck 数据库上，基于几何特征的网络也考虑了灰度图像。在[55]中，使用集成化重发50和 vgg16框架在 kdef 数据库中提取面部特征并对个人表情进行分类。Hasani 和 mahoor [56]提出了一个 dcnn 框架，该框架由三维的先启-再启动层和一个长时短期记忆单元(lstm)组成，这两个单元共同从面部图像中提取空间和时间关系(三维先启-再启动 地标)。几何和区域 lbp 特征被自动编码器合并，然后是基于 kohonen 自组织映射的分类器(自动编码器 自动分类器)来识别面部表情。提出了一种基于 dcnn 和 lstm (时空特征 lstm)编码面部表情特征的时空特征表征学习算法。Pons 和 masip [59]被认为是解决 fer 问题的 dcnns 的集合。Villanueva 和 zavala [60]提出了一个 dcnn 的面部表情分类: 只有快乐和悲伤。Menget al [61]和 liuet al [62]致力于身份感知模型的研究。Menget al [61]使用两个相同的 dcnn 流来联合估计不同的表达式和身份特征(iacnn)来寻找由于针对 fer 任务的个人属性引起的浮雕跨学科变化。另一方面，liuet al. [62]采用深度量学习(2b (n m) softmax)联合优化深度量和 softmax 的损失。Alamet al. [63]为 fer 问题求助于一个稀疏深度的同时循环网络(s-dsrn) ，并在模型中包含了一个退出率。Benitez-quirozet al. [64]提出了一种基于判别颜色特征和 gabor 变换算法(颜色特征 gabor 变换)的面部动作单元变化时间不变性算法。在文献[65]中，提出了一种深度综合多块集结卷积神经网络模型。它有两个分支。一个分支提取整体特征，而另一个分支从分割的表示图像斑块中提取局部特征。然后，结合两个特征向量，使用 dcnn 和 eti-pooling 对表达式进行分类。Zhanget al. [66]开发了一个广泛的学习系统。多级 dcnn 被开发用于提取面部图像中的中层和高层特征，以解决面部图像的综合问题[67]。在[68]中，一种注意力集中的 dcnn 被命名为一种深刻的情感来解决佛的问题。在[69]中，提出了一种基于心理机制的深度图网络。首先，利用分割技术将人脸图像分割成若干个小的关键区域。此外，这些关键区域然后转换成相应的面部表情区域。其次，从这些区域中提取局部外观特征进行进一步的 aus 分析。然后，以相关区域为顶点，以每两个路标之间的距离为边，构造 aus 人脸图来表示表情。最后，将人脸图中邻接矩阵的学习混合特征融入到基于图的卷积神经网络中，将局部外观信息和全局几何信息结合起来。Kopaczkaet al. [7]提出了一个高分辨率的热面部图像数据库，以完成佛的任务。此外，为了提高鲁棒性，他们对现有的红外地标检测方法进行了扩展，并分析了一种深度学习方法在这一任务上的性能。

# 提出的方法

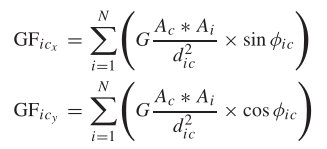
本部分简要概述了使用 gf 的图像的边缘描述符，然后详细描述了我们提出的针对 fer 任务的 dcnn。

## 边缘描述

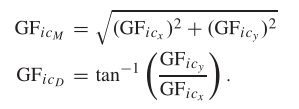
Roy 和 bhattacharjee [70]指出，图像的每个像素值都平行于一个普遍的物体，因此，它被认为是物体的质量。图像的 gf 被其邻近像素上的中心像素所使用。万有引力定律指出，每个人的质量(m1)都会吸引宇宙中其他物体的质量(m2) ，这个力指向两个物体质量中心之间的一条直线(d) ，这个力(gf)与物体的质量成正比，与物体分离的平方成反比。在数学上，gf 是用以下公式计算的:



其中 g 是万有引力常数，它的值是6.67259 × 10-11。设 a 为灰度图像。让我们考虑一个局部3掩模的中心像素 ac。意思是 ac 周围有8个相邻的人工智能像素。从万有引力定律可以清楚地看出，所有相邻的八个像素都会对空间产生作用力。因此，相邻像素点对交流电的作用力可以用基特性来表示。然后，当基性与 x 轴成一定角度时，基性的 x 和 y 分量分别为 gficx = gfic × sinφ 和 gficy = gfic × cosφ。可以分别使用(2)和(3)计算 gfick 和 gfick。图2显示了输入图像和边缘强度，即 x 和 y 方向的 ms



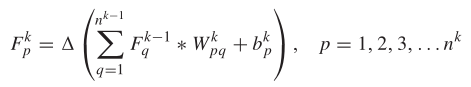
其中 n 是掩码相邻像素的总数，d2ic 是第 i 像素和中心像素之间的平方欧几里得度量。Gfic 的 gfim 和 gfid 的计算公式如下:



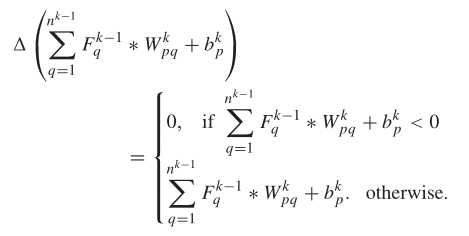
方程(4)和(5)可以重复使用，每个像素作为中心像素，找出 m 和 d 的灰度图像的梯度

## 建筑结构的计划 dcnn

通过对目标任务的损失函数进行反向传播学习，使得 dcnns 能够自动学习目标任务的特征，并且能够更准确地描述目标任务。现有的基于 dcnn 的模型，如 vgg-16和 vgg-19，是建立在一个单一的分支上，与卷积层顺序连接，通常侧重于均匀缩放的接受域，忽略边缘信息。因此，他们缺乏足够的空间结构特征收集面部外观。介绍了解决这一问题的多卷积网络。在这一部分，我们将介绍一种新的 dcnn 来完成 fer 的任务。建议的 dcnn 的结构如图3所示。它由两个分支组成。第一个分支能够从图像的 m 和 d 中提取重要的局部特征，比如边缘、直线、曲线和角点，就像浅 dcnn 设计的那样。另一方面，第二个分支负责提取整体特征，从而区分一个表达式和其他表达式，因为主要的 dcnn 被考虑。由于考虑了图像的 m 和 d，所提出的 dcnn 能够提取出与单个表情相关的特征。Dcnn 的第一个分支由三个依次连接的卷积层组成，即两个 maxpooling、一个 average pooling 和 zero-padding，它们依次连接。另一方面，dcnn 网络的第二个分支包含五个卷积、三个最大池、一个平均池和一个上采样层。此外，这两个分支是连接和转发到两个密集层面部表情的分类。每一层的详细描述及其参数如图4所示。此外，还采用了5 × 5和4 × 4的丰富语境信息过滤器。这些过滤器可以让网络学习真正的边缘变化。卷积层最大的优点是能够自动提取特征，kth 卷积层由 nk 特征映射图组成，其中 p = 1,2,3，... ，nk 和 k 代表特定的卷积层。每个特征映射，即 fk-1q，其中 q = 1,2,3，... ，nk-1来自(k-1)第卷积层，与过滤器 wkpq 卷积，并加上 biasbkp。此外，卷积特征映射被反馈到非线性激活函数校正线性单元(rlu)。方程(6)说明了如何获得一个卷积特征映射。



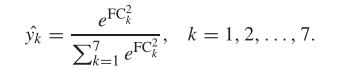
其中，∗表示卷积操作。激活函数的职责是重新计算从一个节点输入到不同激活节点的加权和。这里采用relu是因为它可以在一定程度上缓解消失梯度问题，从而降低成本/损失函数的值。它在复杂数据库上的计算速度更快，性能更好[71]。relu激活函数的数学表示如(7)所示。将max-pooling应用于(6)得到的卷积特征映射，通过提供卷积特征映射的抽象表示形式来克服过拟合问题。Max-pooling从每个feature map中计算每个patch的最大值，聚焦在patch中最先呈现的feature。为了简化模型，它还减少了参数的数量。此外，它还提供平移、旋转和比例不变的特征图



连接从 l16和 l29层获得的特征映射 fm16和 fm29。然后，将连接后的特征映射图 fm16和 fm29扁平化，再将其作为输入 fc0输入到第一个完全连接层。将第一完全连接层的输出 fc1馈入第二完全连接层以生成 fc2作为输出。两个完全连接的层所涉及的数学运算用下面的公式表示:

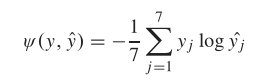


其中 wi 和 bi 是完全连接层的权重和偏置。从实验中可以观察到，由于一些可学习的参数与它们相关，在两个完全连接的层中都会出现过拟合问题。在这项工作中，采用压差技术来解决全连接层中出现的过拟合问题。第二完全连接层 fc2的输出进一步馈入软最大层。最大软层由7个神经元组成，产生一个概率向量 y = [ y1 y2 y3 y4 y5 y6 y7]。概率向量由七个概率值组成，本研究考虑了七类面部表情。K 的概率值由以下方程式求得:



### 网络训练

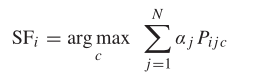
提出的网络独立训练的 m 和 d 的梯度的人脸图像和估计的概率每类。该方法首先采用 gloot 均匀化方法初始化网络权值，然后采用 adam 优化方法避免局部最优[72] ，同时引入学习率衰减方法提高训练效果。在 adam 优化中，初始学习率和学习率衰减分别被指定为0.00001和1e-4。分类交叉熵被用来发现多元分类的损失，它是一个度量我们的模型的误差的方法。分类交叉熵是用下面的公式计算出来的:



其中 y (= [ y1，y2，y3，y4，y5，y6，y7])是实际标签的一个热编码向量。批量大小16被认为是在培训拟议的 dcnn，因为网络可以占用较少的内存在我们的系统和60年代被认为是 jaffe，ck ，kdef，和 raf 数据库，而200年代被用于 fer2013。

### 评分级融合（Score-Level Fusion）

评分级融合: 为了估计七个基本表达式的最终预测，评分级融合技术[73]是在 m 和 d 的梯度上进行的。数学上，分数级融合是通过以下方程式完成的:

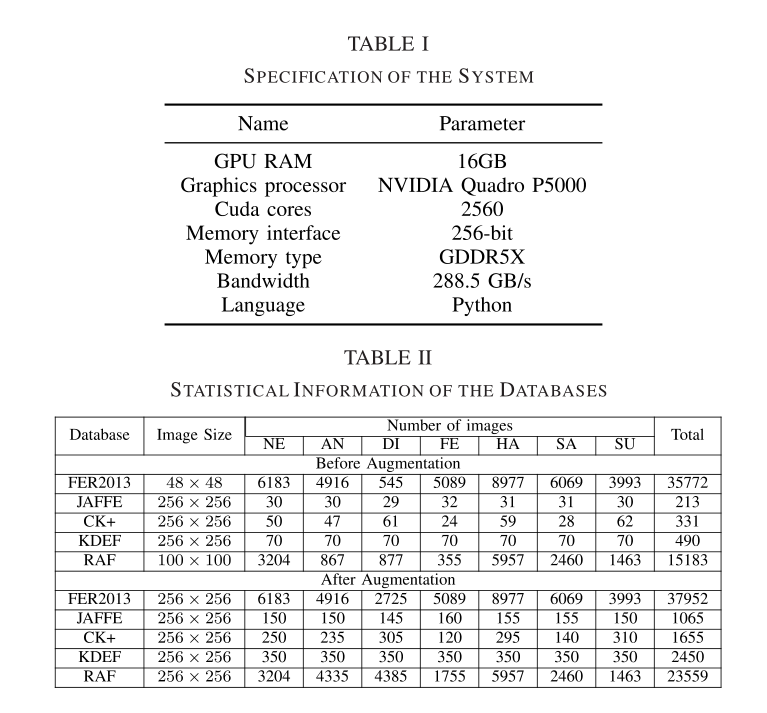


其中 c 表示各种表达式，i 表示输入样本，n 表示两种不同的模态，在这项研究中，输入图像的 m 和 d，被认为是两种不同的模态。评分级融合更容易衡量个体的模式评分，并且它能够给出更好的表现。

# 实验结果和讨论

## 环境设置

本研究以 keras 架构及Anaconda开发平台作为训练及测试模式。Python 语言被使用，因为许多深度学习库都是使用它开发的。该系统的规格见表一。

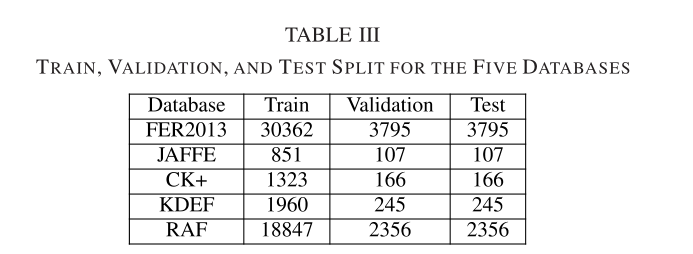


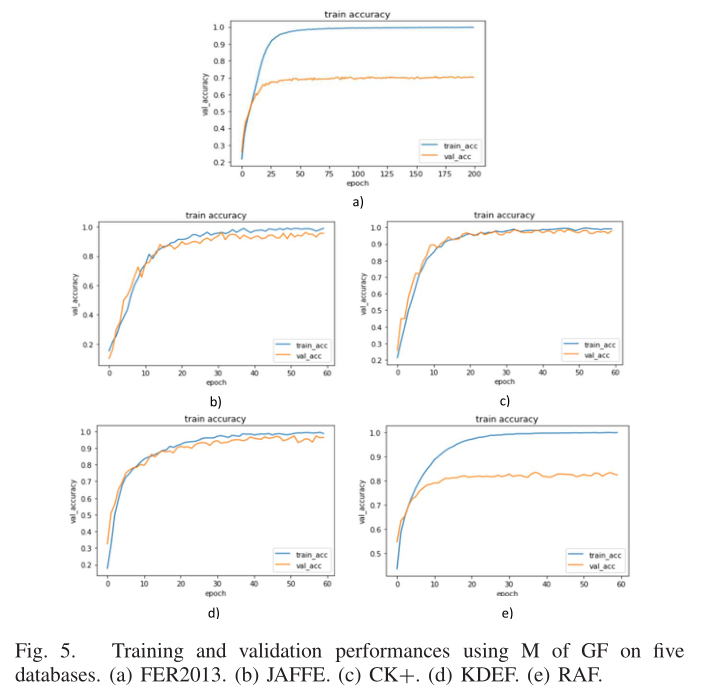
## 数据库描述

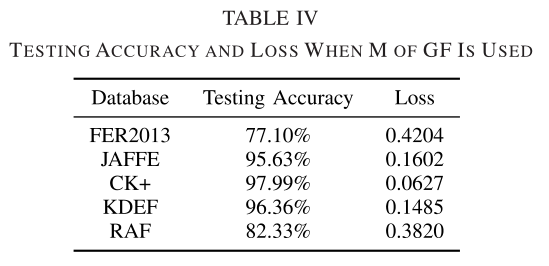
在这项工作中，五个著名的基准数据库，即 fer2013[32] ，jaffe [33] ，ck [34] ，kdef [35]和 raf [36][37] ，被用于评估所提议的网络，因为这些数据库包含七个基本的通用面部表情，即 ne，an，di，fe，ha，sa 和 su。表二的顶部描述了上述数据库的统计信息。

## 数据增强

拥有一个大型数据库对于衡量 dcnn 模型的性能非常重要。此外，我们可以防止 dcnn 模型学习不相关的特征，因为不相关或部分相关的特征会对模型性能产生负面影响。然而，通过增加我们已有的数据，性能可以在一定程度上得到改进。Dcnn 可以是不变的平移，视点，大小，或照明。因此，一些图像处理技术，如旋转图像5 o 顺时针和逆时针方向，水平翻转，并加入高斯噪声，被认为是扩大了总数的图像七个面部表情的 jaffe，ck ，和 kdef 数据库，以增加这些数据库的多样性。另一方面，增加面部表情的 fer2013和 an，di，和 fe 表情 raf 只是由于他们的不平衡等级。表二底部描述了数据增加后五个数据库的统计信息。此外，每个数据库分为三个部分: 培训、验证和测试。所有实验均采用十倍交叉验证技术，以评估该方法的性能。换句话说，在10个子集中，8个子集用于训练，一个子集用于验证，剩下的子集用于测试。这项研究报告了平均分类结果。表三显示了在训练，验证和测试过程中使用的每个折叠面部图像的数量。





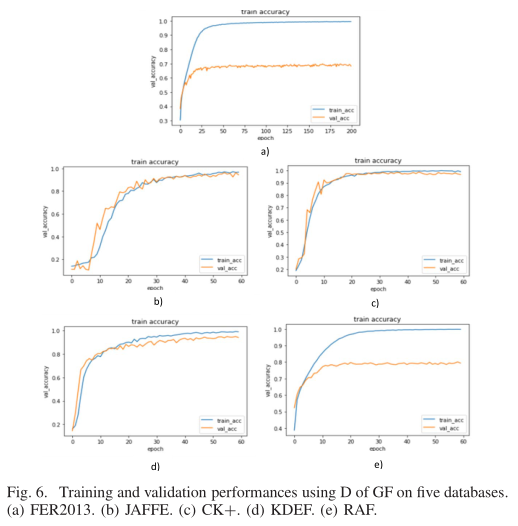


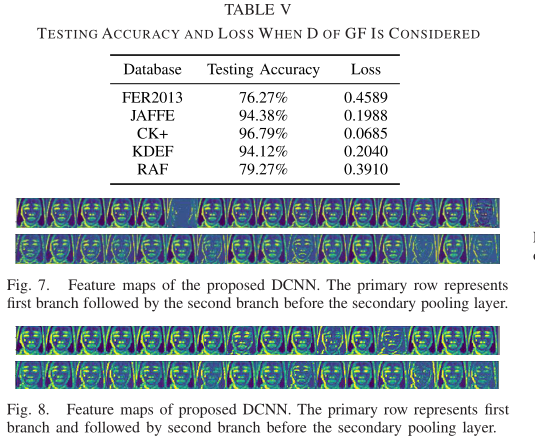
## 使用M的实验结果

在第一个实验中，只有 m 后面的 dcnn 被考虑。换句话说，该模型的上半部分只用于训练和验证。因此，模型的上半部分是在上述五个数据库中实现的。图5显示了相对于每个数据库上的历元的训练和验证性能。每个数据库的平均测试准确度和平均损失见表四。从表四可以清楚地看出，三个数据库的结果都很好: jaffe，ck 和 kdef。然而 fer2013和 raf 数据库的结果相对较差，但这些结果可以接受。

## 使用D的实验结果

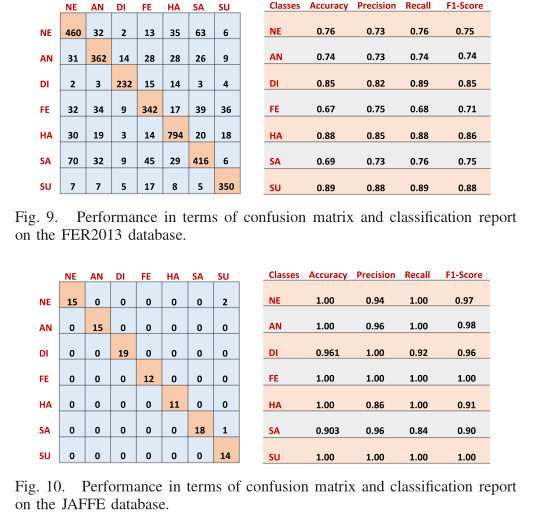
在第二个实验中，只有 d 和 dcnn 被采用。换句话说，建议模型的下半部分只考虑用于训练和验证。因此，该模型的下半部分是在上述五个数据库上执行的。图6显示了每个数据库上迭代的训练和验证精度。每个数据库的平均测试准确度和平均损失见表五。从表五可以看出，测试结果与表四的趋势相同。然而，如果 d 后跟拟议的 dcnn，那么性能就会下降。

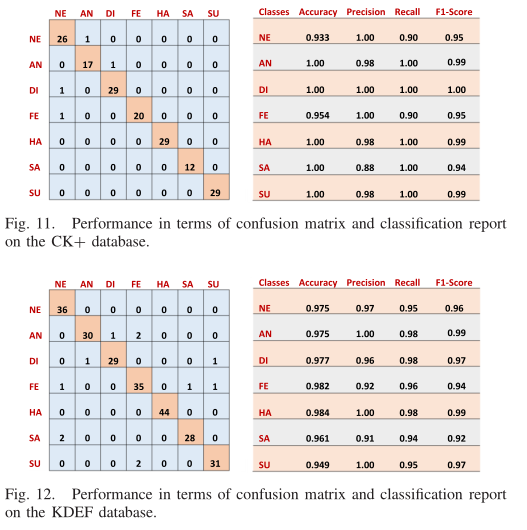


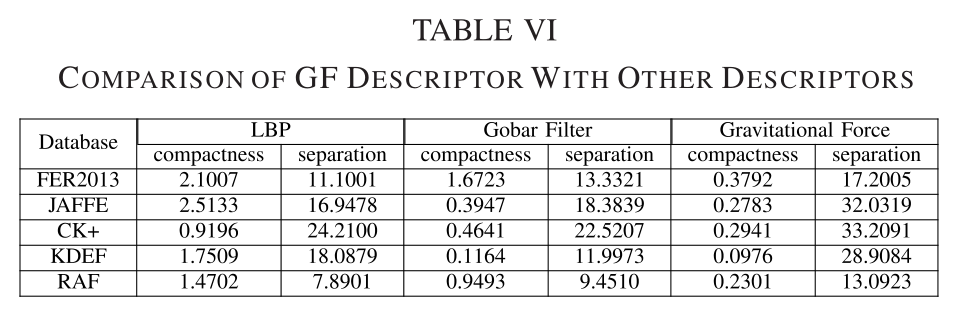


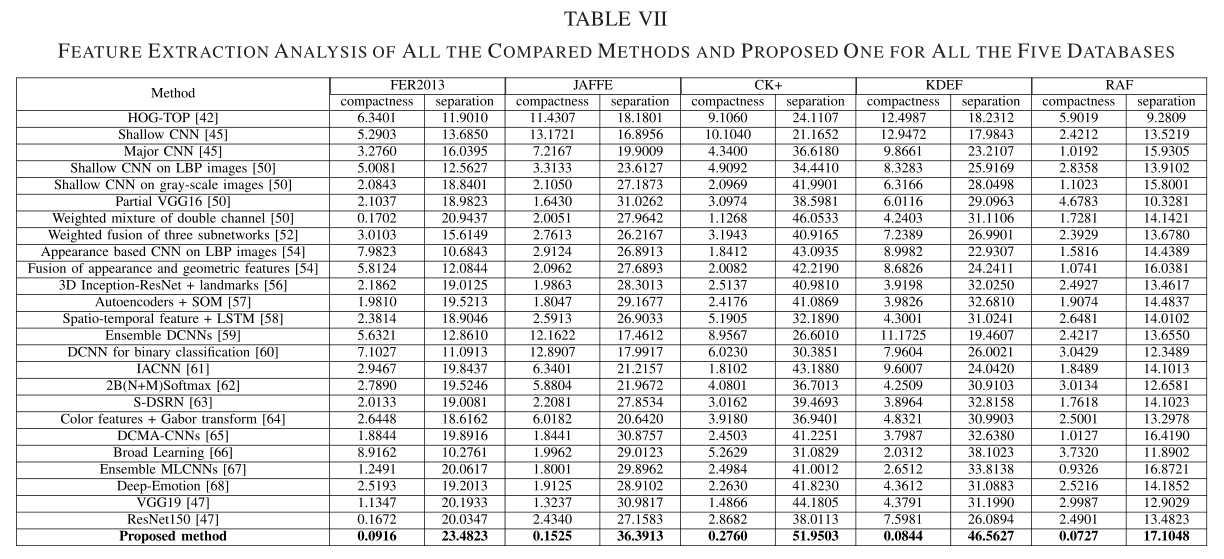
## 同时使用M和D的实验结果

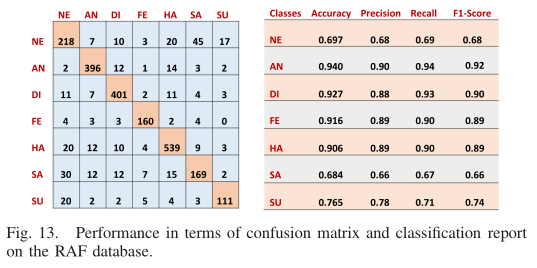
在第三个实验中，使用了图3所示的完整模型。建议的模型在五个数据库中运行。图7和图8显示了由所提出的 dcnn 获得的中间特征图。图9-13分别显示了从 fer2013、 jaffe、 ck 、 kdef 和 raf 数据库中获得的十分之一的混淆矩阵和分类报告混淆矩阵。对每个数据库的报告分类报告分别进行十倍平均，得到报告的分类报告。表八最后一行报告了拟议模型在五个数据库中获得的平均准确度。











## 模型分析

提出的方法包括两个步骤。第一步利用特征描述子提取边缘信息，第二步利用 dcnn 调整边缘信息提取局部和整体特征。Bhattacharjee 和 roy [31]已经证明了 gf 特征描述符的性能优于其他边缘信息抽取技术。因此，我们没有再做同样的实验。然而，gf 特征提取器的性能与两个著名的纹理测度 lbp 和 gabor 人脸描述器进行了比较。将一幅256 × 256像素的人脸图像作为输入分别输出到上述三个描述子，并产生一个1 × 65536的特征向量作为输出。然后将得到的特征向量映射到加法多面体上，得到 d 值为256 × 256 = 65536的点。这个过程会在数据库的所有脸部图像中重复。在这个过程的最后我们会得到一个 d 维的多面体，其中一个面的图像将被表示为一个点。如果两个类的点高度重叠，那么特征描述符就不够好。当两个类的点高度重叠时，精度会降低。在这里，重叠是基于紧凑度和分离度计算的。紧凑性和分离性决定了聚类结果的质量。当点彼此接近时，簇具有良好的紧凑性，当簇不重叠时，簇具有良好的分离性。换句话说，紧性和分离性的理想值分别为零和无穷大。本文首先考虑了七个基本表达式，然后用 k- 平均法将这些点划分为 k = 7个簇。上述三个特征描述符在五个数据库中的紧凑度和分离度值计算如表六所示。从表 vi 中可以看出，gf 的紧凑度值比 lbp 和 gabor 脸描述符要小。另一方面，在三个特征描述子中，gf 的值分离度最高。因此，我们可以得出结论: gf 比 lbp 和 gabor 人脸描述符更好。换句话说，特征向量信息量更大，在使用 gf 时能够区分面部表情。同样的实验是在提取了 dcnn 和其他最先进的模型的特征之后进行的，紧凑度和分离度的数值见表七。从表七可以得出两个结论: gf 和拟议的 dcnn 联合产生的特征，具有更独特的功能比单独由 gf 产生的特征，因为紧凑度值较小和分离值较高。因此，我们可以说，由于同样的原因，gf 后跟拟议的 dcnn 比最先进的方法更好。

## 比较结果

1. **翻译：**
2. 在最后一个实验中，我们提供了与25个最先进的算法的比较结果，例如，hog-top [42] ，shallow cnn [45] ，major cnn [45] ，shallow cnn on lbp images [50] ，shallow cnn on gray-scale images [50] ，partial vgg16[50] ，weighted mixture of double channel [50],三个子网络的加权融合[52] ，基于外观的 cnn 对 lbp 图像[54] ，外观和几何特征的融合[54] ，3-d inception-resnet landmarks [56] ,自动编码器 som [57] ，时空特征 lstm [58] ，ensemble dcnns [59] ，dcnn 二进制分类[60] ，iacnn [61] ，2b (n m) softmax [62] ,s-dsrn [63] ，color features gabor transform [64] ，dcma-cnns [65] ，broad learning [66] ，ensemble mlcnns [67] ，deep-emotion [68] ，vgg19[47] ，以及 resnet150[47]。然而，这种比较只是基于平均识别精度。上述的一些方法已经在视频中实现了。很少有作品被认为是较少的类别。因此，我们通过保持总体架构不变来改变这项研究的一些算法。表八显示了上述最先进方法取得的平均分类准确率。从表8可以清楚地看出，提出的模型打败了五个数据库中现有的25种方法，这是由于使用了基于 gf 的局部边缘特征以及提出的 dcnn 提取的整体特征，这是我们的主要重点。然而，所有的方法也进行了比较，基于训练和测试时间。一般来说，一个方法的训练时间取决于网络的大小、输入的大小、单元的数量、折叠的数量等。在这项研究中，所有模型都是按照各自的规格实现的。然而，我们考虑十倍交叉验证和60个时代，同时训练所提出的方法在所有数据库，除了 fer2013。所提出的方法仅对 fer2013数据库进行了200个时代的训练。所有最先进的方法所需的培训和测试时间，包括所有五个数据库的拟议方法，见表八。然而，单一交叉验证的测试时间仅在表 viii 中注明。每个图像的测试时间(ttpi)对于数据库的所有图像都是相同的，因为它们的大小是相等的。然而，不同的方法有所不同。从表8可以清楚地看出，除了 fer2013数据库之外，拟议的 dcnn 在所有数据库中平均需要培训和测试时间。提出的方法大约需要960分钟来训练 fer2013的方法，这是相当大的。
3. **原文:** 可修改后右键重新翻译
4. In the last experiment, we provide comparative results against 25 state-of-the-art algorithms, for example, HOG-TOP [42], shallow CNN [45], major CNN [45], shallow CNN on LBP images [50], shallow CNN on gray-scale images [50], partial VGG16 [50], weighted mixture of double channel [50], weighted fusion of three subnetworks [52], appearance-based CNN on LBP images [54], fusion of appearance and geometric features [54], 3-D Inception-ResNet+landmarks [56], autoencoders+SOM [57], spatiotemporal feature+LSTM [58], ensemble DCNNs [59], DCNN for binary classification [60], IACNN [61], 2B(N+M)Softmax [62], S-DSRN [63], color features+Gabor transform [64], DCMA-CNNs [65], broad learning [66], ensemble MLCNNs [67], deep-emotion [68], VGG19 [47], and ResNet150 [47], on five publicly available databases. However, the comparison is done based on average recognition accuracy only. Some of the abovementioned methods were implemented on videos. Few works considered less number of classes. Thus, we change a few of these algorithms for this study by keeping the overall architecture the same. Table VIII shows the average classification accuracies achieved by the abovementioned state-of-the-art methods. It is clear from Table VIII that the proposed model defeats all the 25 existing methods on five databases, and it happens due to the use of GF-based local edge features along with holistic features extracted by the proposed DCNN, which is our main focus. However, all the methods are also compared based on training and testing times. Generally, the training time of a method depends on the size of the network, size of the input, number of epochs, number of folds, and others. In this study, all the models are implemented according to their respective specifications. However, we consider tenfold cross validation and 60 epochs while training the proposed method on all the databases except FER2013. The proposed method is trained for 200 epochs for the FER2013 database only. The training and testing times required by all the stateof-the-art methods, including the proposed method on all the five databases, are reported in Table VIII. However, testing time for onefold cross validation is noted only in Table VIII. Testing time per image (TTPI) is the same for all the images of a database as their size is equal. However, it varies from one method to another. It is clear from Table VIII that the proposed DCNN takes an average training and testing time across all the databases except the FER2013 database. The proposed method takes about 960 min to train the proposed method for FER2013, which is quite large.

# 结论

实验结果表明，该方法能够有效地解决静止图像和静止图像的匹配问题。在实验室控制的环境下，面部表情不同于野外环境，野外环境更自然，更自发。因此，三个数据库，即 jaffe，ck 和 kdef，在实验室控制的环境中开发，被考虑在这项工作中。这项研究还采用了两个数据库，即 fer2013和 raf，在野外建立，以证明所提出的方法的效力超过最先进的方法。提出了一种新颖的 dcnn 框架，用于提取面部表情的整体特征。然而，在使用所提出的 dcnn 模型之前，采用了基于 gf 的边缘描述子来提取低层的局部特征。基于 gf 的边缘描述子产生两个中间的局部特征，即 m 和 d。最后，采用分数级融合技术，利用 m 和 d 将模型的输出结合起来，对 fer2013、 jaffe、 ck 、 kdef 和 raf 分别达到了78% 、98% 、98% 、96% 和83% 的平均识别率。实证结果表明，局部特征和整体特征可以共同增强传递任务。实验结果还表明，该方法考虑了平均时间，性能优于25种基线方法。然而，在实验室控制的环境下，其性能通常不如电脑，值得进一步研究。此外，将所提出的模型部署到实际应用中也是值得研究的。