



北京科技大学
University of Science and Technology Beijing

密级： 公开

本科生毕业设计(论文)

题目： 基于半监督学习的
面部动作单元识别

作者： 赵相尧

学号： 41607668

学院： 计算机与通信工程学院

专业： 计算机科学与技术

成绩： 91

2020 年 05 月

声 明

本人郑重声明：所呈交的论文是本人在指导教师的指导下进行的研究工作及取得研究成果。论文在引用他人已经发表或撰写的研究成果时，已经作了明确的标识；除此之外，论文中不包括其他人已经发表或撰写的研究成果，均为独立完成。其它同志对本文所做的任何贡献均已在论文中做了明确的说明并表达了谢意。

学生签名： 赵相尧 2020年5月25日

导师签名： _____ 年__月__日

摘 要

人脸面部动作单元 (AU) 识别在表情识别领域起着至关重要的作用, 是计算机主动理解人类表情的重要途径。传统的 AU 识别使用的参数量较小的监督学习算法, 数据量需求不大。但随着人们转而使用识别能力更强, 参数也更多的深度网络, 该任务所需的数据量大大提升。然而监督学习算法依赖于已有 AU 标注的数据, 而 AU 的标注是个十分费事费力的过程。因此, 如何使用无 AU 标注的数据训练算法就成为了一个非常有意义的课题。

本文对国内外 AU 识别的研究现状进行了综述, 在此基础上, 结合多任务学习提出了一个半监督学习框架, 即将两个相关任务中的一个的参数固定, 同时提供两个任务的数据进行训练, 这相当于对未固定参数的任务提供了半监督信息, 从理论上来说, 该框架可以提高未固定参数的任务的性能。并对如何固定其中一个任务的参数提出了两种方法, 采用伪样本和真实样本进行预训练。

随后本文将此框架运用于 AU 识别之上, 将 AU 识别与情感识别作为两个相关任务。经过测试, 本文发现该框架可以在不扩大训练集的情况下, 提升模型的泛化性能, 并且模型的参数越多, 该框架对性能的提升幅度越明显。

基于上述模型, 本文还实现了一个实时的 AU 检测系统。该服务部署在本地, 可以从外部设备获取视频信号, 并利用后端的推断模型对输入图像进行人脸的检测、切割和 AU 的识别, 随后推断模型将检测结果交付服务器进程, 最后通过浏览器渲染的页面将检测结果即时反馈给用户。

关键词: 人脸面部动作单元, 半监督学习算法, 多任务学习, 半隐任务学习

Facial Action Unit recognition based on semi-supervised learning

Abstract

Facial Action Unit (AU) recognition plays an important role in the field of expression recognition, and it is an important way for computers to understand human expression actively. The traditional AU recognition uses a supervised learning algorithm with a small number of parameters, which requires little data. However, as people turn to the deep network with stronger recognition ability and more parameters, the amount of data required for this task is greatly increased. However, the supervised learning algorithm relies on the existing AU labeled data, whose annotating process is very laborious. Therefore, how to train a classifier without AU annotation has become a very meaningful topic.

In this paper, the research status of AU recognition at home and abroad is reviewed. On the basis of those research, a semi-supervised learning framework is proposed combining multi-task learning, that is, the parameters of one task in two related tasks which can be connected are fixed, and the data of the two tasks are provided for training at the same time, which is equivalent to semi-supervised information for tasks whose parameters are not fixed. Theoretically, the framework can improve the performance of latter task. Two methods of how to fix the parameters of one task are put forward that pseudo samples or real samples are used for pre training.

Then this framework is applied to AU recognition. After testing, we find that the framework can improve the generalization performance of the model without expanding the training set, and the more parameters of the model, the more obvious the performance improvement of the framework.

Based on the above model, this paper also implements a real-time AU detection system. The service is deployed locally, and can obtain video signals from external devices, and use the back-end inference model to apply face detection, cutting and AU recognition on the input image. Then the inference model will deliver the detection results to the server process, and finally the detection results will be immediately fed back to the user through the browser rendered page.

Key Words: Facial Action Unit, Semi-supervised learning, Multi-task learning, Semi-hidden-task learning

目 录

摘 要.....	I
Abstract.....	III
1 引 言.....	1
1.1 研究背景及意义.....	1
1.2 研究现状.....	1
1.3 本文研究方法 with 目标.....	3
1.4 本文组织结构.....	4
2 相关知识与技术.....	6
2.1 面部动作单元相关基础知识.....	6
2.1.1 面部动作单元的概念.....	6
2.1.2 面部动作编码系统.....	6
2.2 人脸图像特征提取方法.....	7
2.2.1 哈尔特征.....	7
2.2.2 方向梯度直方图特征.....	9
2.2.3 人脸几何特征.....	11
2.2.4 卷积神经网络.....	11
2.3 半监督学习.....	14
2.4 多任务学习.....	15
3 基于半监督学习的面部动作单元识别框架.....	17
3.1 结合半监督学习与多任务学习的半隐任务学习.....	17
3.2 显任务参数的预先确定.....	20
3.2.1 基于伪样本采样的显任务参数预训练.....	20
3.2.2 使用真实标签预训练显任务参数.....	23
4 面部动作单元识别系统的实现.....	25
4.1 AU 识别系统的整体方案.....	25
4.2 图像的获取.....	26
4.3 人脸检测与切割.....	26
4.4 人脸特征提取.....	27
4.4.1 人脸特征点提取.....	27
4.4.2 HOG 特征提取.....	28
4.4.3 卷积神经网络人脸特征提取.....	29
4.5 输入特征的预测.....	31

4.6 用户交互界面.....	31
4.7 训练阶段使用的双输入双输出模型的搭建.....	32
5 5.实验与分析	35
5.1 实验准备.....	35
5.1.1 实验环境.....	35
5.1.2 评价指标.....	35
5.1.3 数据准备.....	36
5.2 显任务参数的预训练.....	37
5.3 AU 检测对比实验.....	39
5.3.1 使用人脸特征点的 AU 检测	39
5.3.2 使用 HOG 特征的 AU 检测	42
5.3.3 使用卷积神经网络的 AU 检测	44
5.4 结果对比.....	46
5.5 本章小结.....	46
6 总结与展望	48
6.1 本文工作总结.....	48
6.2 未来工作展望.....	48
参考文献	49
在学取得成果	53
致 谢	55

1 引言

1.1 研究背景及意义

随着社会经济和科学技术的发展,人们并不满足于传统的语言命令交互、图形用户界面等人机交互模式,对人机交互提出了越来越高的要求,机器对人类情绪的主动理解成为了一个重要的目标。

计算机可以通过面部表情、语音或脑电等对人类的情绪进行识别,但目前技术上最成熟、应用最广泛的还是面部表情识别技术^[1]。1968年,美国科学家 Albert Mehrabian 提出情绪表达中有 55%的信息是通过面部表情传递的^[2],这更为表情识别的重要性提供了理论依据。

1972年,Paul Ekman 提出人类有六种基本表情:生气、厌恶、恐惧、高兴、伤心、惊讶^[3]。然而这六种基本表情无法完全描述人类复杂的情绪,人们会表现出更多的复合表情,因此研究人员使用面部动作单元(Action Units, AU)去描述人类的面部表情。AU在表情识别领域起着至关重要的作用,它可以描述人脸不同区域的动作,从而捕获人类各种复杂情感的细微差异信息。

AU识别可以辅助对精神分裂症、抑郁症等疾病的判定,有助于对消费者在购买过程中的反应以及体验进行评估,同时还可以实时监测驾驶员的疲劳状态。因此,AU识别的应用十分广泛,在医疗健康、广告传媒、安全监管等多个领域都有着良好的应用前景。

以前的AU识别使用参数量较小的监督学习算法,所需的数据较少^[8]。随着时代的发展,人们转而使用识别能力和泛化能力更强的大规模深度网络,参数增多,所需的数据量大大提升。然而这带来了一个问题,有监督学习算法依赖于已有AU标注的数据,但为大量图像进行AU标注是一个十分费时费力的过程,而且需要熟知FACS的专家才行。训练一个可以识别AU的有素专家大约需要100小时,而通常对一分钟的视频进行AU标注还需要2小时左右^[4]。因此,如何使用无AU标注的数据对AU识别就成为了一个非常有意义的课题。

1.2 研究现状

由于AU标注的困难性,以数据为驱动的方法很难进一步提升AU识别的效果。而由于AU会组合出现,因此可将AU识别看作一个多标签学习问题。多标签学习的一个核心问题是捕获各标签间的关系,对于AU识别来说,

各 AU 由于肌理上的原因存在显式的空间关系，同时与表情也存在概率依赖关系。所以近些年来以知识为驱动的方法越来越多，成为一个新的研究方向。

下面按照标签缺失程度的从小到大对面部动作单元识别方法进行介绍。

(1) 有标签的 AU 识别

最直接的方法是基于图像特征对每个 AU 进行独立识别。人们利用追踪特征点、Gabor 变换、主动外观模型 (Active Appearance Model, AAM) 等方法提取面部特征，然后使用支持向量机 (Support Vector Machine, SVM) [5]、线性链式条件随机场 [6] 等方法来检测每一个 AU 是否出现。这些方法需要为每一个 AU 单独建立一个识别器，并且没有利用任何有关面部的先验知识。

另一种方法是对 AU 组合进行识别。如 Lucey 等人 [7] 使用 AAM 和 SVM 识别几种特定的 AU 组合。这种方法的缺陷在于 AU 组合后标记空间迅速膨胀，而且只能识别事先人为给定的 AU 组合，因此只适用于需要检测少量特定 AU 组合的情况。

考虑到 AU 之间具有共生和互斥等内在联系，人们提出了基于模型的 AU 识别方法，如使用概率图模型如动态贝叶斯网络 (Dynamic Bayesian Network, DBN) [8]、受限玻尔兹曼机 (Restricted Boltzmann Machine, RBM) [9] 等来捕获 AU 之间的概率关系，以提高识别器的效果。

以上提到的这三类方法都必须使用带有 AU 标注的数据，因此不适用于无 AU 标注的情景。

(2) 有表情标签辅助的 AU 识别

由于表情可以被编码成 AU 的组合，当大量样本没有 AU 标注的时候，表情标签可以弥补 AU 标签缺乏的不足，并且对表情的标注更加容易，所以表情辅助可以有效降低 AU 识别的成本。

基于表情辅助的 AU 识别的一个基本思路是使用概率图模型，将表情看作 AU 之外的另一个节点，这个节点是多值的（也可以用多个二值节点将表情编码）。DBN 和 RBM [10] 可以有效地捕获 AU 节点和表情节点依赖关系的概率图模型。

(3) 标注不完整的 AU 识别

AU 标注不完整有两种情况，一种情况是单个样本 AU 标注的部分分量缺失，另一种情况是单个样本无任何 AU 标注但含有其他信息如表情，在现实问题中，这两种情况都有可能出现。

对于部分样本有完整的 AU 标注和表情标注的情形，可以先利用贝叶斯网络、分组稀疏诱导先验压缩感知、受限玻尔兹曼机等方法在拥有完整标注

的样本上捕获 AU 和表情之间的约束关系，随后利用最大后验推理^[11]、标记分配之间的平滑度和预测标记与对应真实标记之间的一致性^[12]、先验分布与合页损失共同约束^[13]等方法补全其他样本缺失的标注。这些方法用到的 AU 先验分布和标签间的一致性或平滑性都是从标注完整的样本中学到的，如果标注完整的样本过少，则学习到的 AU 和表情间的约束就不能反应普遍规律。

对于没有 AU 标注的情形，就需要从领域知识中获取表情与 AU 之间的关系。如可以先根据前人总结的各表情下每个 AU 发生的概率生成伪样本，再利用 RBM 从伪样本中学习 AU、表情之间的约束^[11]。

1.3 本文研究方法 with 目标

本文旨在设计并实现一个 AU 实时检测系统，它是一个针对输入图像的特征提取和判别预测系统。但与传统的监督学习判别模型不同的是，该系统在建立过程中引入了半监督学习的过程，从而在不扩大训练集的情况下提高了判别模型的泛化能力（图 1-1）。具体研究主要包括如下几个方面：

（1）半监督学习框架

半监督学习算法很多，本文需要搭建一个符合本文研究任务的半监督学习框架，阐明其数学原理，再将该框架运用于具体的问题，指导本文在建模过程中运用未标注的数据，实现模型泛化能力的提升。

（2）人脸特征的提取

在将输入送进判别模型之前，需要先提取图像的特征，以减少与任务无关的噪声对判别结果的影响，还可以降低输入数据的维度，提升模型的运算速度。本文使用传统机器学习方法，分别对人脸的几何特征和纹理特征进行提取，还使用了深度卷积神经网络对图像特征进行提取。

（3）特征判别模型

模型对输入的特征进行判别，并输出对每个标签进行预测的概率。在模型的训练阶段，基于搭建的半监督学习框架，引入半监督信息，提升其性能。

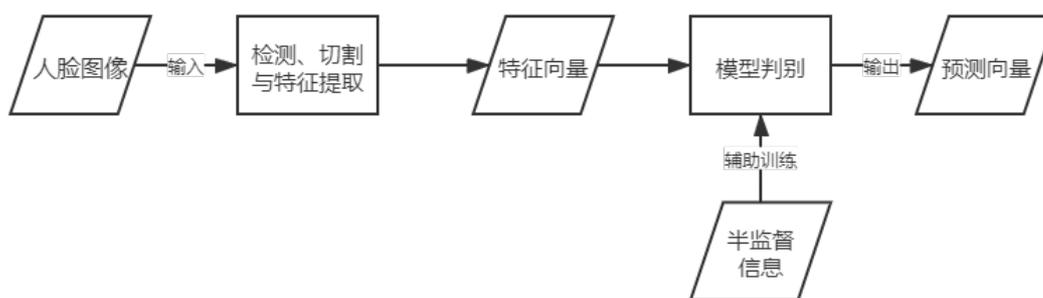


图 1-1 模型整体结构示意图

本文的整体研究目标为，提升在小数据集上训练的模型的泛化能力。具体目标为，在小人脸数据库上，利用半监督学习算法，在不引入其他数据的情况下，提升 AU 检测模型的泛化能力。

模型借助 TensorFlow、OpenCV 等框架，使用 Python 语言进行实现。人脸检测使用了基于 Haar 特征的级联分类器，人脸特征的提取分别使用了基于级联回归树的人脸特征点坐标、方向梯度直方图特征和深度卷积神经网络。判别模型使用了多层感知机模型。

1.4 本文组织结构

包括引言和结论，本文共分为六个章节，其组织结构如下：

(1) 第一章，引言。

本章包含本文的研究背景及意义、研究现状、本文研究方法与目标和本文组织架构四个小节，解释了本研究所处的背景，阐明了本研究的现实意义，介绍了前人在此话题下做出的研究，并对本文的研究方法与目标做出了说明，最后对本文的组织架构进行了梳理。

(2) 第二章，相关知识与技术。

本章介绍了本文所研究的具体问题的相关知识和技术。本章首先介绍了面部动作单元的定义，使读者对本文设计所要检测的目标有一个清晰的认识。随后本章介绍了人脸图像特征的提取的方法，使读者了解本文在设计中使用到的算法的具体流程。最后本文介绍了半监督学习和多任务学习的概念，为第三章的内容进行铺垫。

(3) 第三章，基于半监督学习的面部动作单元识别框架。

本章结合半监督学习和多任务学习提出了半隐任务学习的学习框架，并在推导过程中结合本文所要解决的具体问题进行了解释，使读者更明白每一

步的用意。随后，本文对框架推导过程中新出现的一个问题提出了两种解决方案。

(4) 第四章，面部动作单元识别系统的实现。

在提出了模型的学习框架后，就需要结合具体问题对模型进行实现。本章首先从俯视角介绍了 AU 识别系统的整体方案，随后具体介绍模型的每一部分的实现方法。

(5) 第五章，实验与分析。

本章使用第四章实现的模型，在一个小数据集上利用三种不同的特征提取方法，对半监督学习是否提高了模型的效果进行了验证。

(6) 第六章，总结与展望。

本章对本文的研究成果进行总结分析，并对未来的工作进行展望。

2 相关知识与技术

本章将首先对面部动作单元和其编码系统进行介绍，使读者对本文所要研究的问题有一个具体的了解，随后本章将对本文使用到的人脸图像特征的提取办法进行介绍，最后本章介绍了半监督学习和多任务学习的概念，这是本文所使用的半监督学习框架的基础。

2.1 面部动作单元相关基础知识

2.1.1 面部动作单元的概念

AU 的定义由 Ekman 等人^[14]于 1978 年在其提出的面部动作编码系统 (Facial Action Coding System, FACS) 中给出，依据面部解剖学原理，FACS 将每个人脸局部肌肉的运动都描述成一个 AU，随后将所有解剖学上可能的面部表情编码成不同 AU 的组合。还有一些其他与面部肌肉运动和人类情感相关的编码系统，比如 EMFACS^[15]、FACSAID^[16]和 Affex^[17]。

AU 组合可以分为加性组合和非加性组合^[16]。在加性组合当中，每个 AU 的外观变化是相互独立的，不会改变每个 AU 的基本特征。例如，AU1 的外观变化是抬起眉毛内角，AU2 的外观变化是抬起眉毛外角，而 AU1+2 使得眉毛的内外侧向上拉起，是 AU1 和 AU2 共同作用的结果。非加性组合会使某个 AU 的原始特征发生变化。例如，AU1 单独出现时眉毛内侧向上拉起，AU4 单独出现时眉毛会压低并聚拢，而 AU1+4 的作用效果是眉毛内侧被抬起并被拉到一起，AU4 的原始特征被改变。

2.1.2 面部动作编码系统

FACS 可以将解剖学上人类能做出的所有表情解构成 AU 组合及其产生的时间段，每个 AU 对应一个或多个面部肌肉的收缩或松弛。对于那些尚未明确肌肉运动基础的动作，FACS 将其称为动作描述符 (Action Descriptors, AD)。

FACS 是一个通过面部表情对面部动作进行划分的系统，它最早由瑞典解剖学家 Carl-Herman Hjortsjö^[18]提出，随后被 Ekman 等人^[14]采纳完善，该版本的 FACS 将人脸划分成不同的 44 个动作单元，主要由与面部运动相关的 30 个 AU 和 14 个 AD 组成。2002 年，Ekma 等人和 Hagar^[16]共同对 FACS 版

本进行了修订，共包含 41 个 AU，其中 27 个是与面部肌肉运动相关的 AU，14 个是没有指定与面部动作相关的杂项 AU。

目前，FACS 已经成为了对面部表情进行系统性分类的通用标准，并在心理学和动画领域被证明是有效的^[19]。

2.2 人脸图像特征提取方法

2.2.1 哈尔特征

哈尔特征 (Haar-like features) 是第一种可用于即时的人脸检测运算的特征，由 Papageorgiou 等人^[20]提出，其基本思想是并不使用图像的强度 (即图像每一个像素点的 RGB 值) 作为特征，而是使用哈尔小波变换后的结果作为特征，该特征也因此得名。

其特征计算方法为利用特征模板建立一个检测窗口，在窗口的指定位置计算每一个矩形区域的像素和，并取其差值。图 2-1 列示了特征计算时使用的特征模板，这些模板共可以提取四种特征，即边缘特征、线性特征、圆心环绕特征和特定方向特征，每个模板的特征值为其白色矩形区域像素和减去其黑色矩形区域像素和。

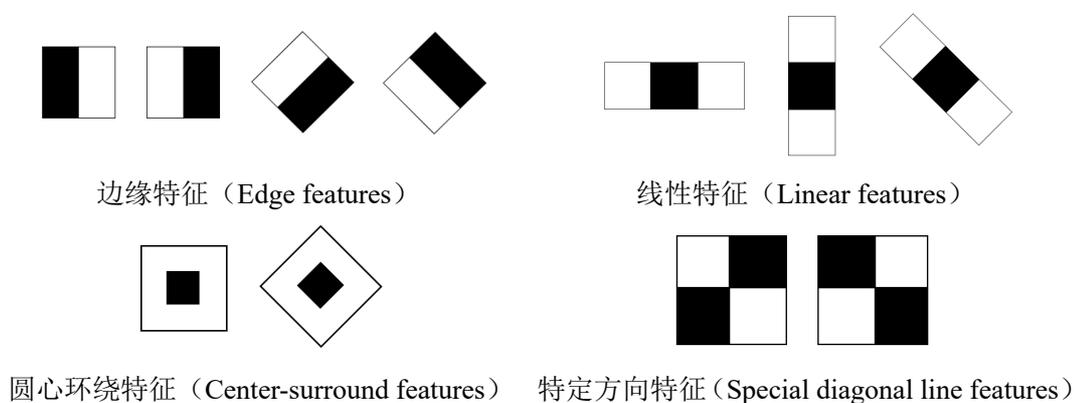


图 2-1 哈尔特征的特征模板

这些模板可以任意改变其比例和尺寸，例如，在一个 24×24 的图片上，就可以构建出一万六千多种检测窗口。这些检测窗口将在输入图像上滑动，在图像的每一个子区域都计算一个哈尔特征值。然后这个特征值会与一个预先计算好的阈值进行比较，将目标和非目标区分开来。以图 2-2 为例，第二列的第一行展示了一种边缘特征的特征模板，它区分的目标为眼睛，如第二行所示，当其在图像上滑动至眼睛的位置时，由于眼睛区域较黑，两颊区域较白，因此其计算出的特征值超过了阈值，因而该检测窗口认为该位置存在

目标，即眼睛；而第三列的第一行展示了一种线性特征的特征模板，它区分的目标也是眼睛，当它滑动至眼睛的位置时，由于眼睛区域较黑，鼻梁区域较白，其计算出的特征值超过了阈值，因而认为该位置存在眼睛。

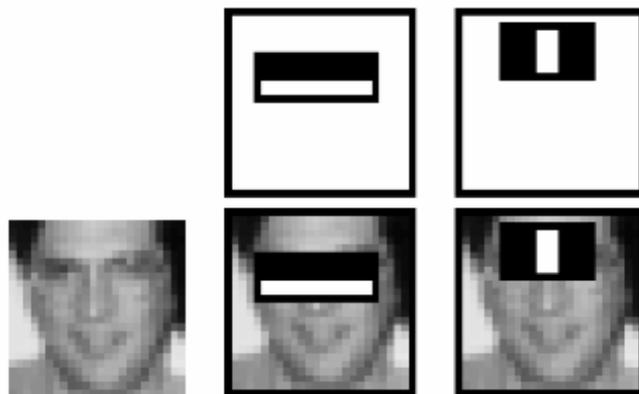


图 2-2 哈尔特征检测窗口对人脸局部特征的检测^[21]

Viola 和 Jones[21]利用哈尔特征实现人脸检测，一个哈尔特征对应一个弱分类器，它只能检测人脸的局部特征，检测人脸的正确率仅仅比随机猜测强一点点。为了达到一个可信的判断，在其检测框架中将一大群这样的特征组合成一个级联分类器，最终形成一个强分类群。

构建级联分类器的算法为 Adaboost 算法，其核心思想为只要一个分类器的检测效率在 50%以上，通过不断的判读最终会得到一个理想的结果。其算法具体描述为：

算法 2-1 Adaboost 算法

输入： 训练集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n), y_i \in \{0, 1\}\}$

输出： 一个强分类器

1: 初始化：设置样本权重 $w_{1,i} = \frac{1}{n}$ ，其中 n 为样本个数

2: **for** $t = 1, 2, \dots, T$ **do**

3: 归一化权重： $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$

4: 对每一个特征 f ，训练一个分类器 $h_f(x)$ ，该特征在训练集上的加权错误率为： $\epsilon_f = \sum_i w_{t,i} |h_f(x_i) - y_i|$

5: 找出加权错误率最小的分类器 $h_k(x)$ ，其加权错误率为 ϵ_k

6: 为每个样本更新权重 $w_{t+1,i} \leftarrow w_{t,i} \beta_k^{1-\epsilon_i}$ ，其中 $\beta_k = \frac{\epsilon_k}{1-\epsilon_k}$ ，若样本 i 被正确

分类，则 $\epsilon_i = 0$ ，否则 $\epsilon_i = 1$

7: **end for**

最终得到的强分类器为:

$$h(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t, \text{ 其中 } \alpha_t = \log\left(\frac{1}{\beta_t}\right) \\ 0 & \text{其他} \end{cases} \quad (2-1)$$

每次提取哈尔特征时都需要计算矩形区域内的像素之和, 严重拖累了人脸检测的速度。可以考虑为图像建立积分图, 记为 I , 即每一个元素是原始图像在对应位置左上角所有像素的和, 这样在计算矩形 $ABCD$ 内的像素和时, 只要进行四次查表, 大大节省了检测的时间。

$$\text{像素和} = I(A) + I(C) - I(B) - I(D) \quad (2-2)$$

同时, 还可以采用筛选式级联方法, 将各个弱分类器按其权重(权重代表了分类器的强弱)从大到小排序后分割为几组, 每张人脸图像在某一组分类器出被判定为负类后就不再进入下一组, 进一步减少了检测的时间。

2.2.2 方向梯度直方图特征

方向梯度直方图(Histogram of Oriented Gradients, HOG)特征是计算机视觉、模式识别领域很常用的一种描述图像局部纹理的特征, 其概念最早由 Robert K. McCannell 在其专利中提出^[22]。它由图像的局部区域的方向梯度直方图组成。其基本思想是, 在一副图像中, 局部目标的表象和形状能够被梯度或边缘的方向密度分布很好地描述。

HOG 的计算包含四个基本步骤:

(1) 伽马和色彩的归一化

为了减少光照因素的影响, 首先需要将整个图像进行归一化。在图像的纹理强度中, 局部的表层曝光贡献的比重较大, 而这种压缩处理能够有效地降低图像局部的阴影和光照变化。但在实际实验中, 发现该处理对最终结果并没有提升^[23], 因此在实践中这一步一般都被跳过。

(2) 梯度计算

随后使用一维算子 $[-1,0,1]$ 及其转置分别计算每个像素处的梯度在 x 方向和 y 方向上的分量(式 2-3 和 2-4), 利用梯度的计算公式计算梯度的模值(式 2-5), 可用近似公式提高计算速度, 同时计算梯度的方向角(式 2-6), 由于方向角的符号并没有意义, 因此一般采用无符号的方向角, 将其映射至 $(0, \pi)$ 区间(式 2-7)。

$$G_x(x, y) = I(x + 1, y) - I(x - 1, y) \quad \forall x, y \quad (2-3)$$

$$G_y(x, y) = I(x, y + 1) - I(x, y - 1) \quad \forall x, y \quad (2-4)$$

$$M(x,y) = \sqrt{G_x^2(x,y) + G_y^2(x,y)} \approx |G_x(x,y)| + |G_y(x,y)| \quad (2-5)$$

$$\theta(x,y) = \arctan\left(\frac{G_x(x,y)}{G_y(x,y)}\right) \quad (2-6)$$

$$\theta(x,y) = \begin{cases} \theta(x,y) + \pi & \theta(x,y) < 0 \\ \theta(x,y) & \text{其他} \end{cases} \quad (2-7)$$

(3) 划分细胞单元并统计梯度直方图

接着将图像分成小的连通区域，这些连通区域被叫做细胞单元（cell）。以方向角为分组依据，按模加权统计每个 cell 内不同梯度的个数，形成每个 cell 的梯度直方图。具体操作为，划分分组，如将 $(0, \pi)$ 划分成长度为 $\pi/9$ 的9个区间，随后在每个区间内将梯度的模累加，这样就得到了一个9维向量，即为该 cell 的描述符。

(4) 划分块并进行对比度规范化

将每几个 cell 组成一个块（block），图 2-3 展示了一个由 4 个 8×8 像素组成的 cell 组成的 block。

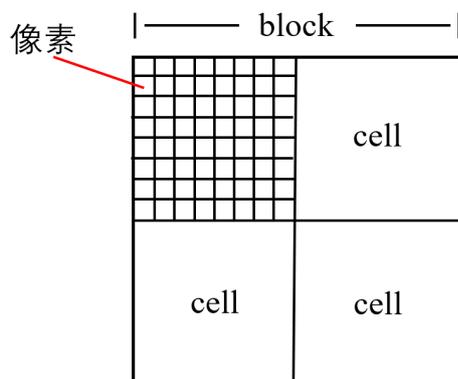


图 2-3 cell 和 block 示意图

将 block 中每个 cell 的描述符串联起来，并对串联起来的向量执行规范化，这个向量就是该 block 上的 HOG 特征。设 v 为原始向量，则其规范化公式为^[24]：

$$v \leftarrow \frac{v}{\sqrt{\|v\|_2^2 + \epsilon^2}}, \text{ 其中 } \|v\|_k \text{ 指 } k \text{ 范数} \quad (2-8)$$

而图像上的所有 block 的 HOG 特征串联起来即为该图像的 HOG 特征。

2.2.3 人脸几何特征

人脸由眼睛、鼻子、嘴、下巴等局部构成，对这些局部和它们之间结构关系的几何描述，可作为识别人脸的重要特征，这些特征被称为几何特征。对人脸几何特征进行提取的方法有很多，最常见的算法为主动形状模型（Active Shape Model）和主动表观模型（Active Appearance Model）^[25]。

但本文在这里介绍一种基于级联回归树的人脸特征点（一系列可以表征人脸的整体形状和其重要位置的点的集合）提取算法^[26]，该算法的精度较传统方法更高，且可在毫秒内完成对特征的提取。回归树是决策树的一种，全称为分类与回归树（Classification and Regression Tree, CART），它是一棵二叉决策树，这样的决策树等价于递归地二分每个特征，将输入空间划分为有限个单元^[27]。CART 给定输入随机变量 X 后输出随机变量 Y 的条件概率分布：当其用于分类时，其叶子结点特征的取值为“正类”和“负类”；当其用于回归时，其叶子节点为一具体的数值。最后，CART 在这些叶子节点上确定预测的概率分布。级联回归树的思想在于后一个回归树将原始特征和前一个回归树的输出作为输入，对前一次的估计结果做出一些修改。记原始图像为 I ，第 t 个回归树为 r_t ，其输出为 \hat{S}^t ，则：

$$\hat{S}^{t+1} = \hat{S}^t + r_t(I, \hat{S}^t) \quad (2-9)$$

将其用于人脸特征点的提取时，首先需要提供一系列已经标注好人脸特征点坐标的人脸图像，假定训练集为 $\{(I_1, S_1), \dots, (I_i, S_i), \dots, (I_n, S_n)\}$ ，其中 I_i 为脸的图像，而 S_i 为脸特征点坐标组成的向量。对训练集进行归一化处理使用梯度下降树（Gradient Tree Boosting）算法学习第一个回归树，而之后的每一个回归树都是在前一个回归树的基础上进行训练，更新公式为：

$$\hat{S}_i^{t+1} = \hat{S}^t + r_t(I, \hat{S}^t) \quad (2-10)$$

$$\Delta \hat{S}_i^{t+1} = S_{\pi i} - \hat{S}_i^{t+1} \quad (2-11)$$

其中 πi 为对训练集的一次随机取样。当 T 个回归树组成的级联回归树的总损失（此处使用均方误差损失）下降到预先设定的程度时，迭代终止。

2.2.4 卷积神经网络

卷积神经网络（Convolutional Neural Networks, CNN）是一类包含卷积计算且具有深度结构的前馈神经网络，是深度学习的代表算法之一。它仿造生物的视知觉机制构建^[28]，得益于其层内的卷积核参数共享和层间连接的稀疏性，它能够以较小的计算量提取格点化特征，有稳定的效果且对数据没有额

外的特征工程要求^[29]。它相较于传统的图像处理算法的优点之一在于避免了对图像复杂的前期预处理过程，尤其是人工参与图像预处理过程，可以直接输入原始图像进行一系列工作。其已经在图像分析和处理领域取得了众多突破性的进展，包括图像特征提取分类、场景识别等。

卷积神经网络的主要结构有：

(1) 输入层

卷积神经网络的输入层可以处理多维数据。常见地，一维卷积神经网络的输入层接收一维或二维数组；二维卷积神经网络的输入层接收二维或三维数组；三维卷积神经网络的输入层接收四维数组。在本研究中，卷积神经网络的输入层为三维输入数据，即平面上的二维像素点和 RGB 通道。

(2) 卷积层

卷积层的功能是对输入数据进行特征提取，其内部包含多个卷积核，卷积核个数称为卷积层的深度，其覆盖区域被称为“感受野”，与前一层中位置接近的区域的多个神经元相连。组成卷积核的每个元素都对应一个权重系数和一个偏差量，在工作时，它会有规律地扫过输入特征，在感受野内对输入特征做矩阵元素乘法求和并叠加偏差量^[30]（图 2-4）。

这个过程由卷积核大小、步长和填充控制。卷积核大小就是其覆盖的区域，可以指定为小于输入图像尺寸的任意值，卷积核越大，可提取的输入特征越复杂^[31]。卷积步长定义了卷积核相邻两次扫描时跨过的距离，卷积步长为 1 时，卷积核会逐个扫过特征图的像素，步长为 n 时会在下一次扫描跳过 $n-1$ 个像素^[32]。由于经过一次卷积运算后，特征图的尺寸会减小，因此需要在特征图通过卷积核之前人为增大其尺寸，以抵消计算中的尺寸减小，这就是填充，常见的填充方法为按 0 填充和重复边界值填充。

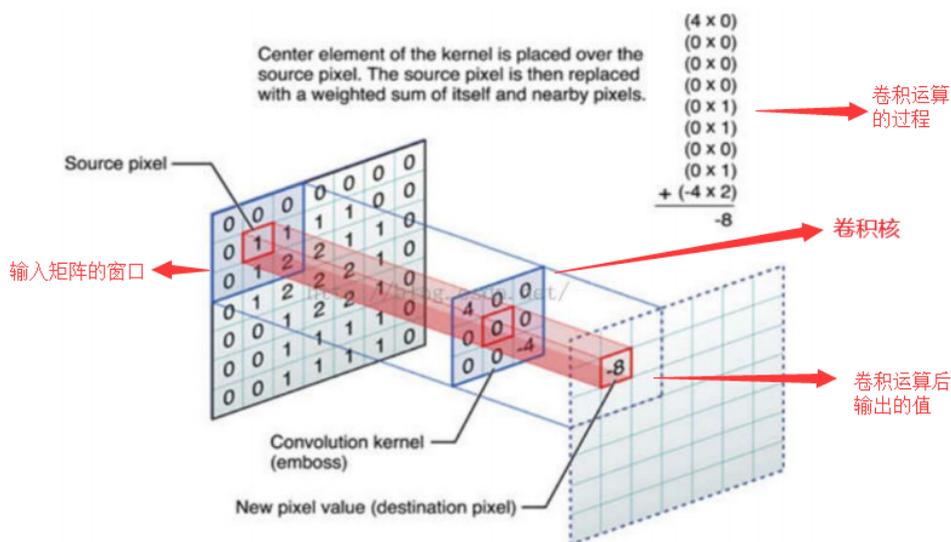


图 2-4 卷积核移动示意图^[33]

在卷积运算结束后，可以采用激励函数将卷积运算后的特征图的每一个像素进行映射。

$$a_{i,j,k}^l = \sigma(z_{i,j,k}^l) \quad (2-12)$$

其中, i,j,k 为像素的坐标, l 为当前卷积层, z 表示卷积运算的输出, a 表示激活函数运算后的输出, σ 为激活函数。

(3) 池化层

卷积层输出的特征图会被传递至池化层进行特征选择和信息过滤。池化层包含预设定的池化函数，其功能是将特征图中单个点的结果替换为其相邻区域的特征图统计量。池化过程的控制参数与卷积层类似，分别为池化大小、步长和填充控制。

此处介绍最大池化（max-pooling）函数。最大池化即取局部接受域中值最大的点。如下图所示是一个 2×3 大小的池化窗口，其锚点为 $[0,1]$ ，当其在特征图 x 上以1为步长移动时，它将锚点处的值设置为整个池化窗口的最大值，在其运行到边缘部分时，边缘外的像素值由填充参数决定。

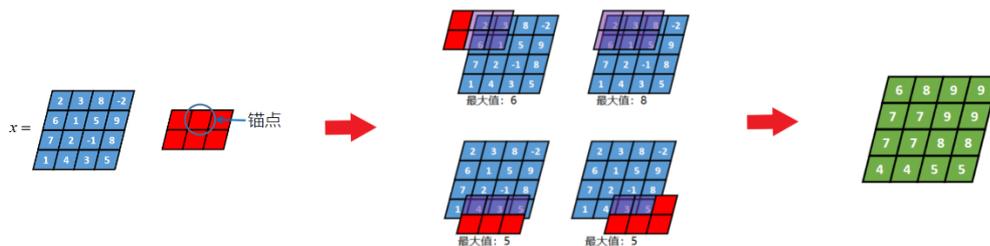


图 2-5 最大池化示意图

(4) 输出层

卷积神经网络中输出层的上游通常是全连接层，其结构和工作原理与传统前馈神经网络中的输出层相同，在此不再赘述。

卷积神经网络和其他神经网络一样，使用反向传播算法进行训练，卷积层的梯度误差为：

$$\delta^{l-1} = \left(\frac{\partial z^l}{\partial z^{l-1}} \right)^T \delta^l = \delta^l \text{rot180}(W^l) \odot \sigma'(z^{l-1}) \quad (2-13)$$

其中， l 表示在模型中的层数， $z^l = a^{l-1}W^l + b^l$ 是卷积运算的输出， a^l 是第 l 层激活函数的输出， W^l 为第 l 层的卷积核， b^l 为 l 层的偏置项， σ 为激活函数。 \odot 为 Hadamard 积，表示两个同阶矩阵对应元素相乘。 rot180 为旋转算子，表示将矩阵旋转 180° ，即将矩阵先上下翻转，再左右翻转。

权重的梯度为：

$$\frac{\partial J(W, b)}{\partial W^l} = a^{l-1} \delta^l = \sum_i \sum_j (\delta_{ij}^l a_{i+p-1, j+q-1}^{l-1}) \quad (2-17)$$

偏置项的梯度为：

$$\frac{\partial J(W, b)}{\partial b^l} = \sum_{u,v} (\delta^l)_{u,v} \quad (2-18)$$

2.3 半监督学习

半监督学习(Semi-Supervised Learning)是监督学习与无监督学习相结合的一种学习方法，它同时使用未标记数据和标记数据，让学习器利用未标记样本来提升学习性能。使用半监督学习时，在带来比较高的准确性的同时还可以减少人工的需求，因此，半监督学习正越来越受到人们的重视。

半监督学习的成立依赖于模型假设，当模型假设正确时，无类标签的样例才能够帮助改进学习性能。半监督学习常见的假设有三个^[34]：

(1) 平滑假设 (Smoothness Assumption)

位于稠密数据区域的两个距离很近的样例的类标签相似，也就是说，当两个样例被稠密数据区域中的边连接时，它们在很大的概率下有相同的类标签；相反地，当两个样例被稀疏数据区域分开时，它们的类标签趋于不同。

(1) 聚类假设 (Cluster Assumption)

当两个样例位于同一聚类簇时，它们在很大的概率下有相同的类标签。这个假设的等价定义为低密度分离假设，即分类决策边界应该穿过稀疏数据区

域，而避免将稠密数据区域的样例分到决策边界两侧。

(3) 流形假设 (Manifold Assumption)

假设数据分布在一个流形结构上，当两个样例位于低维流形中的一个小局部邻域内时，它们具有相似的类标签。

许多实验研究表明当数据不满足这些假设时，无类标签的样例不仅不能对学习性能起到改进作用，反而会恶化学习性能，导致半监督学习算法的性能下降^[35]。

半监督学习按照统计学习理论的角度包括直推学习 (Transductive Learning) 和纯 (pure) 半监督学习两类模式。直推学习假定学习过程中所考虑的未标记样本是待预测数据，学习的目的为在这些未标记样本上获得最优泛化性能；而纯半监督学习假定训练数据中的未标记样本并非待预测数据，希望学得的模型能适用于训练过程中未观察到的数据。前者利用训练集内给定的有标签和无标签样本进行训练，预测训练集中无标签样例的类标签；后者处理整个样本空间中所有给定和未知的样例，利用训练集中的有标签样本和无标签样本，以及未知的测试样例一起进行训练，不仅预测训练数据中无类标签的样例的类标签，更主要的是预测未知的测试样例的类标签。

从不同的学习场景看，半监督学习可分为四大类，即：半监督分类、半监督回归、半监督聚类、半监督降维，本文适用的学习场景为半监督分类，即在无标签样例的帮助下训练有标签的样本，获得比只用有标签的样本训练得到的分类器性能更优的分类器，弥补有类标签的样本不足的缺陷，其中类标签取有限离散值。

2.4 多任务学习

在 1997 年之前，很多的机器学习系统仅专注于单一任务，但有很多任务应用的领域是相关的，而每个机器学习模型的训练都费时费力，因此研究者们提出了这个问题：可否将某个领域或任务上学习到的知识或模式应用到不同但相关的领域或问题中^[36]？1997 年，机器学习杂志发表了一个关于迁移学习的特刊，迁移学习领域的主要研究者发表了一系列文章研究如何训练多个学习任务的问题，其中^[37]提出一种学习多个任务的算法并介绍多任务学习的过程。

多任务学习是一种归纳迁移方法，充分利用隐含在多个相关任务训练信号中的特定领域信息。在反向传播过程中，多任务学习允许共享隐层中专用用于某个任务的特征被其他任务使用，同时来自相关任务的训练信息，也可以

被隐层使用，提取更多优化度量指标的其它信息——共享相关任务之间的表征，可以使我们的模型更好地概括原始任务。实验结果表明，多任务学习可以很好地提高模型的泛化性能，在现实中有很多可以应用多任务学习的场景。

关于多任务学习为什么可以提高泛化能力并没有明确的回答，但存在一些潜在的解释，比如：某种程度上，不相关任务对于其他任务来说可以视为噪声，那么它就可以通过作为噪声源来提高泛化能力^[38]；再如，只有一个任务时，模型可能难以区分相关与不相关的特征，但其它任务可以为这些特征的相关性或不相关性提供额外的证据，因此多任务学习可以帮助模型将注意力集中在重要的特征上^[39]。

3 基于半监督学习的面部动作单元识别框架

本章将介绍基于半监督学习算法的面部动作单元识别框架的研究过程及设计思路，并体现在探索过程中遇到的难题及解决问题的过程。

3.1 结合半监督学习与多任务学习的半隐任务学习

内隐学习是认知心理学的一个重要概念，指的是在无意识中获得某种知识，学习了某种规则，这种学习的效果可以通过某种测试表现出来，但是意识层面却无法觉知这种规则，不能外显的把这种规则说出来^[40]。从前文所述中我们知道，AU 是基于解剖学对人脸肌肉运动进行的编码，是人脸面部所能表现出的面部动作的原子单元，所有的表情都可以被 AU 的组合表现出来。人类并不能直接识别 AU，但可以识别表情。同时，人类还会对他人面部肌肉的微小动作做出反应，但人类却无法显意识地表达出 AU 在人脸出现的规则，可以这么说，人类通过内隐学习学到了 AU 的识别。

于是，在机器学习领域，我们也可以使用这种思想，提出隐任务学习（Hidden-Task Learning）的概念^[41]。其基本思想是，我们拥有两个互相串联的任务，这两个任务是对同一个事实的不同描述，但前一层任务的抽象程度较高，我们称其为隐任务，后一层任务的具体程度更高，我们称其为显任务。在显任务学习的过程中，隐任务的知识其实也被学到了。那么，我们就可以通过显任务的学习来学习隐任务。在这里，显任务就是表情识别任务，而隐任务就是 AU 识别任务。

假定我们拥有训练集 $X^v = \{(x_1^v, y_1^v), \dots, (x_n^v, y_n^v), \dots, (x_N^v, y_N^v)\}$ ，其中 $x_n^v \in \mathbb{R}^d$ 是样本的特征，而 $y_n^v = [y_{n1}^v, \dots, y_{nk}^v, \dots, y_{nK}^v] \in \{0,1\}^K$ 是样本在显任务中的标签。在这里 x 是输入的人脸特征，而 y^v 则是图像的表情标签。

对于隐任务 t ，记从 x 到隐任务输出的映射函数为 $h(x, \theta_t)$ ， θ_t 是函数的参数。假设参数集合 $\Theta = \{\theta_1, \dots, \theta_t, \dots, \theta_T\}$ ，就可以定义隐任务的映射函数为：

$$H(x, \Theta) = [h(x, \theta_1), \dots, h(x, \theta_t), \dots, h(x, \theta_T)]^T \quad (3-1)$$

T 是隐任务的个数，该函数将 x 映射为 T 个隐任务的输出。在这里，隐任务的输出即为 AU 的概率。

与之相似的，假设显任务的个数为 K ，第 k 个显任务的映射函数为 $v(H(x, \Theta), \phi_k)$ ，它将 $H(x, \Theta)$ 映射成显任务的输出，其中 ϕ_k 是第 k 个显任务的参数。记显任务的参数集合为 $\Phi = \{\phi_1, \dots, \phi_k, \dots, \phi_K\}$ 。在这里，显任务的输出

即为表情的概率。

当我们在学习显任务的时候，目标函数为：

$$\min_{\Theta, \Phi} \mathcal{L}^v(\Theta, \Phi, X^v) + \beta \mathcal{R}(\Theta) \quad (3-2)$$

其中， $\mathcal{R}(\Theta)$ 是正则化项， $\mathcal{L}^v(\Theta, X^v)$ 是显任务在 X^v 上的经验风险：

$$\mathcal{L}^v(\Theta, \Phi, X^v) = \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K \ell(v(H(x, \Theta), \phi_k), y_{nk}^v) \quad (3-3)$$

若我们使用梯度下降法对其进行优化，则我们可以求出梯度为：

$$\frac{\partial \mathcal{L}^v}{\partial \Phi} = \frac{\partial \mathcal{L}^v}{\partial v} \frac{\partial v}{\partial \Phi} \quad (3-4)$$

$$\frac{\partial \mathcal{L}^v}{\partial \Theta} = \frac{\partial \mathcal{L}^v}{\partial v} \frac{\partial v}{\partial H} \frac{\partial H}{\partial \Theta} \quad (3-5)$$

到这一步为止，对于表情来说，我们构建了一个全监督学习算法，而对于 AU 识别来说，我们构建了一个无监督学习算法。因为当我们只利用 X^v 时，我们却在最小化的过程中可以优化 Θ ，但是这个 Θ 并不具有现实的意义，它映射出的输出只是模型中显任务层之前的一层的输出而已，我们无法给出具体的解释。这也与我们在现实中遇到的内隐学习的情景是一样的，即在我们成长过程中，父母老师在教会我们辨认人脸的活动时，提供的“数据标签”都是表情标签，并没有 AU 的标签，即我们只拥有显任务训练集。但是随着我们看到的人脸表情越来越多，我们也开始会辨认人脸的细微动作，即学到了 AU 的知识，但我们并没有办法显式地将其表达出来。

而对于计算机进行的传统 AU 识别任务，我们要反过来考虑。因为，之前的 AU 识别算法都是利用全监督学习算法，它们使用的数据是带 AU 标注的，即只有隐任务训练集。而且它们的输出是 AU，即隐任务的输出，所以它们只在学习隐任务。

因此，我们给出隐任务训练集 $X^h = \{(x_1^h, y_1^h), \dots, (x_m^h, y_m^h), \dots, (x_M^h, y_M^h)\}$ ，其中 $y_n^h \in \{0, 1\}^T$ 是样本在隐任务中的标签。在这里 x 是输入的人脸特征，而 y^h 则是图像的 AU 标签。当我们学习隐任务的时候，目标函数为：

$$\min_{\Theta} \mathcal{L}^h(\Theta, X^h) + \beta \mathcal{R}(\Theta) \quad (3-6)$$

其中， $\mathcal{R}(\Theta)$ 是正则化项， $\mathcal{L}^h(\Theta, X^h)$ 是隐任务在 X^h 上的经验风险：

$$\mathcal{L}^h(\Theta, X^h) = \frac{1}{MT} \sum_{m=1}^M \sum_{t=1}^T \ell(h(x_m^h, \theta_t), y_{mt}^h) \quad (3-7)$$

我们可求出其梯度为：

$$\frac{\partial \mathcal{L}^h}{\partial \Theta} = \frac{\partial \mathcal{L}^h}{\partial H} \frac{\partial H}{\partial \Theta} \quad (3-8)$$

到此为止，我们针对隐任务构建了一个全监督学习算法，此时我们学到的 Θ 是有现实意义的，它映射出的输出就是 AU 被激活的概率。而对于显任务来说，我们无法学到关于它的任何知识，因为函数 v 并没有在目标函数中出现，我们自然也无法学到它的参数 Φ 。从另一个角度来说，这是因为隐任务相比显任务更抽象，而显任务更具体，我们可以从抽象的事物中演绎推理出具体的事物，却无法从抽象的事物中归纳总结出具体事物，而监督学习做的事情正是归纳。这就犹如现实中的脸盲症，患者可以记住人面部的细节，如高鼻子、黑眼睛，但却无法将这些细节组合起来辨认出他人。

以上我们运用了多任务学习的思想，接下来，我们将引入半监督学习的思想。在本文设定的情境下，我们同时拥有带 AU 标签和带表情标签的数据，即我们同时拥有训练集 X^v 和 X^h ，因此我们可以同时学习两个任务，这个时候，目标函数为：

$$\min_{\Theta, \Phi} (1 - \alpha) \mathcal{L}^h(\Theta, X^h) + \alpha \mathcal{L}^v(\Theta, \Phi, X^v) + \beta \mathcal{R}(\Theta) \quad (3-9)$$

其中， α 是两个经验风险的权衡项，它衡量了两个任务在总任务中的重要程度。对于上式，我们称其为半隐任务学习（Semi-Hidden-Task Learning）。当我们在最小化目标函数时，我们同时在寻找最优的参数 Θ 和 Φ ，并且此时优化得到的两个参数都有现实意义，它们映射出的输出分别为 AU 被激活和表情出现的概率。

对于该式的梯度求解只要使用导数的加法规则即可，记总经验风险为 L ，则：

$$\frac{\partial L}{\partial \Phi} = \alpha \frac{\partial \mathcal{L}^v}{\partial v} \frac{\partial v}{\partial \Phi} \quad (3-10)$$

$$\frac{\partial L}{\partial \Theta} = \alpha \frac{\partial \mathcal{L}^v}{\partial v} \frac{\partial v}{\partial H} \frac{\partial H}{\partial \Theta} + (1 - \alpha) \frac{\partial \mathcal{L}^h}{\partial H} \frac{\partial H}{\partial \Theta} \quad (3-11)$$

此时，对于两个任务来说，该算法都是一个半监督学习算法，因为对于每个任务它都提供了带数据标签和不带数据标签的数据。从理论上来说，该算法可以提高模型的性能：对于 AU 识别任务，表情标签提供了 AU 的聚类依据；而对于表情识别任务，AU 标签通过注意力机制让其排除不重要的特征。

进一步来说，由于本文的任务仅是 AU 识别，我们希望在有限的数据集上可以让误差信息更多地优化 Θ 而不是 Φ 。于是我们假设 Φ 可以在训练阶段前由有关隐任务输出和显任务输出之间关系的先验知识获得，那么仅针对隐任

务学习的目标就是通过最小化以下目标函数来学得最优的隐任务参数 Θ :

$$\min_{\Theta} (1 - \alpha)\mathcal{L}^h(\Theta, X^h) + \alpha\mathcal{L}^v(\Theta, X^v) + \beta\mathcal{R}(\Theta) \quad (3-12)$$

注意，该式与式 3-9 不同的地方在于，该式中我们并没有优化 Θ 。

3.2 显任务参数的预先确定

在上一小节的最后我们提出，为了充分利用有限的数据集对隐任务参数 Θ 的优化，需要预先确定显任务的参数 Φ 。在本节，本文将提出两种办法预先确定 Φ 。

3.2.1 基于伪样本采样的显任务参数预训练

如果我们拥有一系列以 AU 为特征，以表情为标注的样本，那么我们就可以单独训练显任务的参数，训练方法为直接将 AU 特征当成隐任务的输出。为此我们考虑利用前人总结的领域知识，以领域知识为规则，生成一系列符合其约束的样本，将这些生成的样本，称为伪样本。在这里，我们利用的领域知识为 AU 和表情的概率分布。

由于表情可以表示为 AU 的组合，因此我们最先想到的便是 AU 在表情条件下的边缘概率，即 $P(AU_i = 1|\text{表情})$ 。Du 等^[42]收集了大量含有基本表情的人脸图片并做了统计分析，总结出了各表情条件下每个 AU 发生的边缘概率，在这里只列出与表情有关的 AU（表 3-1），表中的 0.2 表示该 AU 在该表情下发生的概率小于 0.2，而表中的 0.7 表示该 AU 在该表情下发生的概率大于 0.7，而其他数值则表示具体的概率。

表 3-1 各 AU 在各表情条件下的边缘概率

	生气	恶心	恐惧	高兴	伤心	惊讶
AU1	0.20	0.20	0.70	0.20	0.60	0.70
AU2	0.20	0.20	0.57	0.20	0.20	0.70
AU4	0.70	0.31	0.70	0.20	0.70	0.20
AU5	0.20	0.20	0.63	0.20	0.20	0.66
AU6	0.20	0.20	0.20	0.51	0.50	0.20
AU7	0.70	0.20	0.20	0.20	0.20	0.20
AU9	0.20	0.70	0.20	0.20	0.20	0.20
AU10	0.26	0.70	0.20	0.20	0.20	0.20
AU11	0.20	0.20	0.20	0.20	0.26	0.20
AU12	0.20	0.20	0.20	0.70	0.20	0.20
AU15	0.20	0.20	0.20	0.20	0.70	0.20
AU17	0.52	0.70	0.20	0.20	0.67	0.20
AU20	0.20	0.20	0.70	0.20	0.20	0.20
AU23	0.29	0.20	0.20	0.20	0.20	0.20
AU24	0.70	0.26	0.20	0.20	0.20	0.20
AU25	0.20	0.20	0.70	0.70	0.20	0.70
AU26	0.20	0.20	0.33	0.20	0.20	0.70

在某种表情下，有些 AU 并不单独出现，而总是在 AU 的组合中出现，即 $P(AU_i = 1|表情) < P(AU_i = 1|表情, AU_k = 1)$ 。彭国柱^[43]基于 EMFACS 获取了各表情下常出现的 AU 组合的领域知识，这体现了表情条件下 AU 的共生关系。

表 3-2 各表情下常出现的 AU 组合

表情	常出现的 AU 组合
生气	4+5, 4+7, 4+5+7, 17+24, 23
恶心	1+2+4, 20
恐惧	9, 10
高兴	12, 6+12, 7+12
伤心	1, 1+4, 15, 6+15, 11+17, 11+15
惊讶	1+2+5, 1+2+26, 1+2+5+26

除此之外，由于各 AU 之间存在解剖学联系，肌肉的行为模式导致各 AU 之间存在着与表情无关的共生和互斥关系。如 AU23(嘴唇收紧)与 AU24(嘴唇按压)都与口轮匝肌相关，人们很难只做出其中一个动作，所以

$P(AU_{23}=1|AU_{24}=1) > P(AU_{23}=1)$, $P(AU_{24}=1|AU_{23}=1) > P(AU_{24}=1)$, 再如 AU_{12} (嘴角外拉) 由颧大肌产生, AU_{15} (嘴角下压)由降口角肌产生, 二者的肌肉相冲突, 因此二者几乎不同时出现, 即 $P(AU_{12}=1|AU_{15}=1) < P(AU_{12}=1)$, $P(AU_{15}=1|AU_{12}=1) < P(AU_{15}=1)$ 。将 AU 间的共生互斥关系总结在下表^[44]。

表 3-3 AU 之间的共生和互斥关系

共生 AU	互斥 AU
AU1-AU2-AU5,	AU12-AU15, AU12-AU17,
AU4-AU7-AU9,	AU2-AU6, AU2-AU7, AU2-AU9,
AU15-AU17-AU24,	AU15-AU25, AU17-AU25
AU23-AU24,	AU23-AU25, AU24-AU25

根据这些领域知识, 我们可以开始生成伪样本。由于这些先验概率对 AU 的约束由强到弱为表情下的 AU 组合大于 AU 间的共生互斥关系大于 AU 在表情条件下的边缘概率^[44], 因此进行一次伪样本采样的流程可以描述为: 给定表情, 并对于本次采样指定一个该表情下常出现的 AU 组合, 首先对 AU 进行随机排列以确定本次的采样顺序, 随后开始采样; 对于每个 AU, 若其在指定的 AU 组合里, 则对其以大于 0.7 的概率进行采样; 否则在已经被采样的 AU 里进行查找, 寻找与其相关的第一个 AU, 若找到的是共生 AU, 则对其以大于 0.7 的概率进行采样, 若找到的是互斥 AU, 则对其以小于 0.2 的概率进行采样; 若没有找到与其相关的 AU, 则对其以该表情条件下的边缘概率进行采样。

算法 3-1 伪样本采样算法

输入: 表情 e , 先验知识
输出: 一个表情 e 下的 AU 激活情况伪样本

- 1: 初始化: 对 AU 随机排序, 指定一个该表情下的 AU 组合
- 2: **for** $t = 1, 2, \dots, T$ **do**
- 3: 开始采样 AU_t
- 4: **if** AU_t 在表情组合中 **then**
- 5: 以大于 0.7 的概率采样 AU_t
- 6: **else if** AU_1 至 AU_{t-1} 中存在 AU_t 的相关 AU **then**
- 7: **if** 第一个相关 AU 是共生 AU **then**
- 8: 以大于 0.7 的概率采样 AU_t
- 9: **else**
- 10: 以小于 0.2 的概率采样 AU_t
- 11: **end if**
- 12: **else**
- 13: 对 AU_t 以表情 e 条件下的边缘概率进行采样
- 14: **end if**
- 15: **end for**

在通过对每个表情运行多次该伪样本采样算法, 我们就获得了一系列以 AU 激活情况为特征, 表情出现情况为标签的伪样本, 我们将其作为训练集, 就可以单独预训练显任务, 以获得参数 Φ 。

3.2.2 使用真实标签预训练显任务参数

在对研究现状进行文献综述的时候我们提到, 有许多研究者利用概率图模型捕获 AU 和表情之间的相互依赖关系, 那么我们也可以选择不使用由心理学家统计出的领域知识, 而是在我们已经拥有的数据集上寻找同时拥有 AU 标注和表情标注的数据, 并用它们的真实标签来预训练显任务。这暗含了 X^v 和 X^h 具有重叠部分的假设, 而就本文所使用的面部数据库来说, 该情况是真实成立的。

具体而言, 记 $x^v = \{x_i | (x_i, y_i^v) \in X^v\}$, $x^h = \{x_i | (x_i, y_i^h) \in X^h\}$, 则如果我们有 $x^v \cap x^h \neq \emptyset$, 那么我们就可以取同时原始特征 x 同时在显任务训练集和隐任务训练集中的样本集 $X^{vh} = \{(x_i, y_i^v, y_i^h) | x_i \in x^v \cap x^h\}$, 我们不使用它的原始特征, 而只使用它们的标注数据 y_i^v 和 y_i^h , 即我们有单独用来训练显任务层的训练集 $Y = \{(y_i^v, y_i^h) | (x_i, y_i^v, y_i^h) \in X^{vh}\}$ 。我们将 Y 中的 y_i^v 视作特征, 将 y_i^h 视作标签, 利用下式对显任务层进行训练:

$$\min_{\Phi} \mathcal{L}^v(\Phi, Y) + \beta \mathcal{R}(\theta) \quad (3-13)$$

注意此时 $H(\mathbf{x}, \boldsymbol{\theta})$ 已经被 Y 替代, 所以我们可以单独确定 Φ :

$$\mathcal{L}^v(\Phi, Y) = \frac{1}{NK} \sum_{n=1}^N \sum_{k=1}^K \ell(v(\mathbf{y}^n, \phi_k), y_{nk}^v) \quad (3-14)$$

4 面部动作单元识别系统的实现

本章首先对本文所要建立的系统的整体方案进行介绍，随后对每一部分的实现进行介绍。

4.1 AU 识别系统的整体方案

上一章对本文所试图建立的 AU 识别系统的预测模型部分的数学框架进行了介绍，但整个系统是由多个部分构成的，接下来对该系统的整体方案进行介绍（图 4-1）。

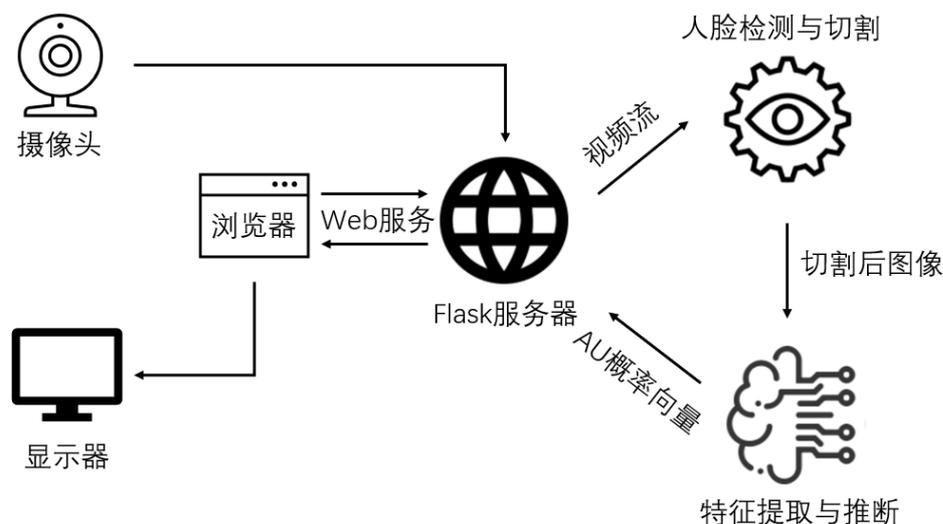


图 4-1 AU 识别系统技术流程图

(1) 人脸图像的获取

用户将外部设备如摄像头捕获的实时图像上传至系统。

(2) 人脸检测与切割

为了提高对输入图像进行识别的准确率，需要剔除不必要的噪声。对于 AU 识别任务来说，整个系统最需要的输入为人面部区域的图像，而其背景则是一种噪声，因此，本系统需要先对输入图像进行人脸检测，并对人脸区域进行切割，还要调整其尺寸至统一大小，以适应识别系统的输入。

(3) 面部特征的提取

直接将像素强度作为特征输入计算量巨大，且噪声较多，因此需要先对

面部图像进行特征提取，将提取到的面部特征作为 AU 识别系统的输入。

(4) 输入特征的预测

提取好的特征将送入训练好的模型进行推断，模型将输出一个一维向量，向量的每一个元素为对应 AU 的激活概率，当其大于设定的阈值时，视为该 AU 激活。

(5) 结果可视化

系统对接收到的图像进行面部区域的编号和标注，随后将标注后的图像和对应人脸的 AU 激活情况输出到外部设备如显示器供用户查看。

4.2 图像的获取

本设计采用 B/S 架构，原先拟定的实现思路为浏览器获取摄像头画面，并将视频流使用 POST 方法上传至服务器进行检测，服务器返回检测结果的 JSON 数组并在页面绘图，可视化检测结果。但由于浏览器获取摄像头画面的传统方法 FLASH 已不再被支持，而 W3C 推荐标准中通用的网页即时通信 (WebRTC) 方法获取摄像头画面时仅支持 HTTPS 协议，需要申请域名和 SSL 证书，因此本文只能将服务搭建在本地。

本文使用 OpenCV 的 VideoCapture() 方法获取摄像头画面，并送入检测系统，根据检测结果在图像上绘图并推流至 Web 页面。后端使用 Python 开发，使用的服务器框架为 Flask。

4.3 人脸检测与切割

OpenCV^[46]是一个基于 BSD 许可发行的跨平台计算机视觉库，可以运行在 Linux、Windows、Android 和 Mac OS 操作系统上。它轻量级而且高效——由一系列 C 函数和少量 C++ 类构成，同时提供了 Python、Ruby、MATLAB 等语言的接口，实现了图像处理和计算机视觉方面的很多通用算法。

OpenCV 自带了一个由 300 万张人脸图片训练好的基于 Haar 特征的人脸级联分类器，并可通过 Python 接口调用。其实时性很高，本文采用该方法进行人脸的检测和切割，使用的库版本为 OpenCV4.2.0。由于进行 AU 识别时一般是正脸，因此在检测时不考虑侧脸的情况。切割后的图像统一调整其尺寸至 96×96，随后送入人脸特征提器。

4.4 人脸特征提取

本文共尝试了三种人脸特征的提取方法，分别为人脸特征点、HOG 特征和卷积神经网络，这三种方法对特征的提取能力由弱到强，其参数数量也从少到多。在最终的实用系统中，本文选择了效果最好的卷积神经网络作为特征提取方式。

4.4.1 人脸特征点提取

Dlib^[47]是一个现代化的 C++工具箱，其中包含用于在 C++中创建复杂软件以解决实际问题的机器学习算法和工具。它广泛应用于工业界和学术界，其开源许可证允许用户在任何应用程序中免费使用它。

Dlib 包含了一个预训练好的实时人脸特征点提取工具，它使用的算法就是本文第二章介绍的基于级联回归树的人脸特征点提取算法，其返回图中每个人脸的 68 个特征点的坐标（图 4-2），本文使用 Dlib 19.19.0 进行人脸特征点的提取。

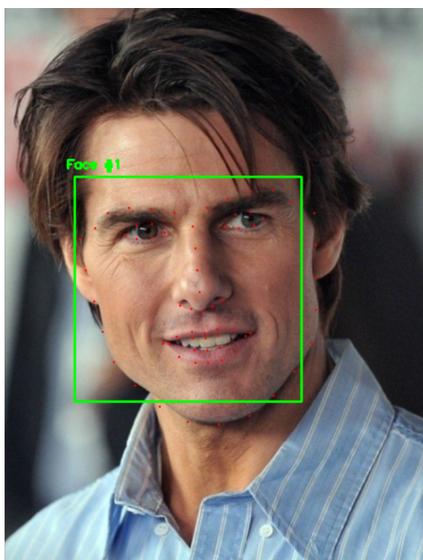


图 4-2 Dlib 提取的人脸特征点

在完成特征点的提取后对提取的特征点坐标进行归一化处理，归一化后的坐标可以忽视不同人脸的尺寸差异，同时还可以在训练过程中提升模型的收敛速度。

$$x_i \leftarrow \frac{x_i - \bar{x}}{\sqrt{\sum_{j=1}^{68} (x_j - \bar{x})^2 / 68}}, i = 1, 2, \dots, 68 \quad (4-1)$$

$$y_i \leftarrow \frac{y_i - \bar{y}}{\sqrt{\sum_{j=1}^{68} (y_j - \bar{y})^2 / 68}}, i = 1, 2, \dots, 68 \quad (4-2)$$

4.4.2 HOG 特征提取

OpenCV 提供了一系列对像素进行操作的函数，在 Python 中，OpenCV 的库名称为 cv2。

对于输入的图像，针对每个像素，本文首先使用基于 Sobel 算子的 cv2.Sobel() 函数计算其在 x 方向和 y 方向上的梯度分量，随后使用 cv2.phase() 计算两分量的方向场（即梯度的方向），而计算梯度的模时，本文使用了式 2-8 中的近似公式。图 4-3 展示了一副计算完梯度后的人脸图像在每一个像素处的梯度。

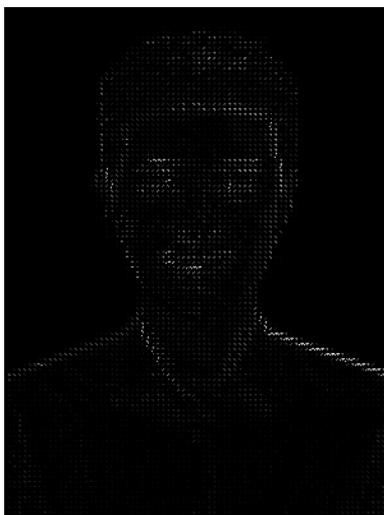


图 4-3 图像梯度可视化

本文定义每个 cell 大小为 8×8 ，整张图像为一个 block，由于输入图像尺寸为 96×96 ，因此该操作就是将整张图像划分为 144 个 cell。对于角度，本文则将其划分为 9 个长度为 $\pi/9$ 的区间。随后本文在每个 cell 上统计其梯度直方图，将其串联起来后便得到了该图像的 HOG 描述符，即一个 1296 维的向量。

4.4.3 卷积神经网络人脸特征提取

本文实现卷积神经网络所用的库为 Tensorflow^[48]，它是一个基于数据流编程的符号数学系统，被广泛应用于各类机器学习算法的编程实现，并且它支持包含 GPU 加速的版本。本文所用的 Tensorflow 版本为 2.0.0b。

VGGNet^[49]是由牛津大学 Visual Geometry 小组和 Google DeepMind 公司研究员一起研发的深度卷积神经网络。它探索了卷积神经网络的深度和其性能之间的关系，通过反复地堆叠 3×3 的小型卷积核和 2×2 的最大池化层，成功的构建了一个深度卷积神经网络。

其最大的创新点在于网络层数的增长并不会带来参数量上的爆炸，因为参数量主要集中在最后三个全连接层中。同时，两个 3×3 卷积层的串联相当于 1 个 5×5 的卷积层，3 个 3×3 的卷积层串联相当于 1 个 7×7 的卷积层，即 3 个 3×3 卷积层的感受野大小相当于 1 个 7×7 的卷积层（图 4-4）。但是 3 个 3×3 的卷积层参数量只有 7×7 的一半左右，同时前者可以有 3 个非线性操作，而后者只有 1 个非线性操作，这样使得前者对于特征的学习能力更强。

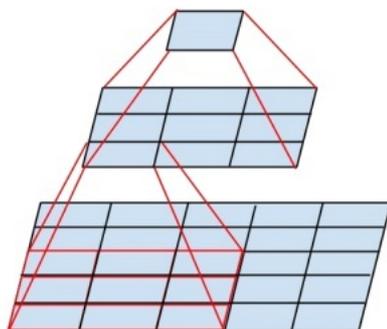


图 4-4 小卷积核 对 大卷积核的替代

VGGNet 包含很多级别的网络，深度从 11 层到 19 层不等，比较常用的是 VGGNet-16 和 VGGNet-19。VGGNet 把网络分成了 5 段，每段都把多个 3×3 的卷积网络串联在一起，每段卷积后面接一个最大池化层，最后面是 3 个全连接层和一个 softmax 层（图 4-5）。

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

图 4-5 不同种类 VGGNet 结构图^[49]

本文不需要 VGGNet 这样参数达到上亿的深度神经网络，因此本文仿照其思路构建了一个由五个 3×3 的卷积网络和两个最大池化层组成的神经网络，本文称其为 SmallVGGNet。

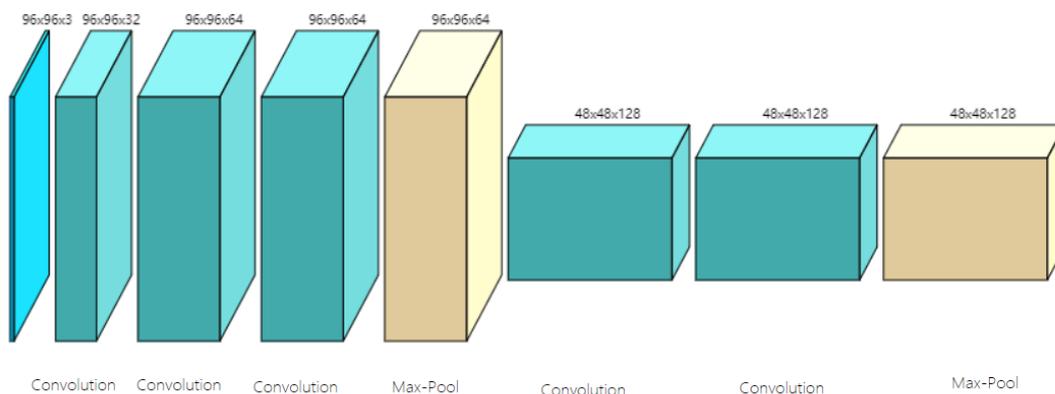


图 4-6 SmallVGGNet 结构图

如图 4-6 所示,输入的人脸图像为 96×96 的三通道图像,图像最先进入一个由 32 个 3×3 卷积核构成的卷积层,随后串联了两个带有 64 个 3×3 卷积核的卷积层,图像在经过一个池化窗口大小为 2×2 的最大池化层后尺寸变为 48×48 ,接着图像连续经过两个深度为 128 的卷积层和一个深度为 128 的最大池化层,卷积核的尺寸和池化窗口的大小与前四层相同,每个卷积层所使用的激活函数为 `relu` 函数。最后,特征图被拉成一个 1024 维向量,这就是该网络从输入图像中提取到的特征。

每个卷积层之后都紧跟一个批归一化 (Batch Normalization) 层,在训练时该层可以利用本批数据的平均值和标准差对数据进行归一化,减少出现梯度消失的情况,从而加快收敛。

每个最大池化层之后都紧跟一个 Dropout 层,在训练时,该层通过在每个训练批次中忽略一半的隐层节点(使其值为 0)减少隐层节点间的相互作用,可以明显地减少过拟合现象。所谓隐层节点间的相互作用是指某些节点只有依赖其他节点才能发挥作用,即对局部特征的依赖,通过 Dropout 操作,可以减缓这种依赖,提高模型的泛化能力。

4.5 输入特征的预测

对输入特征进行预测,即通过 $H(\mathbf{x}, \Theta)$ 前一小节提取到的特征映射为 AU 和表情的概率,本文采用感知机模型作为此处的映射函数。本文所考察的 AU 为与表情相关的 17 个 AU (即 AU1、AU2、AU4、AU5、AU6、AU7、AU9、AU10、AU11、AU12、AU15、AU17、AU20、AU23、AU24、AU25、AU26),于是本文构建了一个由 17 个神经元构成的全连接层作为 $H(\mathbf{x}, \Theta)$ 。由于 AU 可以同时出现,因此隐任务是一个多分类多标签分类任务,所以该层的激活函数为 Sigmoid。当输出大于预先设定的阈值(此处为 0.5)时,认为其对应的 AU 被激活。

4.6 用户交互界面

对输入特征进行预测后,后端将对图像进行绘图和标注工作,随后利用 Flask 的 Response 方法返回一个 image 对象,浏览器在接受到该对象后对页面进行渲染(图 4-7),将结果显示给用户。用户界面的其他部分也由浏览器渲染而成。

AU识别系统

A simple AU recognition system



图 4-7 结果返回界面示意图

4.7 训练阶段使用的双输入双输出模型的搭建

在预测阶段，我们使用的是一个单输入单输出模型，但是在训练阶段中，我们需要用到显任务来帮助隐任务的训练，因此我们需要构建一个双输入双输出模型，模型的构建使用 tensorflow.Keras 的 Functional API 完成。

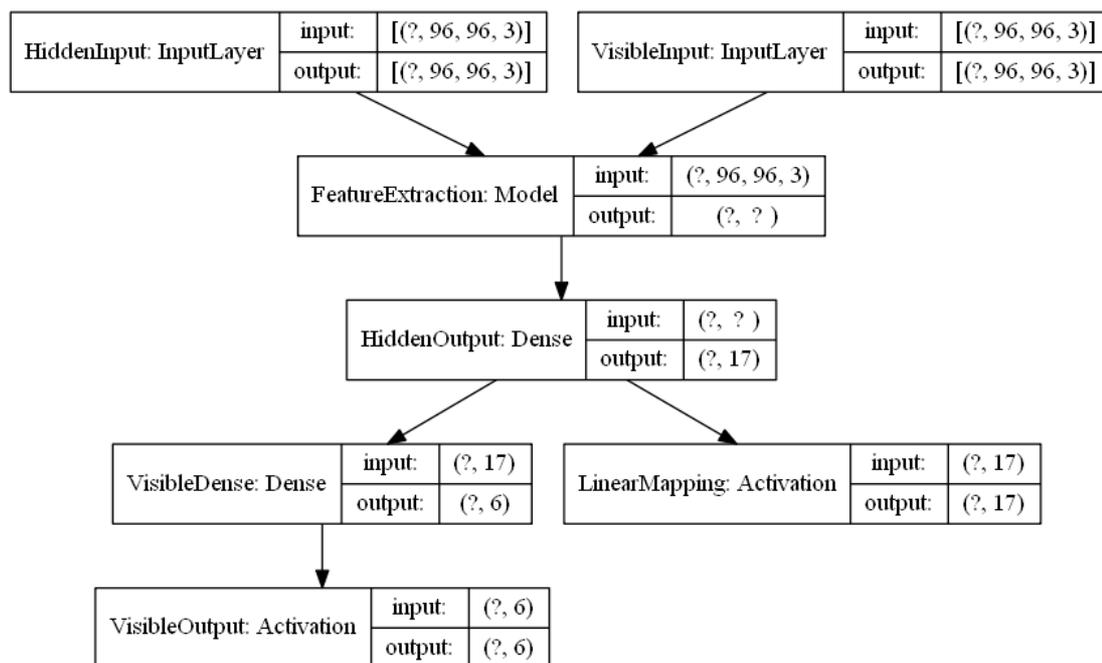


图 4-8 训练阶段使用的双输入双输出模型示意图

图 4-8 展示了本文训练阶段使用的双输入双输出模型，每个矩形框的左侧，冒号前的字符串为该层的名称，冒号后的字符串为该层的类型；每个矩形框的右侧，input 后的元组表示该层的输入尺寸，output 后的元组表示其输出尺寸，“？”表示输入的该维度长度未知。

该模型最顶层的两个层为 InputLayer（输入层），该层接受指定尺寸的输入，在此处为三通道的 96×96 图像。这两个输入层分别命名为 HiddenInput 和 VisibleInput，表示它们分别接受来自隐任务训练集 X^h 和显任务训练集 X^v 的输入。

模型的第三层命名为 FeatureExtraction，它的作用是提取输入的图像的特征。它的类型为 Model，因为在此处我们不知道其内部结构，将其视为一个黑箱，它的输出也是未知的，当我们将其替换为本章第二小节介绍的三种特征时，其输出的特征向量分别为 128 维、1296 维和 1024 维。

模型的第三层为一个 Dense 层，即全连接层，该层的输入是 FeatureExtraction 层的输出，输入的尺寸随着 FeatureExtraction 层使用的特征提取方法的变化而变化，但其输出必定为一个 17 维的向量。该层的激活函数为 Sigmoid 函数，从该层的名称 HiddenOutput 我们可以看出，该层的输出已经是隐任务的输出，这个 17 维向量的每一个元素分别代表了对应 AU 的激活概率。

第四层的右侧是出于理解方便而增加的一层，它不对输入作任何变换，

直接原样输出，通过增加这一层，我们可以清楚地从图中看出模型双输出的结构。

第四层的左侧与第五层结合就是显任务的映射函数，我们也使用一个感知机模型作为 $v(H(x, \theta), \Phi)$ ，因此我们在 VisibleDense 层使用一个 Dense 层。在 TensorFlow 的语法中，某层紧跟一个 Activation（激活）层与在该层直接使用激活函数对其输出进行映射是等价的，为了理解清楚本文将 Activation 层单独列示。由于人脸一次只能出现一种表情，因此 VisibleOutput 层的激活函数为 Softmax 函数。由于本文所有的数据库中只包含六种基本表情，因此显任务的输出为一个 6 维向量。

根据式 3-10 和 3-11，在反向传播的过程中，VisibleOutput 的经验风险将同时对 VisibleDense 和 HiddenOutput 的参数产生影响，而 HiddenOutput 的经验风险将只对 HiddenOutput 的参数产生影响。

5 5.实验与分析

本章将对实验环节进行阐述。首先将对实验准备工作进行说明，包括实验环境、评价指标和实验数据集，其次分别对几种特征提取方式和是否引入显任务进行对比，最后将对本章内容进行总结。

5.1 实验准备

5.1.1 实验环境

本文中所涉及的实验均在笔者的个人计算机上完成，该设备配备一颗 Intel Core i7-9750H 处理器，六核心十二线程，主频 2.6GHz，内存容量为 32GB。同时，该设备搭载有一颗 GeForce RTX 2060 GPU，拥有 6144MB 显存和 1920 个 CUDA 核心，根据 nVidia 公布的数据，每个 CUDA 核心的算力为 7.5。对本实验来说，6144MB 的显存略有不足，在实验过程中，需要开启显存与内存的交换才可以正常进行训练，这也拖累了一定的速度。

实验环境的操作系统版本为 Windows 10 1909，语言环境为 Python 3.7.1，所使用的库的版本在本文第四章均有介绍。

5.1.2 评价指标

查准率 (Precision, P) 与查全率 (Recall, R) 是常见的度量学习器性能的指标。对于二分类任务来说，将真实标签为正例且预测标签也为正例的样本个数记为 TP，将真实标签为反例但预测标签为正例的样本个数记为 FP，将真实标签为正例但预测标签为反例的样本个数记为 FN，将真实标签为反例且预测标签也为反例的样本个数记为 TN。则可以定义三个指标为：

$$P = \frac{TP}{TP + FP} \quad (5-1)$$

$$R = \frac{TP}{TP + FN} \quad (5-2)$$

而一个多分类问题可以当成多个二分类问题的组合，此时我们可以用两种做法来衡量模型在这个多分类任务上的整体性能。第一种方法是针对每个二分类任务分别计算其查准率和查全率，再计算它们的平均值，这样就得到了宏查准率 (macro-P) 和宏查全率 (macro-R)，即：

$$macroP = \bar{P} \quad (5-3)$$

$$macroR = \bar{R} \quad (5-4)$$

第二种方法是计算出每个二分类任务的 TP、NP、TN、FN，再利用它们的平均值计算微查准率 (micro-P) 和微查全率 (micro-R)：

$$microP = \frac{\overline{TP}}{\overline{TP} + \overline{FP}} \quad (5-5)$$

$$microR = \frac{\overline{TP}}{\overline{TP} + \overline{FN}} \quad (5-6)$$

基于此，我们还可以分别定义宏 F1 (macro-F1) 和微 F1 (micro-F1)，它们实际上是对应的 P 和 R 的调和平均数，是对两个指标的综合考虑：

$$macroF1 = \frac{2 \times macroP \times macroR}{macroP + macroR} \quad (5-7)$$

$$microF1 = \frac{2 \times microP \times microR}{microP + microR} \quad (5-8)$$

除此之外，还有一个非常常用的指标精度 (Accuracy)，它是分类正确的样本数占样本总数的比率。

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (5-9)$$

5.1.3 数据准备

CK+数据库^[50]是由卡耐基梅隆大学开发的一个免费人脸表情数据库，它包含有 123 个实验对象做出六种基本表情的 593 个图像序列，并且每个序列的最后一帧都有 AU 的标签。在本任务中，本文利用了该数据库中带 AU 标签的 593 张图像和带表情标签的 327 张图像。



图 5-1 一个生气表情的图像序列

5.2 显任务参数的预训练

基于本文第四章内容，我们构建一个输入为一个 17 维的特征向量，输出为 6 维向量，不包含隐藏层的“双层感知机模型”（图 5-2），其输出层的激活函数为 Softmax 函数。

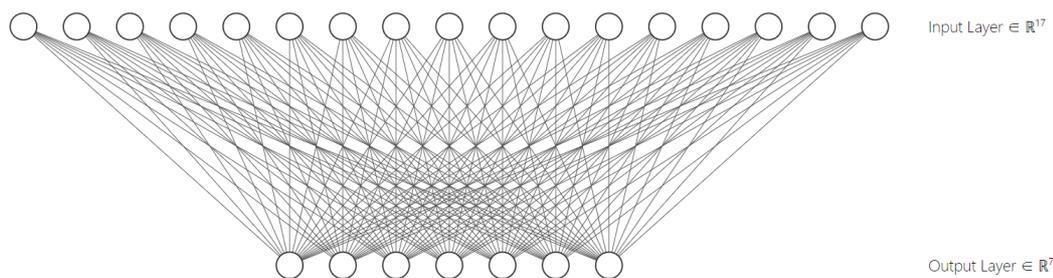


图 5-2 显任务层

首先我们使用本文 3.2.1 节中所介绍的伪样本采样算法，对六种表情各采样出 5000 个伪样本，将其投入显任务层的训练，并保存训练好的权重待用。

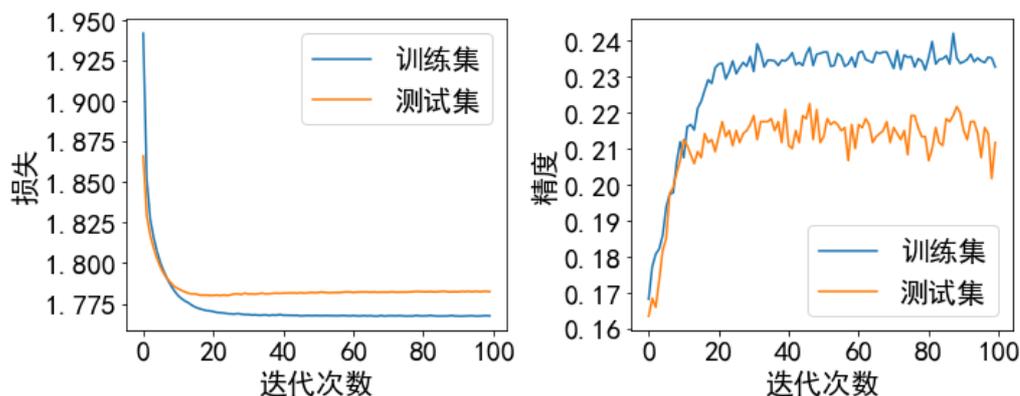


图 5-3 使用伪样本训练显任务层的结果

由于这些伪样本都是生成的，因此它们必然满足同分布假设，所以它们的经验误差和泛化误差十分接近，在训练集和测试集上的精度也都很接近。从图 5-3 中我们可以看出，使用伪样本训练显任务层参数的效果较差，在 100 个迭代次数之后，其经验误差仍高达 1.77，而精度只有 0.22 左右。从现象上看这似乎是欠拟合了，但经实验，在一定程度内增加显任务层的复杂度并不能进一步提升显任务层的性能，这说明伪样本中并没有包含足够的信息，或者其并没有完全反映出 AU 和表情之间的依赖关系，噪声过多。

接下来我们使用 CK+数据库中同时具有 AU 标签和表情标签的真实样本对显任务层进行训练，剔除掉非基本表情的轻蔑后，符合该条件的样本在 CK+数据库中 共有 309 个。本文将 其以 8:2 的比例划分为含 247 个样本的训练集和包含 62 个样本的测试集。

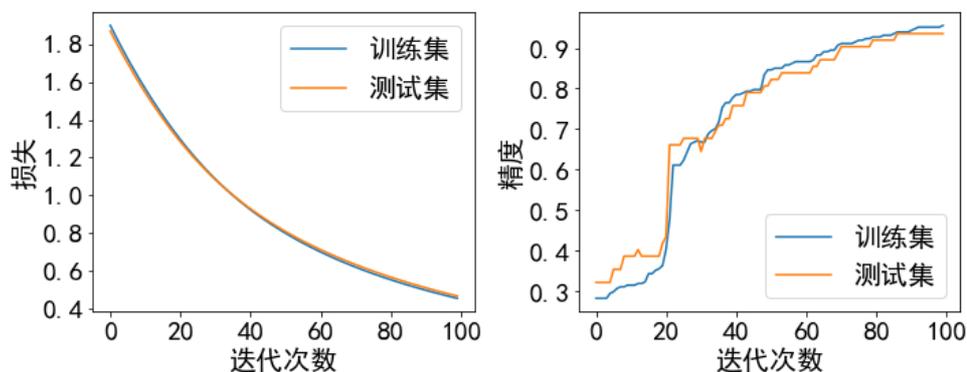


图 5-4 使用真实标签训练显任务层

从图中我们可以看出，在经过 100 个迭代次数的训练后，经验误差和泛化误差的走势和数值都很接近，训练集和测试集上的精度也表现出了一样的

相似性，且最终都超过了 0.9，这说明在本文采用的数据库中，AU 和表情之间存在着确定的依赖关系，并且在数据库中这种依赖关系是同分布的。

5.3 AU 检测对比实验

本文分别基于三种特征提取方法，对不利用半监督学习和利用半监督学习时的学习情况进行对比，并且在利用半监督学习的情况中，分别采用本文第三章所提到的两种显任务参数预训练方案和不锁定显任务参数的训练方案。

5.3.1 使用人脸特征点的 AU 检测

在本节中，FeatureExtraction 层的输出为 68 个人脸特征点坐标组成的 136 维特征向量，它也是 HiddenOutput 层的输入。

首先我们排除显任务层，对 AU 检测任务进行一个监督学习的训练作为对照组。首先我们发现在所有的指标上，模型在训练集和测试集上的走势都是相似的。在经过 100 个迭代次数的训练后，模型的经验误差下降至了 0.3759，而泛化误差下降到了 0.3949，模型的精度在 0.8 左右（图 5-5），但这并不代表模型的准确率很高，从宏查全率和微查全率我们可以看出正例中只有很少的部分被挑了出来，模型的精度之所以很高，是因为 AU 的数量很多，但人脸同时出现的 AU 却并不多，样本集是一个稀疏矩阵，拥有很多的反例，模型通过将大部分样本都预测为反例提高了精度。宏查准率和微查准率的相差较大，这主要是由于两种度量指标的计算方法不同，在计算微查准率时，各 AU 的 TP 和 FP 有多有少，在计算平均值的时候这些数量上的多少被相互抵消了，而对每个分类来说，平均查准率还是比较低的。

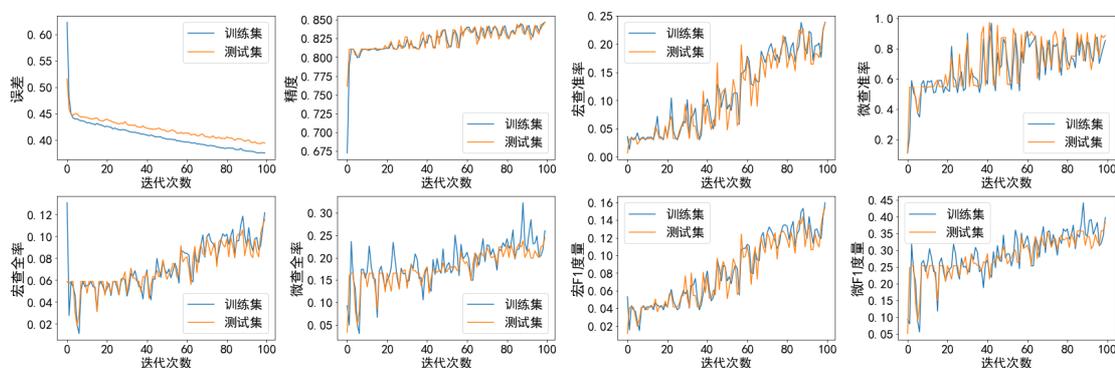


图 5-5 人脸特征点单任务训练结果

接下来我们使用三种方式来加入显任务：第一种方式是不预先确定显层的参数，而是在训练中同时对 Θ 和 Φ 进行优化（图 5-6）；第二种方式是使用伪样本预训练显任务参数，并在训练隐任务参数时将显任务参数锁定（图 5-7）；第三种方式是使用真实标签预先训练显任务参数，在训练隐任务参数时将显任务参数锁定（图 5-8）。引入显任务时，我们设置两个任务的经验风险的权重（ α ）均为 0.5。针对每种方式，我们都训练了 100 个迭代次数观察其结果。注意，以下的都是隐任务（AU 识别任务）的结果图，显任务的结果图并未在此展示。

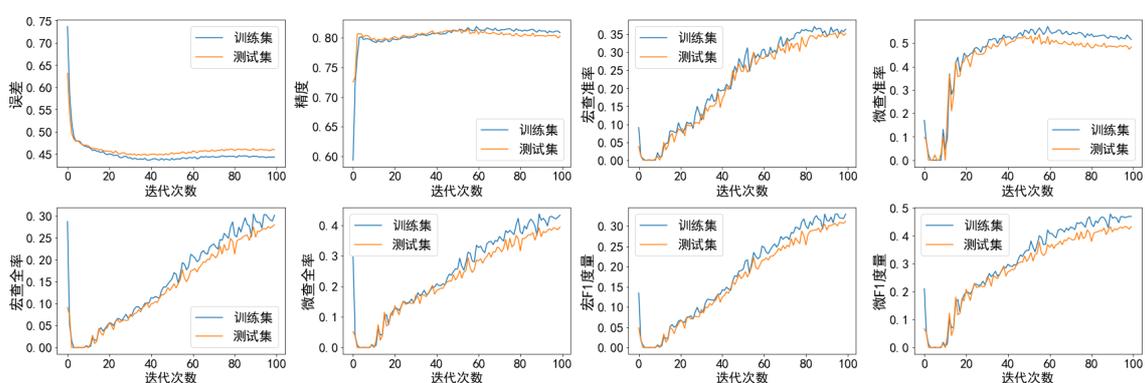


图 5-6 人脸特征点引入显任务但不预训练显任务参数

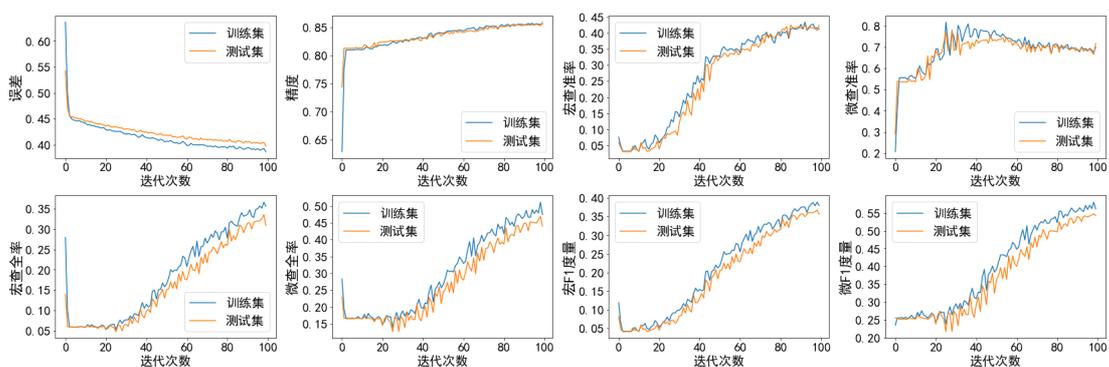


图 5-7 人脸特征点使用伪样本预训练显任务参数

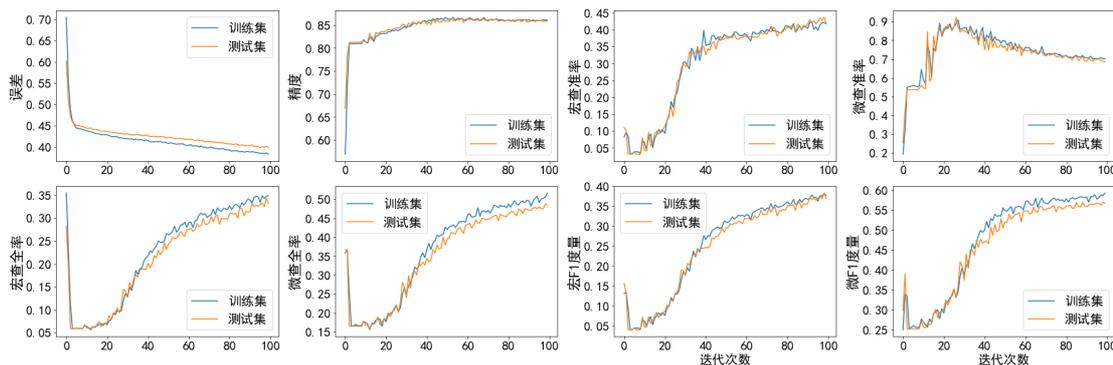


图 5-8 人脸特征点使用真实标签预训练显任务参数

最终，我们将四种情况下模型在测试集上的各项评价指标汇总在一张表中（表 5-1），从表中我们可以发现，结果显示，相比不引入显任务的情况，模型在经验误差和泛化误差并没有下降，精度也没有很大提高的情况下，查准率和查全率都获得了显著的提高，同时，以 F1 度量来衡量，这三种方式的提高幅度依次增大。根据 3-9 式，当不固定 ϕ 进行训练的时候，每次下降的梯度方向在同时优化 ϕ 和 θ ，所以第一种方法所获得的 θ 并不是针对 L^h 最优的参数，而是和 ϕ 作为一个组合，针对总损失 L 的最优参数。而伪样本训练的显任务参数虽然效果不好，却也可以对模型的效果有所提升，这说明伪样本也包含了一定量的信息。

表 5-1 人脸特征点不同情况度量指标对比

度量指标	单任务	不预训练显任务	用伪样本训练显任务	用真实样本训练显任务
泛化误差	0.3949	0.4596	0.3973	0.3988
精度	0.8467	0.8025	0.8588	0.8579
宏查准率	0.2375	0.3526	0.4233	0.4204
宏查全率	0.1154	0.2796	0.3092	0.3314
宏 F1	0.1544	0.3114	0.3564	0.3697
微查准率	0.8872	0.4836	0.7181	0.6859
微查全率	0.2448	0.3939	0.4399	0.4843
微 F1	0.3824	0.4332	0.5432	0.5664

虽然模型通过引入半监督信息获得了效果的提升，但总体来说它的效果还是较差，宏 F1 和微 F1 在分别上升至 0.4 和 0.6 左右后便无法上升，这和选择的特征有很大的关系。特征点包含的信息有限，多数是针对几何特征的，而 AU 检测需要比较精细的面部特征，更适合选用纹理特征。

5.3.2 使用 HOG 特征的 AU 检测

由于特征点作为 AU 识别的特征并不合适，因此本文利用 HOG 特征作为特征，继续对是否引入半监督学习的不同情景进行对比实验。

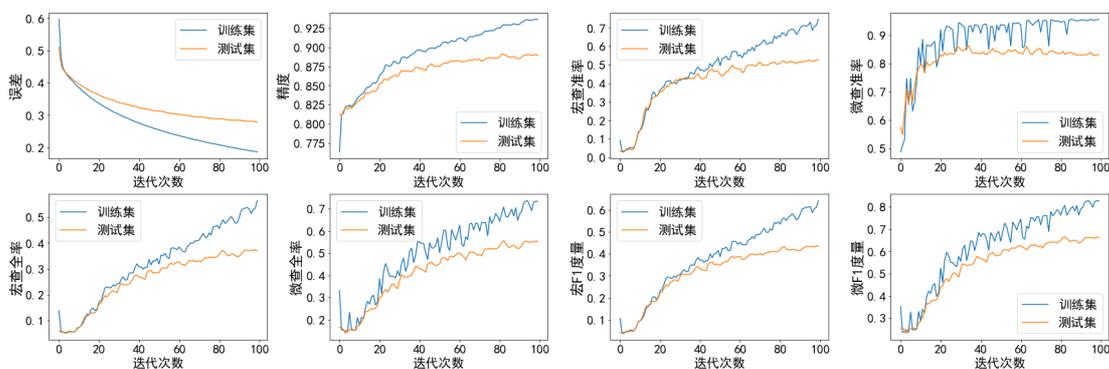


图 5-9 HOG 特征单任务训练结果

我们仍然是首先训练一个单任务的监督学习模型作为对照组，模型的经验误差为 0.1867，而泛化误差则达到了 0.2791，相比上一节，误差总体上有所减小。从图中我们可以发现，与上一节对比最明显的就是随着特征换为纹理特征，模型可以提取的信息增多，模型在训练集和测试集的性能表现差异变得很大，模型的各项度量指标在测试集上均显著落后于训练集，模型的过拟合程度增加了。

使用三种不同的方法引入显任务后，我们发现虽然模型的经验误差和泛化误差都有所加大，但是这二者之间的差距缩小了，由相差 49.5%变为了相差 15.3%到 25.8%，具体的误差情况为不预训练显任务的情况下经验误差为 0.3296，泛化误差为 0.4146，使用伪样本预训练显任务的情况下经验误差为 0.2642，泛化误差为 0.3516，使用真实样本预训练显任务的情况下经验误差为 0.2425，泛化误差为 0.2797。

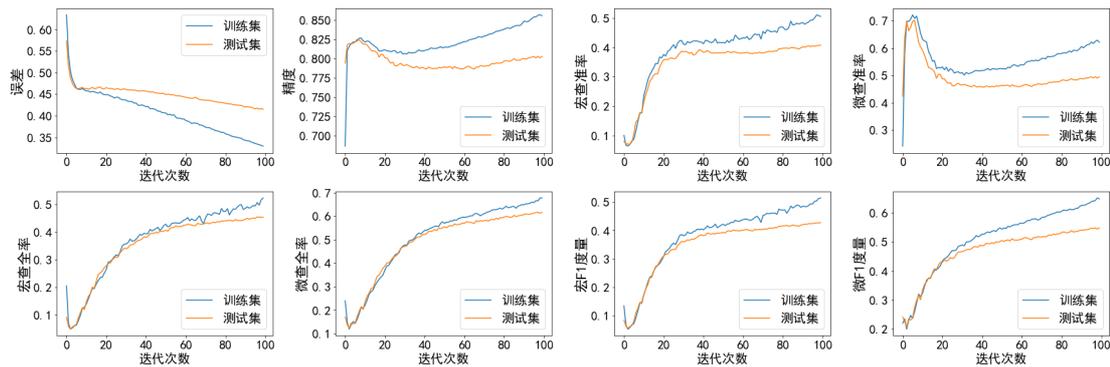


图 5-10 HOG 特征引入显任务但不预训练显任务参数

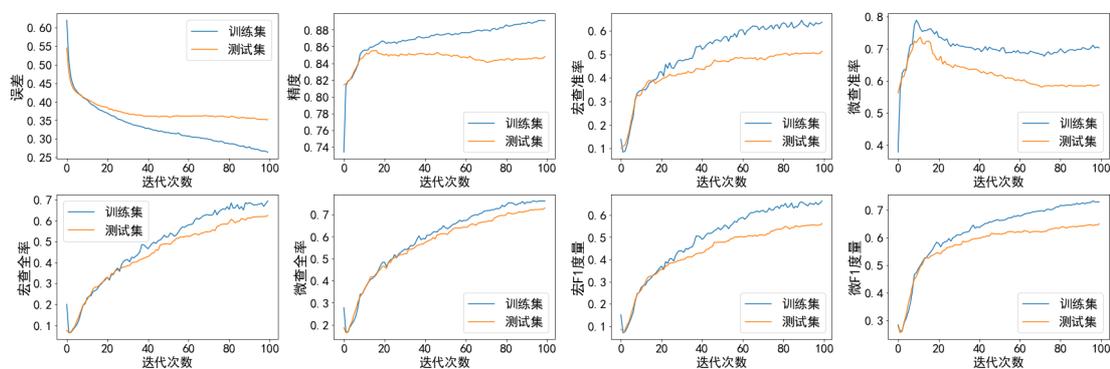


图 5-11 HOG 特征使用伪样本预训练显任务参数

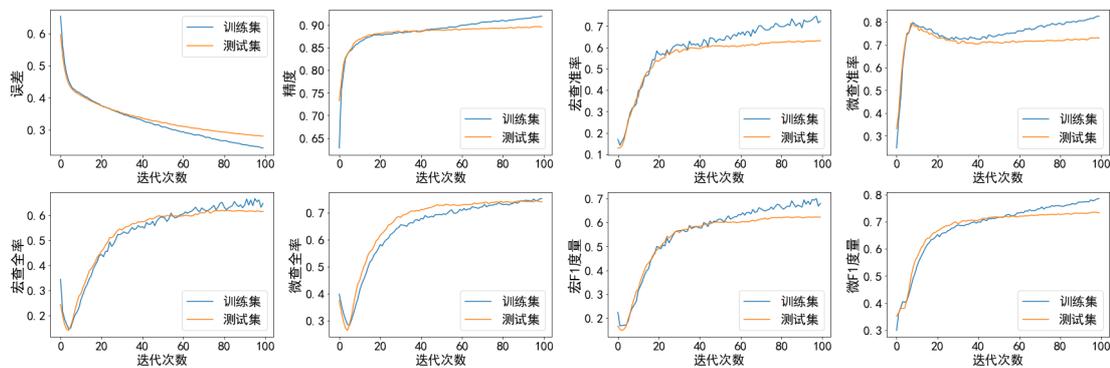


图 5-12 HOG 特征使用真实标签预训练显任务参数

模型的其他评价指标在训练集和测试集上的性能差距也都变小了，尤其是使用真实标签训练显任务时（图 5-12）。在不增加 AU 标注的数据量的情况下，模型的泛化能力得到了很大的提升（表 5-2），而这个提升的效果也依旧是使用真实标签预训练显任务的情况最大，引入显任务但不预训练显任务的情况最小。并且，这种性能的提升并没有显著牺牲模型在训练集上的情况，而是测试集向训练集的单方面靠拢（图 5-10 至 12），这说明模型在训练集上

获取到了更符合总体分布而非样本分布的信息。总体来说，引入半监督学习后，模型的整体性能有了较大提升。

表 5-2 HOG 特征不同情况度量指标对比

度量指标	单任务	不预训练显任务	用伪样本训练显任务	用真实样本训练显任务
泛化误差	0.2791	0.4146	0.3516	0.2797
精度	0.8896	0.8029	0.8479	0.8962
宏查准率	0.5272	0.4075	0.5126	0.6314
宏查全率	0.3702	0.4547	0.6244	0.6157
宏 F1	0.4336	0.4272	0.5618	0.6224
微查准率	0.8306	0.4952	0.5877	0.7289
微查全率	0.5509	0.6158	0.7297	0.7414
微 F1	0.6612	0.5482	0.6504	0.7338

5.3.3 使用卷积神经网络的 AU 检测

以上我们基于传统的特征提取方法进行了实验，但是传统的机器学习模型参数并不足够多，满足于较小的数据集，对半监督学习的需求并不是那么迫切，引入半监督学习提升的效果也仍然有限。那么，对于使用了深度神经网络，参数达到几百万个的现代机器学习模型来说，引入半监督学习的效果又如何呢？我们基于前文提到的 SmallVGGNet 进行了实验。

首先我们构建一个单任务的监督学习算法，并进行了 100 个迭代次数的训练，从结果中我们可以看出，得益于卷积神经网络强大的特征提取和拟合能力，模型得到了 0.03 的经验误差，而泛化误差却高达 0.53，二者相差了 17 倍。至于其他所有的评价指标，模型在训练集上都得到了超过 0.98 的成绩，而在测试集上这些成绩却并不太好，基本都在 0.55 以下（图 5-13）。

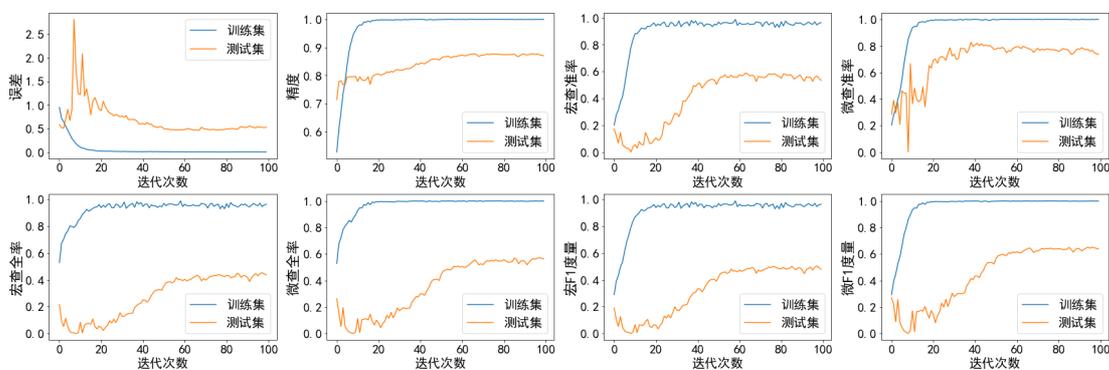


图 5-13 卷积神经网络单任务训练结果

接下来我们引入隐任务进行实验（表 5-3）。我们首先不预训练显层，同时优化 θ 和 ϕ ，经实验我们发现这种情况下模型的性能不仅没有得到任何的提高，反而还有所下降（图 5-14）。

表 5-3 卷积神经网络不同情况度量指标对比

度量指标	单任务	不预训练显任务	用伪样本训练显任务	用真实样本训练显任务
泛化误差	0.5305	0.5850	0.4918	0.3746
精度	0.8702	0.8336	0.8671	0.8975
宏查准率	0.5333	0.4489	0.5461	0.5930
宏查全率	0.4369	0.4328	0.5736	0.6695
宏 F1	0.4792	0.4394	0.5578	0.6283
微查准率	0.7354	0.5853	0.6636	0.7136
微查全率	0.5640	0.5648	0.7070	0.7920
微 F1	0.6371	0.5724	0.6832	0.7495

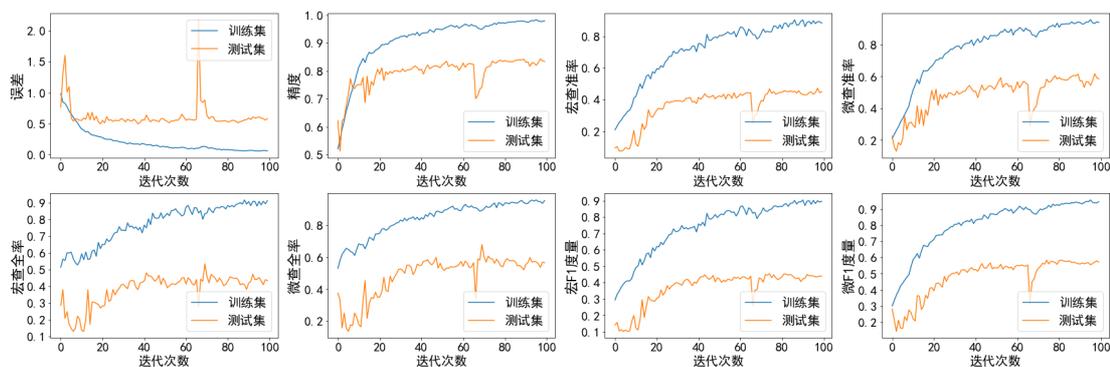


图 5-14 卷积神经网络引入显任务但不预训练显任务参数

但当我们使用伪样本预训练好显任务参数后，模型的泛化性能得到了显著的提升（图 5-15），经验误差和泛化误差的差距被缩小到了 9 倍，同时泛化误差相比不使用显任务时下降了 8%，泛化误差的下降在之前的实验中是从来没有出现的。除此之外，模型的宏 F1 上升了 16.4%，微 F1 上升了 7.9%。

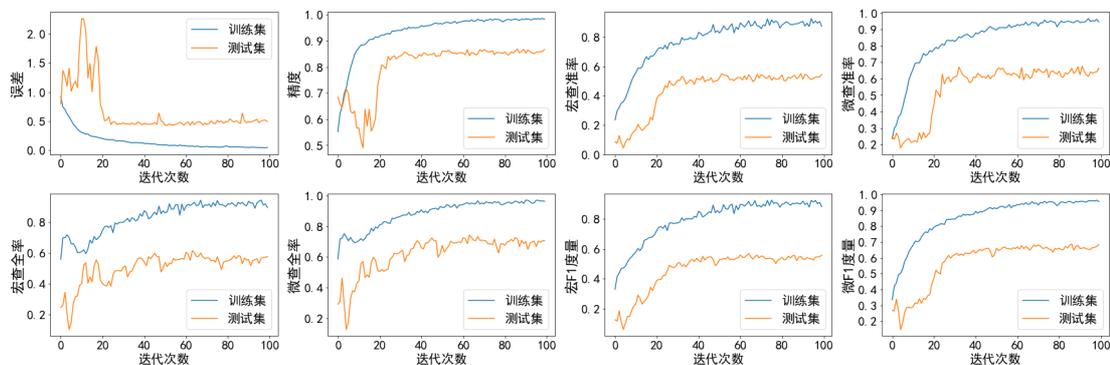


图 5-15 卷积神经网络使用伪样本预训练显任务参数

当我们使用真实标签训练显任务的参数后，模型的性能又有了进一步的提升（图 5-16）。模型的泛化误差相比不采用半监督学习时下降了 29.38%，宏 F1 上升了 31.1%，微 F1 上升了 17.6%。

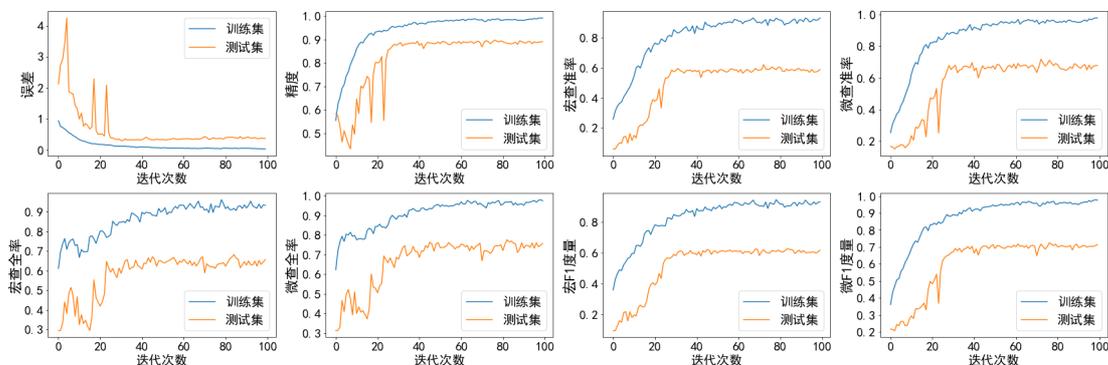


图 5-16 卷积神经网络使用伪样本预训练显任务参数

5.4 结果对比

结合表 5-1、5-2、5-3 我们通过对比发现，在三种特征提取方法中，卷积神经网络提取图像特征获得了最好的泛化能力，在每种引入显任务的方法下，其精度、宏 F1 和微 F1 都是最高的。

从所有评价指标在测试集上的提升幅度来看，使用人脸特征点作为特征时所有评价指标的平均提升幅度最大，为 74.46%，HOG 特征次之，为 23.38%，最后是卷积神经网络的 21.9%。人脸特征点方案是由于其原本效果就很差，提升空间较大，且就算提升了其绝对效果也很差。从所有评价指标提升的绝对值来看，卷积神经网络方法提升了 0.1125，而 HOG 特征提升了 0.1009，提升幅度相差 10%。因此引入显任务后，使用卷积神经网络的情况下显任务对泛化能力的提升是最大的。

并且考虑到卷积神经网络方法的各项评价指标在训练集上都能达到 0.98 以上，而 HOG 特征的则几乎都在 0.8 以下，前者的过拟合程度更高，因此，使用卷积神经网络的情况下显任务对过拟合程度的抑制是最大的。

5.5 本章小结

本章在使用不同特征提取方法的情形下对半监督学习算法对模型性能的提升做了测试，从测试中我们可以发现几个有意思的事实：

一是随着模型使用的特征提取器的特征提取能力的增加，半监督学习算法对模型的泛化性能的提升也不断增加。可以从注意力机制的角度出发说明

这个问题，当提取的特征包含的信息越多的时候，引入的半监督信息作为一种证据可以让隐任务更多地注意到与任务有关的特征，但当提取的特征包含的信息较少的时候，就算引入了半监督信息，它也只能“巧妇难为无米之炊”，无法将隐任务的注意力引导到根本不存在的信息上。

二是伪样本训练的显任务参数效果并不好，但却仍然能提高隐任务的性能。对此本文提出两种解释，首先是伪样本虽然效果较差，但其是基于领域知识生成，还是包含了一定的信息的，如果所采用的领域知识更加全面准确，那么伪样本训练的显任务参数效果应该也是很好的。其次是显任务在此时可以作为一种噪声源提高模型的泛化能力。

三是当显任务参数并不锁定的时候，引入半监督信息并不一定能提升隐任务的模型性能，有所提升时其提升的幅度也是最小的。这点可以从半隐任务学习的框架来进行解释，当两层参数都不锁定的时候，由于模型是基于总损失对二者同时进行求导，其梯度方向是同时优化两个参数的方向，最后得到的最优参数其实是二者的最优组合。

6 总结与展望

6.1 本文工作总结

本文主要基于半监督学习理论提出了一种结合多任务学习的半隐任务学习框架，即将两个可以相互串联的相关任务中的一个任务的参数固定，并同时提供两个任务的数据进行训练，这相当于对未固定参数的任务提供了半监督信息，从理论上来说，该框架是可以提高未固定参数的任务的性能的。

本文随后将此框架运用于人脸面部动作单元的认识，且对此方法进行了具体的实现。在小数据集的测试当中，本方法的确提高了识别任务的泛化性能，在不增大数据量的情况下提升了模型的效果，达成了研究的目的。并且，当模型的参数越来越多的时候，该方法对模型性能的提升越多。在深度学习逐渐流行的今天，这种方法是十分具有现实意义的。

6.2 未来工作展望

本文所提出的基于半隐任务学习的半监督学习框架在实践上成功地提升了原任务的泛化性能，但本文认为其还有许多可以改进和提升的地方，以下是笔者总结的一些问题：

(1) 由于批梯度下降法的算法所限，每批向隐任务和显任务投入的数据量必须相等，而在一个迭代次数中，投入两个任务的数据总量也必须相等，在两个训练集大小不一的情况下，只能舍弃一部分样本，或者让一些样本重复出现，这些都拖累了潜在的泛化性能提升。笔者设想，可以采用更灵活的设计，模型既可以投入两批数据，也可以只投入一批数据，在只投入一批数据的时候，模型只利用式 3-6 对参数进行优化。

(2) 本文所采用的伪样本采样算法并没有很好地模拟出真实样本的情况，而真实标签则只反映了数据集中 AU 和表情的依赖关系。本文并没有探究伪样本效果较差的原因是生成算法出现了问题，还是前人总结的领域知识不够全面。就伪样本采样算法的改进笔者设想，在前人的工作中有学者使用概率图模型捕获 AU 和表情的关系，那么也可以使用概率图模型来生成指定表情的 AU 样本。

(3) 本文仅在一个小数据集上进行了测试，并没有使用其他数据集测试模型的效果。未来可以寻找更多的数据集，测试模型的跨库泛化能力。

参考文献

- [1] 宗仁. 情绪识别如何拯救你的生命[EB/OL]. 硬创公开课, 2016-05-04. https://www.sohu.com/a/73422929_114877.
- [2] Mehrabian A. Communication without words[J]. *Psychology Today*, 1968, 2(4): 53-56.
- [3] Ekman P, Friesen W V, Ellsworth P. Emotion in the human face: guidelines for research and an integration of findings[M]. Oxford: Pergamon Press, 1972.
- [4] Littlewort, Gwen & Frank, Mark & Lainscsek, Claudia & Fasel, Ian & Movellan, Javier. Automatic Recognition of Facial Actions in Spontaneous Expressions[J]. *Journal of Multimedia*, 2006, 1(6): 22-35.
- [5] Valstar M, Pantic M. Fully automatic facial action unit detection and temporal analysis[C]. 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW 06). IEEE, 2006 149.
- [6] Lauren Van der Matten, Hendrik E. Action unit classification using active appearance models and conditional random fields[J]. *Cognitive processing*, 2012, 13(2): 507-518.
- [7] Lucey S, Ashraf A B, Cohn J. Investigating spontaneous facial action recognition through AAM representations of the face[M]. *Face recognition*, London: IntechOpen, 2007.
- [8] Tong Y, Liao W, Ji Q. Facial action unit recognition by exploiting their dynamic and semantic relationships[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2007, 29(10): 1683-1699.
- [9] Wang Z, Li Y, Wang S, et al. Capturing global semantic relationships for facial action unit recognition[C]//*Proceedings of the IEEE International Conference on Computer Vision*. 2013: 3304-3311.
- [10] Wang J, Wang S, Ji Q. Facial action unit classification with hidden knowledge under incomplete annotation[C]//*Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*. ACM, 2015: 75-82.
- [11] Adria Ruiz, Joost Van der Weijer, Xavier Binefa. From Emotions to action units with hidden and semi-hidden-task learning[C]//*2016 23rd International Conference on Pattern Recognition (ICPR)*. IEEE, 2016: 4089-4094.
- [12] Song Y, McDuff D, Vasisht D, et al. Exploiting sparsity and co-occurrence structure for action unit recognition[C]//*Automatic Face and Gesture Recognition (FG)*, 2015 11th IEEE International Conference and Workshops

- on Automatic Face and Gesture Recognition: volume I. IEEE, 2015: 1-8.
- [13] Wu S, Wang S, Pan B, et al. Deep facial action unit recognition from partially labeled data[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017, 3951-3959.
- [14] Ekman P, Friesen W. Facial Action Coding System: A Technique for the Measurement of Facial Movement[M]. Palo Alto: Consulting Psychologists Press, 1978.
- [15] Friesen W, Ekman P. EMFACS-7: Emotional Facial Action Coding System[J]. University of California at San Francisco, 1983, 2(36): 1.
- [16] Ekman P, Friesen WV, Hager JC. Facial Action Coding System: The Manual on CD ROM[EB/OL]. Salt Lake City: A Human Face, 2002. <http://www.cs.cmu.edu/afs/cs/project/face/www/facs.htm>.
- [17] Izard, C E. Measuring emotions in infants and children[M]. Cambridge Studies in Social and Emotional Development. New York: Cambridge University Press, 1982: 114-116.
- [18] Hjortsjö CH. Man's face and mimic language[M]. Lund: Studentlitteratur, 1970.
- [19] Hamm J, Kohler CG, Gur RC, Verma R. Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders[J]. Journal of Neuroscience Methods, 2011, 200 (2): 237-256.
- [20] C P Papageorgiou, M Oren, T Poggio. A general framework for object detection[C]//Sixth International Conference on Computer Vision, Bombay, India, 1998: 555-562.
- [21] Viola, Paul, Jones, Michael. Rapid Object Detection using a Boosted Cascade of Simple Features[C]//IEEE Conference Computer Vision Pattern Recognition, 2001.
- [22] Mcconnell, Robert K. Method of and apparatus for pattern recognition[P]. 1986. doi:US4567610 A.
- [23] Navneet D, Triggs B. Histograms of Oriented Gradients for Human Detection[C]//Proceedings of CVPR'05, San Diego, USA, 2005.
- [24] 傅红普, 邹北骥. 方向梯度直方图及其扩展[J]. 计算机工程, 2013, 39(05): 212-217.
- [25] 唐坤. 面部特征点定位算法研究(硕士学位论文)[D]. 殷福亮, 指导. 大连: 大连理工大学, 2013.
- [26] Kazemi and J. Sullivan, One millisecond face alignment with an ensemble of regression trees[C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, 2014: 1867-1874.

- [27] 李航. 统计学习方法[M]. 北京: 清华大学出版社, 2012: 67-72.
- [28] Fukushima K. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position[J]. *Biological Cybernetics*, 1980, 36(4): 193-202.
- [29] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang L, Wang G, Cai J, 2015. Recent advances in convolutional neural networks[EB/OL]. <https://arxiv.org/abs/1512.07108>.
- [30] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述[J]. *计算机学报*, 2017, 40(06): 1229-1251.
- [31] Goodfellow I, Bengio Y, Courville A. Deep learning[M]. Cambridge: MIT press, 2016: 326-366.
- [32] Dumoulin V, Visin F. A guide to convolution arithmetic for deep learning[EB/OL]. <https://arxiv.org/abs/1603.07285>.
- [33] 李金洪. 深度学习之 tensorflow 入门、原理与进阶实战[M]. 北京: 机械工业出版社, 2018.
- [34] 刘建伟, 刘媛, 罗雄麟. 半监督学习方法[J]. *计算机学报*, 2015, 38(08): 1592-1617.
- [35] 周志华. 机器学习[M]. 2016年1月第1版. 北京: 清华大学出版社, 2016.
- [36] Pratt, S Thrun. Guest Editors Introduction[J]. *Mach Learn*, 1997, 28(7): 5.
- [37] R. Caruana. Multitask Learning[J]. *Mach Learn*, 1997, 28(7): 41-75.
- [38] L Holmstrom, P Koistinen. Using additive noise in back-propagation training[J]. *IEEE Trans Neural Network*, 1992, 3(1): 24-38.
- [39] Liu S, Johns E, Davison A J. End-to-End Multi-Task Learning with Attention[C]//Proceedings of CVPR2019, 2019: 1871-1880.
- [40] 郭秀艳. 内隐学习研究综述[J]. *华东师范大学学报*, 2004, 22(1): 49-55.
- [41] A Ruiz, J V Weijer, X Binefa. From Emotions to Action Units with Hidden and Semi-Hidden-Task Learning[C]//2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015: 3703-3711.
- [42] DU S, TAO Y MARTINEZ A M. Compound facial expressions of emotion[J]. *Proceedings of the National Academy of Sciences*, 2014, 115(15): 1454-1462.
- [43] 彭国柱. 基于弱监督和半监督学习的面部动作单元识别 (硕士学位论文) [D]. 王上飞, 指导. 合肥: 中国科学技术大学, 2019.
- [44] LI Y CHEN J, ZHAO Y et al. Data-free prior model for facial action unit recognition[J]. *IEEE Transactions on Affective Computing*, 2013, 4(2): 127-141.

- [45] Zhang Yong, Dong Weiming, Hu Bao-Gang, Ji Qiang. (2018). Classifier Learning with Prior Probabilities for Facial Action Unit Recognition[C]// CVPR 2018, 2018: 5108-5116.
- [46] Bradski G. The OpenCV Library[EB/OL]. Dr. Dobb's Journal of Software Tools, 2000. <https://github.com/opencv/opencv/wiki/CiteOpenCV>.
- [47] Davis E, King. Dlib-ml: A Machine Learning Toolkit[J]. Journal of Machine Learning Research, 2009, 10: 1755-1758.
- [48] Martín Abadi, Ashish Agarwal, Paul Barham et al. TensorFlow: Large-scale machine learning on heterogeneous systems[EB/OL]. Software available from tensorflow dot org, 2015. <https://www.tensorflow.org/>.
- [49] K Simonyan, A Zisserman. Very deep convolutional networks for large-scale image recognition[EB/OL]. 2014, <https://arxiv.org/abs/1409.1556>.
- [50] Lucey P, Cohn J F, Kanade T, Saragih J, Ambadar Z, Matthews I. The Extended Cohn-Kanade Dataset (CK+): A complete expression dataset for action unit and emotion-specified expression[C]. Proceedings of the Third International Workshop on CVPR for Human Communicative Behavior Analysis, 2010: 94-101.

在学取得成果

一、 在学期间所获的奖励

2017.11 优秀共青团干部 共青团北京科技大学委员会

2018.11 优秀学生干部 北京科技大学

2018.11 三好学生 北京科技大学

二、 在学期间发表的论文

三、 在学期间取得的科技成果

致 谢

一场疫情的到来，把繁忙的世界突然推入了慢车道，尚怀着一丝不舍与错愕，我们就要毕业了。这是一段异于平常的经历，在经历了一个冬季的揪心，一个春天的迷茫，和一个夏日的复苏后，我们这届毕业生将度过一场少了告别，缺了合影，短了赠言的毕业季。但就是这样一段难熬的时光，却使我感到无比地充实和难忘。回首望去，正是北科大给我带来的成长，让我能以足够的勇气和学识去应对这一场人类的劫难和我自己的前进征程。在完成毕业论文之际，谨向各位指导我的老师、帮助我的同学以及支持我的家人致以最真诚的感谢。

我首先要感谢支瑞聪教授。作为一个双学位的学生，我在计通学院的学习过程中总有些羞涩，不敢与老师和同学沟通。但是支老师以她的耐心细致温柔引导我放开胆子，有问题就问，让我的论文写作过程变得顺畅了许多。同时，支老师还在毕业论文的撰写及答辩技巧方面给予了我很大的帮助，每次让老师帮我查看出现的问题，老师的认真和严谨都让我感到惭愧的同时也深受感动。

其次我要感谢刘梦祎学姐，她在论文撰写期间给予了我很大的帮助，在我每次遇到困难时，均能及时为我解答，并给予耐心指导，让各种难题迎刃而解，从而顺利地完成毕业设计。

最后还要感谢母校北京科技大学的各位老师的培养和指导，感谢参与我毕业论文评审以及答辩的各位老师付出的时间和精力，感谢所有曾经帮助或关心过我的人。

随着毕业论文的完成，大学生活也落下了帷幕。希望总有一天，师长和同窗们可以再次聚首，补齐我们因疫情而憾缺的毕业典礼！

