

基于深度学习的人脸面部表情识别的研究

路晓亚¹

LU Xiaoya

摘要

脸部表情在日常社交中占据着重要地位，目前主要利用计算机对图像中的表情特征区域进行提取和分类，完成脸部表情识别任务。尽管图像经过归一化处理可以消除大部分噪声干扰，却难以解决图像部分信息被遮挡的问题。深度学习中的卷积神经网络能够自动提取更深层次的面部本质特征，从而减小人为和外界环境对脸部识别的影响，提高表情识别的效率和准确率，对基于深度学习的脸部表情识别算法进行了研究。

关键词

深度学习；表情识别；卷积神经网络；人脸检测；特征提取

doi: 10.3969/j.issn.1672-9528.2024.11.047

0 引言

随着人工智能、计算机视觉和信息技术在现代社会的不不断发展，研究者对人机交互有了高效、智能和可靠的要求。研究表明，在人与人交谈时，语言、肢体动作等方式传递给其他人的信息远没有面部表情传递出的信息多。面部表情是使用最广泛、最简单的信息传递方式，传递出的信息量占总体信息量的 55% 以上，能呈现出更真实有效的关键信息，清楚地表达个人情感和个人意愿。

随着深度学习的进步，人脸表情识别已经实现了从传统方法向深度学习方法的转变。深度学习是近些年发展最为迅猛的一个领域，在研究者们的不断钻研努力下，创造出一个又一个新的强大算法^[1]。与传统人脸表情识别方法不同，深度学习不仅可以借助计算机强大算力进行自主学习，无需手工特征提取，并且还可以使用海量数据运算，使实验结果更高效有力。

本文基于 VGG 网络和残差网络，对它们进行改良并提出以下具体方向：首先在 VGG 网络中引入通道 - 空间注意力机制以获取关键特征信息，并在高级残差模块中添加了压缩 - 激励模块，该模块的卷积核按照瓶颈层的方式排列，随后利用细化模块提取更深层次的特征以提升人脸表情识别准确率。其次，在残差网络中利用裁剪掩码模块进行数据增强处理，采用 Ghost 模块改进以减少冗余参数量，同时结合通道注意力机制和多尺度空间注意力机制，最后采用联合损失

函数以增加类间距、减少类内距离^[2]。

1 人脸表情识别

人脸表情识别的流程通常包括人脸检测、特征提取和表情分类。在人类社会的多个领域应用中，存在很高的研究价值人脸检测的主要任务是检测图像中是否存在人脸，并确定人脸的位置，同时提取人脸特征；在表情识别方面，特征提取是至关重要的步骤，将原始数据转化为更高级别的图像表示，比如形状、空间结构等，并对大量数据进行降维处理，因此，提取简单易懂的高级语义特征对于提高表情识别准确率至关重要；表情分类则是根据人脸部分运动特征的不同来区分不同的表情类型。

1.1 人脸检测

人脸检测作为人脸表情识别的环节之一，其目标是从任意给定的一张样本图片中，利用某种手段得到存在的人脸所在位置和关键点坐标。以下为几个常用的人脸检测算法。

(1) 基于 Adaboost 人脸检测，基于 AdaBoost 算法的人脸检测是一种高效的面部识别技术。AdaBoost 是 Adaptive Boosting 的缩写，是一种集成学习方法，它通过整合多个弱分类器构建一个强分类器。其基本原理是整合各种训练集训练出的弱分类器，形成一个新的强分类器，再将多个强分类

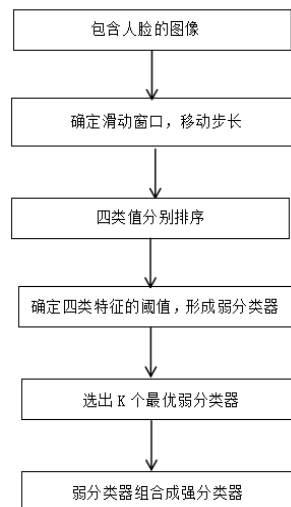


图 1 Adaboost 算法流程图

1. 商丘工学院信息与电子工程学院 河南商丘 476000

[基金项目] 2024 年度河南省科技厅科技攻关项目“智慧教育评价系统中基于深度学习的表情识别技术研究”(242102210111)

器整合在一起，组成一个级联分类器，从而可以精确检测出人脸区域。Adaboost 算法流程图如图 1 所示。

(2) 基于特征的人脸检测，基于特征的人脸检测方法定义是：提取人脸中与表情相关联的位置特征进行人脸检测。基于特征的人脸检测方法的一个主要优点是它们可以在没有大量计算资源的情况下实现较好的检测性能。然而，这些方法可能需要针对特定的应用场景进行特征选择和分类器训练，以获得最佳的检测效果。随着深度学习技术的发展，基于深度特征的检测方法也越来越流行，因为它们能够自动学习更加复杂和抽象的特征表示。

(3) 基于深度学习的人脸检测，随着深度学习在图像领域的不断应用，人脸检测方向上基于深度学习的方法极大地减少了对特征提取的依赖程度和有着处理大容量数据的潜力。一些常见的用于人脸检测的深度学习方法是级联卷积神经网络（Cascade CNN）和多任务卷积神经网络（MTCNN）等神经网络模型。其中，Cascade CNN 是一种深度学习架构，它通过多个阶段的卷积神经网络来逐步提高任务的性能，尤其是在图像识别和分类任务中。该网络一般分为 3 级级联：12-Net、24-Net 和 48-Net，具体流程如图 2 所示。

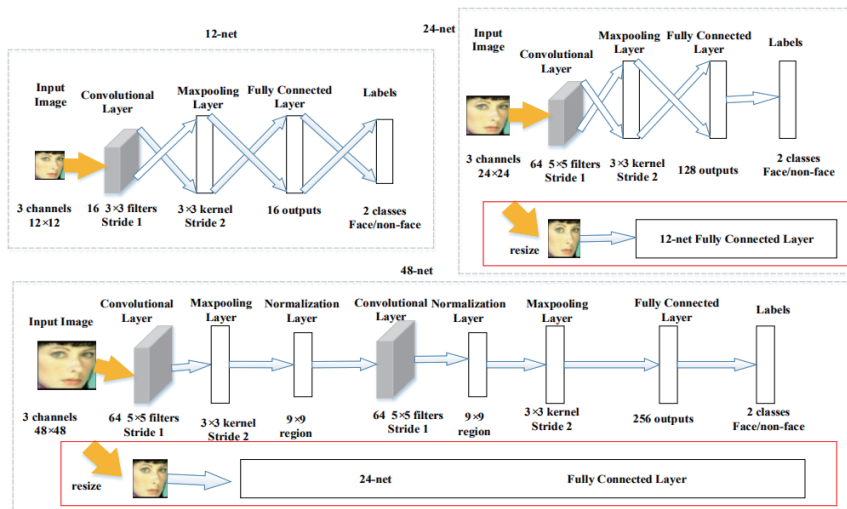


图 2 网络结构

1.2 特征提取

人脸检测特征提取是识别和定位图像中人脸区域的关键步骤之一。在处理人脸图像之前，需要先进行预处理，包括将图像转换为灰度、进行直方图均衡化以及进行归一化等操作，以增强图像的特征信息，减少光照和对比度变化对图像的影响。在提取特征前，通常需要先定位图像中的人脸。可以利用传统技术，比如基于 Haar 特征的 AdaBoost 分类器，也可以采用深度学习技术，比如 MTCNN 来实现。一旦人脸被定位，接下来就是提取特征。传统的特征提取方法包括 Gabor 滤波器、LBP、HOG 等，它们能够捕捉到人脸的局

部纹理和形状信息。深度学习方法则通过训练卷积神经网络（CNN）自动学习特征表示^[3]。深度学习中常用的特征提取方法包括直接法、残差法、映射法、复合法等，其中直接法是指使用 CNN 直接从输入样本图像中提取特征，然后进行人脸表情识别。直接法的优势是可以自主学习特征，不需要手工干预。残差法是将图片中的人脸区域信息分为表情信息和非表情信息（中性表情）两类，分别进行差值比较，然后将得到的差值作为表情特征用于表情识别。残差法的优势是可以加速网络训练和收敛，增加网络的鲁棒性和网络的准确率。映射法是将非峰值表情的图映射为峰值表情，再提取表情特征对表情识别。映射法的优势是降低数据的维度，提高网络的泛化能力。复合法指的是多种特征相结合进行表情判定。复合法的优势是可以利用多种网络的优点，提高网络准确率和鲁棒性。

1.3 表情分类

表情分类是面部表情识别流程过程中的最后一步。表情分类的基本思想是：将样本图像经过人脸检测确定样本中的人脸位置，提取出具有足够多的信息含量特征部分进行识别分类，分类类别一共设置为 7 类（生气、厌恶、恐惧、高兴、

中性、悲伤和惊讶），计算机通过提取出的这些特征选择一种类别作为输入样本图像中的表情类别归属。目前，许多国内外研究人员在进行表情分类时，主要采用支持向量机（SVM）或基于 CNN 的分类算法。

2 基于卷积网络注意力机制的人脸表情识别

目前，大多数基于深度学习的表情识别技术为得到更高的表情识别准确率，选择不断加深网络层数或者改进创新卷积网络模型^[4]。但是网络层的数量和宽度不断加深并不会让模型识别面部表情的准确率得到很大的提升，反而会出现很多问题，如梯度不稳定至消失、计算复杂且量大、

特征表达瓶颈和网络退化等，导致网络性能变差，增加表情识别的错误率等。为了解决这些难题，国内外的学者们已经提出了许多网络结构，比如 VGG、AlexNet、ResNet 等都被用作人脸表情识别领域研究的基础网络。

2.1 网络设计

人脸表情识别算法采用了卷积网络融合注意力机制、残差模块和细化模块。该方法由浅层特征抽取、注意力机制、中间特征抽取以及终端特征抽取 3 个层次构成。首先，将浅层图像中的卷积作为图像的“特征抽取”进行图像分割；注

意机理从信道与时空两个方面对关键区进行加权,降低了对关键区的扰动;基于此,本文提出了一种新的基于残差的神经网络方法,该方法可以有效地降低参数的个数,从而降低了模型的重量;同时,采用多层次、多层次的特征抽取、支路网络与主要网络相结合的方法,实现面部关键点的精细处理^[5]。同时,通过引入共同损耗的方法,增大类之间的间距,减小类内的间距,从而进一步提高识别精度。具体网络模型的流程如图3所示。

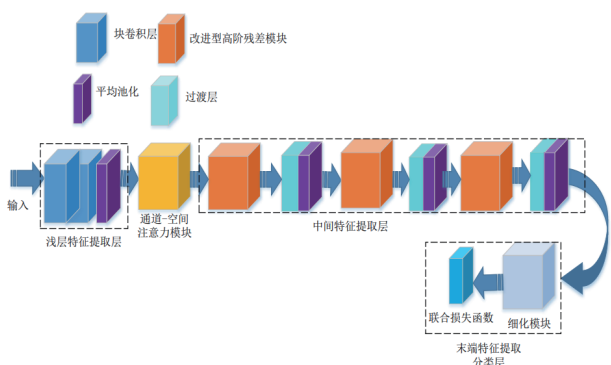


图3 网络模型结构

2.2 浅层特征提取层

离图像较近的神经网络为浅层次特征抽取层,其所抽取的特征是否丰富,将会对随后的图像进行中层及末端的特征抽取,从而对最终的识别精度产生重要的影响^[6]。卷积神经网络作为一种“特征抽取子”,能有效地从图像中提取出图像的局部特性。

2.3 注意力机制

在面部情绪识别领域,常规的卷积神经网络是通过对整个面部的整体结构来进行情绪的预测,但是容易被其它部件所影响,从而改变了整个面部的情绪识别性能。注意力机制的工作思路是采用掩码的形式重新赋予特征权重,权重大小各不相同,将注意力聚焦于关键点特征区域,以此得到更多相关表情的重要特征信息,忽略无关信息。为了让网络更准确地提取表情的关键点信息,本章借鉴CBAM的设计,提出了一种新的注意力机制:通道-空间注意力模块(CSAM)。该方法将注意的渠道与空间的注意机理有机地融合在一起,以丰富网络的结构,强化重要的特征^[7]。与CBAM相比,CSAM和空间注意力机制是并行相加的。其中,通道注意力机制利用全局平均池化层(GAP)压缩提取特征序列,而空间注意力机制经过平均池化和最大池化后各进行一次卷积,从而可以在进行图像融合前保留相对较多的通道信息。

2.4 中间特征提取层

本文提出一种基于深度学习的卷积神经网络模型。但是,

这种情况下,网络中的参数数目会增多,运算复杂度也随之提高。为了加强图像的抗噪性和更好地关注面部表情,本文在图像的中间特征抽取部分使用了3种改进的高阶残差模型。同时,用可分割的深度卷积代替原来的卷积层,扩大了神经网络的宽度,降低了参数的个数。该方法通过在两个修正后的高阶残差模型间增加一种转换层次,从而在保证模型维数不变的前提下,增强模型的转换性能。

2.5 末端特征提取分类层

在现实生活中,由于面部表情具有一定的普遍性,使得分类结果的准确性较低。为此,本文设计了人脸细节的精细处理模型,通过引入一种联合损耗函数,增大了人脸各类间的间距。

3 基于残差网络注意力机制的人脸表情识别

不同的网络模型各有各自独特之处,但也存在一些小问题,比如识别准确度不够高,模型参数量较大等^[8]。为此本章设计了一种网络模型:首先对图片裁剪掩码,进行数据增强处理;然后在改进的Ghost模型中加入了一个“压缩-激发”的环节,以降低系统的噪音干扰;利用多层次的空间注意力机理,进一步提高了图像的细节识别能力。在此基础上,提出了一种基于图像的人脸表情分类方法。

3.1 网络设计

本章以残差网络为基本网络,将切割mask模组、Ghost模组、通道注意力机制、多尺度空间关注机制与共同损耗功能整合于一体,具体流程见图4。在对样本集上的图像实施截掩码运算时,使用了一种随机遮挡方法,提高了遮挡的随机性,从而比只关注某一部分的局部信息更能充分地利用整体的信息;采用信道注意机制对信道进行了加权处理,提高了信道的重要程度;同时采用多层次的空间关注机理,结合多尺度和空间关注,获取更宽的感知野,实现对图像的纹理特征的提取。区别于常规的编码模型,提出了一种基于深度可分的卷积模型来代替数据包的结构,减少了模型中的参数量;并提出了一种基于压缩-激发(SE)的方法,该方法可以降低特征抽取时的噪音影响,是一种新的基于人脸特征的人脸识别算法。

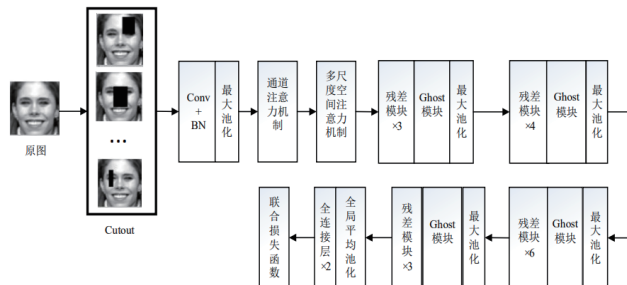


图4 网络模型结构

3.2 通道注意力机制

通道注意力机制（channel attention mechanism, CAM）是一种特征重标定技术，用于强调重要的通道特征并忽略不重要的通道特征。通过分析特征图内部的通道关系来实现。以下是通道注意力机制的提取过程：首先对特征图进行全局平均池化（global average pooling, GAP）和全局最大池化（global max pooling, GMP），得到两个 $1 \times 1 \times C$ 的特征图，分别代表所有位置的平均值和最大值特征。将这两个 $1 \times 1 \times C$ 的特征图通过一个共享的全连接层（或卷积层）进行特征压缩，通常使用降维系数 r 来减少特征维度，然后通过 ReLU 激活函数引入非线性^[9]。并通过平均池化和最大池化得到的权重进行合并（通常是相加），然后通过 sigmoid 激活函数生成每个通道的最终权重。最后，原始特征图的每个通道乘以其对应的权重，实现特征的重标定，增强重要特征并抑制不重要特征。

3.3 多尺度空间注意力机制

在对图像进行面部情绪识别时，通常采用的是一些重要的特征，如：眼睛、嘴唇、眉毛等。由于这些特性中蕴含着更多的纹路，所以当表现出各种情感的时候，面部表情也会随之改变，哪怕是最微小的变化，电脑也能很容易的发现^[10]。

在对样本图进行 Cutout 处理后，对脸部的一些特征进行了遮挡，使得脸部的特征点的提取成为了一个非常重要的问题。然而，由于各幅图像中的面部在整个图像中所占的比重不尽相同，因此，基于多尺度特性和空间关注机理，通过对各感知野权值的提升，实现对各幅图像中各节点的有效提取。具体流程如图 5 所示。

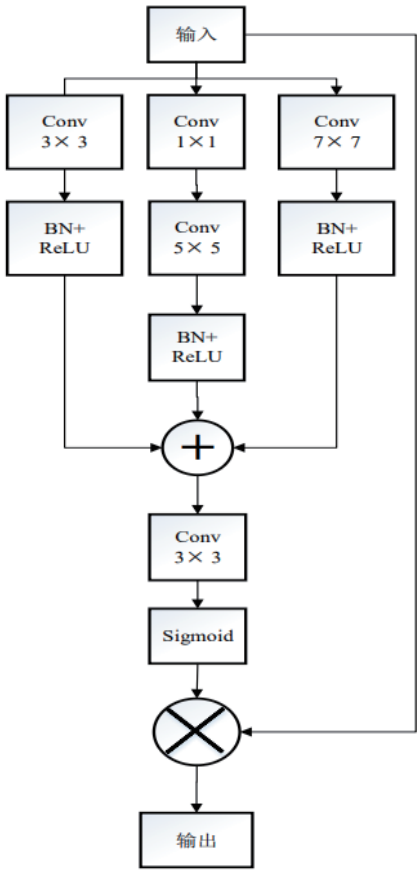


图 5 多尺度空间注意力机制

4 结语

本文通过具体分析人脸表情识别的三个步骤（人脸检测、特征提取和表情分类），针对每一步骤分别探讨了传统算法和基于深度学习算法，在理论上为后续研究打下了基础。除此之外，提出了一个卷积神经网络注意力机制的网络模型，在卷积网络中引入了通道空间注意力机制用于权重分配，以关注表情区域的细微差异特征信息；同时还加入了改进的高阶残差模块，从而减少参数量同时增强对表情区域的关注，提高了网络的识别能力。提出的基于残差网络注意力机制的人脸表情识别方法，以残差网络为基础网络，加入裁剪掩码对图像任意区域的任意大小进行遮挡，提高网络的鲁棒性。

参考文献：

[1] 李阳. 基于多特征融合的面部表情识别方法研究 [D]. 西安: 西安科技大学, 2017.

[2] 褚晶辉, 汤文豪, 张姍, 等. 一种基于注意力模型的面部表情识别算法 [J]. 激光与光电子学进展, 2020, 57(12): 197-204.

[3] 赵家琦, 周颖玥, 王欣宇, 等. 采用支路辅助学习的人脸表情识别 [J]. 计算机工程与应用, 2022, 58(23): 151-160.

[4] 张波, 兰艳亭, 李大威, 等. 基于卷积网络通道注意力的人脸表情识别 [J]. 无线电工程, 2022, 52(1): 148-153.

[5] 罗岩, 冯天波, 邵洁. 基于注意力及视觉 Transformer 的野外人脸表情识别 [J]. 计算机工程与应用, 2022, 58(10): 200-207.

[6] 刘全明, 辛阳阳. 端到端的低质人脸图像表情识别 [J]. 小型微型计算机系统, 2020, 41(3): 668-672.

[7] 罗翔云, 周晓慧, 付克博. 基于深度学习的人脸表情识别 [J]. 工业控制计算机, 2017, 30(5): 2.

[8] 王信, 汪友生. 基于深度学习与传统机器学习的人脸表情识别综述 [J]. 应用科技, 2018, 45(1): 8.

[9] 焦阳阳, 黄润才, 万文桐, 等. 基于图像融合与深度学习的人脸表情识别 [J]. 传感器与微系统, 2024, 43(3): 148-151.

[10] 陈曦, 蔡现龙. 基于深度学习的人脸局部遮挡表情动态识别算法 [J]. 吉林大学学报 (信息科学版), 2024, 42(3): 503-508

【作者简介】

路晓亚（1983—），女，河南商丘人，硕士，副教授，研究方向：计算机技术。

（收稿日期：2024-08-09）