

# 基于深度学习的人脸遮挡检测技术研究

## 摘 要

在安防领域，为了逃避安防系统中监控摄像头的监视，进而实现其逃避法律处罚的目的，很多违法犯罪嫌疑人使用墨镜、围巾、口罩或者直接蒙面的方式来刻意遮挡自己的人脸特征，这种情况下的遮挡给嫌疑人抓捕和案件侦破带来了困难。如果能够利用人脸遮挡检测算法及时发现不法人员，然后自动向监控中心等部门报警，构建预警机制，则可以有效的阻止犯罪，降低犯罪率。

鉴于深度学习技术的强大表达能力和学习能力，本文利用深度学习来研究安防领域中的人脸遮挡检测技术，提出了一种基于级联卷积神经网络的人脸遮挡检测方法。该方法首先利用人头级联卷积神经网络来检测人头，然后针对眼睛和嘴巴经常被遮挡的事实，分别利用眼睛级联卷积神经网络和嘴巴级联卷积神经网络进行眼睛和嘴巴的遮挡检测。实验结果表明本文所实现的人脸遮挡检测算法在测试数据库上取得了令人满意的效果。

在此基础上，本文设计并实现了一个基于人脸识别算法的智能考勤系统，该系统集成了本文所提出的基于级联卷积神经网络的人脸遮挡检测算法，能够检测出认证过程中考勤员工的遮挡现象，并给予相关处理。

**关键词：**人脸遮挡检测，人脸识别，卷积神经网络，Caffe

# **RESEARCH ON FACE OCCLUSION DETECTION *BASED ON DEEP LEARNING***

## **ABSTRACT**

In order to evade surveillance cameras arranged in the public security field, many suspects in the crime are usually in appearance disguise, such as wearing sunglasses or wearing masks, for obscuring the face information. This behavior brings great difficulties for the policemen to catch the criminal suspect. Therefore, if a warning is sent to the monitoring center and thus some actions are adopted when the surveillance system automatically detects the occluded face of a customer, the crime is likely to be prevented.

Based on the powerful ability of feature representation and adaptive learning ability of deep learning, this thesis explores the face occlusion detection based on deep learning, and proposes a cascaded convolutional neural network based face occlusion detection algorithm. Because eyes and mouth are often occluded for many suspects in the crime, three cascaded convolutional neural network are designed for detecting the head, eyes and mouth with occlusion, respectively in the proposed algorithm. For each kind of convolutional neural networks trained in this paper, two convolutional neural networks are cascaded to implement coarse-to-fine occlusion detection. Our face occlusion detection algorithm achieves satisfactory results in two face database used in our experiments.

Furthermore, an attendance record system is designed and implemented which integrates the proposed face occlusion detection algorithm and convolutional neural networks based face recognition algorithm. The attendance record system can detect the occurrence of face occlusion during the process of checking attendance and give an alarm to the attendance personnel with occlusion.

---

**Key Words:** Face Occlusion Detection, Face recognition, Convolutional Neural Network, Caffe

目 录

摘 要 ..... i

ABSTRACT .....ii

第 1 章 绪论 ..... 1

    1.1 论文研究背景..... 1

    1.2 研究现状..... 4

    1.3 论文研究内容 ..... 6

    1.4 论文组织结构..... 7

第 2 章 卷积神经网络介绍 ..... 9

    2.1 卷积神经网络 CNN.....9

        2.1.1 卷积层..... 10

        2.1.2 子采样层 ..... 11

        2.1.3 稀疏连接 ..... 11

        2.1.4 权值共享 ..... 12

        2.1.5 Softmax 回归 ..... 12

    2.2 Caffe 架构 ..... 13

    2.3 本章小结 ..... 15

第3章 基于级联卷积神经网络的人脸遮挡检测方法..... 16

    3.1 算法框架 ..... 16-

    3.2 人头检测的级联网络 ..... 17-

    3.3 眼睛检测的级联网络 ..... 19

    3.4 嘴巴检测的级联网络 ..... 20-

    3.5 级联网络训练库 ..... 21

    3.6 测试结果 ..... 24-

    3.7 本章小结 ..... 26-

---

第 4 章 智能人脸考勤系统 .....  
27-  
4.1 系统设计..... 27-

---

|                  |                     |     |
|------------------|---------------------|-----|
| 4.2              | 人脸识别模块 .....        | 28- |
| 4.2.1            | 基于深度学习的人脸识别算法 ..... | 29  |
| 4.2.2            | PCA 降维 .....        | 31  |
| 4.2.3            | 实验结果 .....          | 31  |
| 4.3              | 系统实现 .....          | 33- |
| 4.4              | 本章小结 .....          | 36- |
| 第 5 章            | 结论与展望 .....         |     |
|                  |                     | 37- |
| 5.1              | 结 论 .....           | 37- |
| 5.2              | 展望 .....            | 38- |
| 参 考 文 献          | .....               |     |
|                  |                     | 39- |
| 致 谢              | .....               |     |
|                  |                     | 42- |
| 攻读学位期间参加的科研项目和成果 | .....               |     |
|                  |                     | 43- |



# 第 1 章 绪 论

## 1.1 论文研究背景

人脸识别技术以其主动性、非侵犯性、用户友好性、非接触性、不易察觉性和唯一性等优点，具有重要的学术研究价值以及广阔的应用前景[1]。人脸识别技术在近几十年中得到了飞速的发展，在中国就已广泛的应用于公安、安全、海关、金融、军队、机场、边防口岸、安防等多个重要行业及领域，以及智能门禁、门锁、考勤、手机、数码相机、智能玩具等民用市场。然而，在实际应用中，人脸识别依然面临着诸多的挑战。现有的人脸识别算法通常在室内可控环境下可以达到满意的结果。然而，在实际应用中，人脸认证系统对于人脸图像的采集一般均在非配合的环境下进行，并且人脸很容易受到光照、遮挡及表情等诸多因素的影响。这些干扰因素导致所获取的人脸图像之间存在较大的差异，给人脸识别带来了极大的困难，使其识别率难以达到理想水平。因此，随着人脸识别应用需求的不断增长，人脸识别这一课题还存在很多地方值得深入研究，特别是无约束环境下如何增强人脸识别系统的鲁棒性得到了广泛的关注和研究，已经成为当前研究的热点。

遮挡是众多影响人脸识别性能的因素中非常重要也是不可避免的一个因素，特别是在安全领域中尤其突出。造成遮挡的原因是多种多样。随着人们日常生活中佩戴眼镜的比例增大，由眼镜遮挡造成的问题也就越来越普遍；环境污染造成了出门佩戴口罩也越来越普遍了；在实际应用中，比如智能门禁、视频监控、保安系统、罪犯识别等，基本均在非配合的环境下进行人脸图像的采集，易被其他

人或者物所遮挡。此外，环境、饰物、刘海或胡须等均可造成遮挡现象的出现。图1-1给出了常用人脸数据库 LFW 数据库<sup>[2]</sup>和 AR 人脸数据库<sup>[3]</sup>中给出的部分遮挡的人脸图像样本，图1-2给出了一些日常生活中的遮挡实例<sup>4</sup>。造成遮挡的这

些干扰因素使得成像设备获取的人脸数据不完整，引起人脸部分信息甚至全部信

息的丢失，导致人脸识别系统无法提取完整有效的人脸信息，影响了整个人脸认证系统检测和识别的准确率。由于遮挡类型多样、位置随机、大小不确定，没有合适的方法对遮挡进行建模，导致遮挡问题处理起来非常困难。



(a)LFW 数据库[2]

(b)AR 数据库[3]

图1-1 人脸遮挡样本

Fig 1-1 Samples of face occlusion





图1-2 实际生活场景中的遮挡实例

如何有效检测和去除遮挡物的影响,成为了人脸检测与识别技术中亟待解决的关键问题。特别是在安防领域,为了逃避安防系统中监控摄像头的监视,进而实现其逃避法律处罚的目的,很多违法犯罪嫌疑人使用墨镜、围巾、口罩或者直接蒙面的方式来刻意遮挡自己的人脸特征(如图1-3所示),这种情况下的遮挡给嫌疑人抓捕和案件侦破带来了困难。如果能够利用人脸遮挡检测算法及时发现不法人员,然后自动向监控中心等部门报警,构建预警机制,则可以有效的阻止犯罪,降低犯罪率。尤其是报警并阻止蒙面人进入银行、政府区域、商场或者其他公共场所,能有效的阻止犯罪。



图1-3 ATM监控捕获到的面部存在遮挡的嫌疑人[13]

Fig 1-3 Suspects of face occlusion captured from ATM

人脸遮挡检测[12, 3139],也称为异常人脸检测[5-8, 14]或遮挡人脸检测[1126-29],是针对安防领域中人脸遮挡问题而提出的、面向应用的课题,是指用视频智能分析的方法检测出视频场景中是否有戴墨镜、戴围巾、戴口罩、戴面具、蒙面等遮挡物遮挡而不能被识别或者分类的人脸。通常,人脸遮挡检测系统利用监控摄像机连续的采集场景中的监控图像,利用计算机视觉技术检测监控图像中的人脸,并分析人脸上的器官是否不全(比如因为蒙面等原因使五官部分或全部被遮挡起来),如果是则向监控中心报警。人脸遮挡检测最初是基于银行大厅[6]、ATM机[7-14]的自动监控系统提出来的,旨在实现监控的智能化,及时发现不法人员并自动向监控中心报警。现在,人脸遮挡检测技术已经被扩展应用到其他安防领域,例如智慧社区的访客系统,基于人脸识别的移动支付等领域。

本文针对人脸识别在实际应用中所存在的遮挡问题，以安防领域为研究背

景，进行深入的探讨和系统的研究，结合信号表示与处理理论的最新发展，把深度学习理论应用到人脸遮挡检测问题，提高人脸识别系统在实际应用过程中的可操作性。因此，本文的研究既有一定的理论意义，也具有较强的实际应用价值。

## 1.2 研究现状

目前，相关文献中已经提出了一些人脸遮挡检测方法，这些方法一般可分为如下三种思路[11]。

(1) 基于特定遮挡类型的人脸遮挡检测方法：该类方法只检测特定的遮挡类型，例如戴头盔、戴墨镜、戴口罩、戴围巾等，遮挡物也相对固定。基于特定遮挡类型的检测方法通常根据不同种类的遮挡物利用 HOG[15]、LBP[16]、SIFT[17-18]、Haar-like[19]等特征提取方法来提取特征信息，然后基于这些提取到的特征信息利用支持向量机 (SVM)[20]、AdaBoost[21]、神经网络[22]等机器学习方法训练相应的分类模型，再使用训练出的分类模型来判断是否为异常人脸以及属于哪种异常类型。

文献[5]提出了基于颜色统计的墨镜检测方法，基于直线检测和颜色统计结合的口罩检测方法。文献[6]在序列图像的人头定位的基础上，依据正常人脸与异常人脸特征的差异来定位和识别异常人脸，实现了基于人脸中心线检测的墨镜检测、基于区域灰度分布特征的眼部区域的墨镜检测和基于直线检测的口罩检测。在文献[23-25]中，针对犯罪嫌疑人经常使用安全头盔遮挡人脸的情况，分别利用改进的霍夫变换、Canny 检测和 Gabor 滤波提出了针对佩戴安全头盔的人脸遮挡检测方法。文献[26]利用提取简单小波特征来实现遮挡物为眼镜的人脸遮挡检测。在文献[29]中，人脸遮挡被分为口罩围巾类的矩形遮挡和墨镜类的椭圆遮挡，然后利用基于自相连神经网络的分类器来检测输入人脸图像的遮挡区域。

基于特定遮挡类型的人脸遮挡检测方法通常能稳定的检测具有这些特定遮挡物的遮挡人脸。然而，在实际的应用当中，由于遮挡物的多变性，使得佩戴同样类型遮挡物的不同人脸表现出来的特征也不完全一样，例如同样是戴墨镜类型的异常人脸，墨镜的形状是多种多样的，可以是圆的、方的、扁的，墨镜的颜色

种类也非常多，例如黑色的、褐色的、红色的等，再加上使用场景中光照因素的影响，墨镜表现出的特征也不尽相同。另外，在实际的监控应用场景中，还可能会出现非预期的遮挡物情况，使得该类方法具有很大的局限性。

(2) 基于肤色的遮挡检测方法：该类方法主要利用遮挡物和人脸中的重要器官的位置、肤色之间的差异来确定遮挡人脸，将不同颜色空间下不同肤色模型应用到人脸检测中来。按照经验，通过分区域计算肤色比来判断是否遮挡，可以适应多样化的人脸遮挡情况。但该类方法通常对光照比较敏感。文献[10]用基于 YCbCr 颜色空间椭圆模型检测算法对区域人脸进行检测，以预先定义的异常行为语义来判断是否出现蒙面或脸部遮挡伪装取款的情况。文献[11]通过在 YCbCr 颜色空间中定义一系列非线性变换的规则，将肤色范围拟合到一个椭圆区域内，然后通过分区域的方法进行遮挡判定。文献[12]在定位人脸后通过计算皮肤面积比来确定人脸是否被遮挡。文献[14]利用基于经验规则的人脸定位方法和自适应的椭圆拟合法来两步定位人脸，然后利用 AdaBoost 来连接皮肤色彩检测和人脸模板匹配，最终实现 ATM 监控下的异常人脸检测。文献[31]利用基于 B 样条的主动轮廓法和椭圆拟合方法来估计人脸位置，然后利用测量到的不同区域的肤色比来确定遮挡类型。

(3) 基于正面裸露人脸的检测方法：该类方法根据眼睛、嘴部、鼻子、眉毛等正常人脸的特征检测正常的人脸及五官，如果发现已知有人的场景中检测不到人脸或五官，再根据不同种类的特征判断是否为异常人脸以及是何种异常[5]。该类方法把人脸遮挡检测问题转变成正常人脸检测和非正常人脸条件下的异常种类判断问题。

文献[7]提出基于与或图技术的多通道人脸检测技术来定位人脸区域，然后依据眼睛、鼻子、嘴等局部特征是否缺失来确定是否是异常人脸。文献[8]将人脸区域分为眼部区域和嘴部区域两部分，训练了基于 Haar 特征的人眼分类器来判断眼部是否存在遮挡，采用基于 Hough 变换的水平直线检测结合分类器搜索的方法来检测嘴部区域的异常。文献[9]提取 Harr-like 特征表示人脸，采用“积分图”方法提高计算速度，采用基于 Adaboost 算法进行正常人脸和异常人脸的分类。文献[11]提出了基于 Adaboost 的人脸部件检测器的遮挡判定方法，使用降级后的正面人眼检测器和嘴巴检测器，通过分区域的方法进行遮挡判定。文献[13]利用人脸器官局部区域识别能力和人脸的全局表征来双重确定是否是遮挡人脸，然后通过训练特定的遮挡检测器来提高检测精度。在文献[27]中，Yoon 和 Kee 第一次针对 ATM 机的安防需求，提出遮挡人脸的概念，并利用集成主成份分析

(PCA) 和 SVM 的方法来区分正常人脸和遮挡人脸。文献[30]通过 AdaBoost

检测器检测两只眼睛和嘴巴，使用神经网络分类正常或者非正常的人脸。

在实际应用场景中，人脸的检测通常是采用提取五官信息特征的方法来进行检测，但是当出现异常人脸情况时，五官往往是被遮挡住的，这样就给人脸检测带来困难，由于视频是动态的，在这种情况下也比较难定位出精确的人脸位置，从而给异常人脸类型的判断带来困难。

### 1.3 论文研究内容

近两年，深度学习的出现使人脸识别取得了突破性进展[32-34]。深度学习的初衷是模拟人脑的认知机理。在某些重要的人脸识别测试集上，深度学习模型已经达到甚至超出了人类的识别水平。这启发我们未来可以在更广的范围内更深入地研究这些模型中神经元响应的性质，并与认知科学的研究成果相结合，相互印证和启发。人脸识别的最新研究成果[35]表明，深度学习得到的人脸特征表达具有手工特征表达所不具备的重要特性，例如它是中度稀疏的、对人脸身份和人脸属性有很强的选择性、对局部遮挡具有良好的鲁棒性。这些特性是通过大数据训练自然得到的，并未对模型加入显式约束或后期处理。这说明深度学习并非单纯拟合数据集，其计算单元有清晰的语义含义，其高识别率可以通过这些特性来解释。

在2014年国际计算机视觉与模式识别会议(CVPR 2014)上，两个采用深度学习的团队——来自脸谱的团队[36]和香港中文大学的团队[37]，在允许利用有标签外部数据且非限定的测试条件下，在LFW 数据库[2]上分别报告了97.35%和97.45%的平均分类精度，比高维 LBP 特征方法的分类错误率降低了50%。上述两个团队均采用了卷积神经网络(Convolutional Neural Network, CNN)的变种架构。其中，脸谱的 DeepFace 方法[36]强调前端的人脸3D 对齐和虚拟正面化预处理，以削弱姿态变化的影响；而香港中文大学的 DeepID 方法[37]则强调采用多个人脸区块分别训练卷积神经网络，并最终融合形成人脸特征表示。最近，该团队进一步开发了DeepID2+ 系统[35]，在上述测试环境下取得了99.47%的正确分类精度，错误率比 DeepID 降低约80%。Google 推出的 FaceNet[38]再次刷新 LFW 数据库[2]上人脸验证的效果记录。

目前关于人脸遮挡检测对于人脸的拍摄角度、采集过程中的光照、遮挡物类型的多变性以及场景的复杂性等鲁棒性较弱，无法同时兼顾多种影响因素，很难满足实际要求。深度学习在人脸识别领域中得到了广泛的研究与应用，并取得

了很大的进展，但并没有相关工作把深度学习和人脸遮挡检测结合起来，因此，本文引入深度学习技术来研究异常人脸识别问题，利用深度学习强大的特征表达能力和学习能力来实现人脸遮挡检测。事实上，在本文的写作过程中，发现已有文献[39]在2015年8月召开的模糊系统与知识挖掘国际会议(2015 12th International Conference on Fuzzy Systems and Knowledge Discovery, FSKD15)上提出了结合多任务学习和卷积神经网络，实现基于多任务卷积神经网络的人脸遮挡检测方法，该方法所用的深度神经网络结构如图1-4所示。文献[39]的出现在很大程度上表明了本文提出的利用深度学习技术解决人脸遮挡检测思路的可行性和创新性。和文献[39]不同，本文利用 Caffe1401框架将深度学习应用到遮挡人脸的识别中，提出了一种先利用级联人头卷积神经网络进行人头定位，然后再基于眼睛级联卷积神经网络和嘴巴级联卷积神经网络进行人脸眼睛、嘴巴部位的遮挡检测方法，进而设计和实现了一个具有人脸遮挡检测功能的基于人脸识别的智能考勤系统。

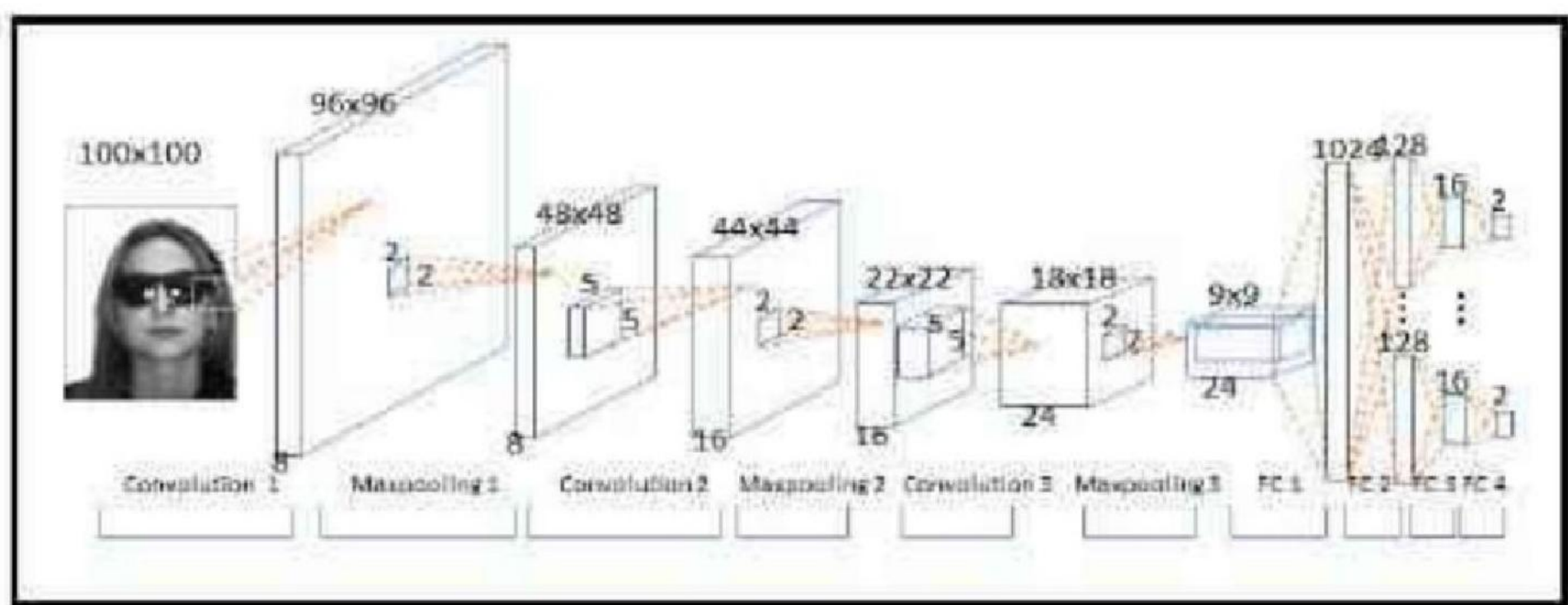


图1-4 基于多任务卷积神经网络 (CNN) 的人脸遮挡检测[39]

Fig1-4 Face occlusion detection based on multi-task convolution neural network

## 1.4 论文组织结构

全文共分五章，具体内容如下：

第1章：绪论，阐述了课题的研究背景和意义，总结了国内外关于人脸遮挡检测的现状。

第2章：深度学习理论介绍，介绍了卷积神经网络相关深度学习理论和 Caffe 深度学习框架的结构。

第3章：基于级联卷积神经网络的人脸遮挡检测，介绍了本文提出的基于级联卷积神经网络的人头检测和基于眼睛和嘴巴的级联卷积神经网络进行人脸遮挡检测方法的具体实现过程。

第4章：智能人脸考勤系统，介绍了集成了本文提出的基于级联卷积神经网络的人脸遮挡检测算法的智能考勤系统的系统框架设计和实现。

第5章：结论与展望，总结了本文的主要工作，指出了系统设计部分存在的不足和有待改进的地方，并对下一步工作进行了展望，提出了有待解决的问题。

## 第2章卷积神经网络介绍

卷积神经网络(Convolutional Neural Networks,CNN)是一种深层神经网络模型。卷积网络最初是受视觉神经机制的启发而设计的,是为识别二维形状而设计的一个多层感知器。这种网络结构对平移、比例缩放、倾斜或者其他形式的变形具有高度不变性。1962年 Hubel 和 Wiesel[41]通过对猫视觉皮层细胞的研究,提出了感受野的概念;1984 年 Fukushima[42]基于感受野概念提出的神经认知机模型,可以看作是卷积神经网络的第一个实现网络,也是感受野概念在人工神经网络领域的首次应用。神经认知机将一个视觉模式分解成许多子模式,然后进入分层式相连的特征平面进行处理,它试图将视觉系统模型化,使其能够在即使物体有位移或轻微变形的时候,也能完成识别。神经认知机能够利用位移恒定能力从激励模式中学习,并且可识别这些模式的变化。在其后的应用研究中, Fukushima 将神经认知机主要用于手写数字的识别。随后,国内外的研究人员提出了多种卷积神经网络形式,在邮政编码识别、车牌识别和人脸识别等方面得到了广泛的应用[47-49]。

卷积神经网络作为深度学习框架是基于最小化预处理数据而产生的,它通过共享局部权值有效降低了网络结构的复杂度,具有局部连接和权值共享特性。卷积神经网络(CNN)已经是现代智能识别语音和图像等信息的研究热点。它的局部连接和权值共享的网络结构使之更类似于生物神经网络,降低了网络模型的复杂度(对于很难学习的深层结构来说,这是非常重要的),减少了权值的数量。尤其在多维图像和大图像中该优点表现最为突出,可以直接将图像输入到网络中,不需要做更多的麻烦的传统特征提取和数据重建。

### 2.1 卷积神经网络 CNN

---

CNN 是第一个真正成功训练多层网络结构的学习算法。它利用权值共享来达到减少学习参数以用来提高 Back Propagation (BP)算法的训练性能。CNN 是一个最小化预处理要求下的深度学习框架。在CNN 中, 图像的一部分(局部感受区域)作为层级结构的最底层的输入, 信息再依次传输到不同的层, 每层通过一

个数字滤波器去获得观测数据的最显著的特征。这个方法能够获取对平移、缩放和旋转不变的观测数据的显著特征，因为图像的局部感受区域允许神经元或者处理单元可以访问到最基础的特征，例如定向边缘或者角点。

卷积神经网络是一个多层的神经网络，每层由多个二维平面组成，而每个平面由多个独立神经元组成。

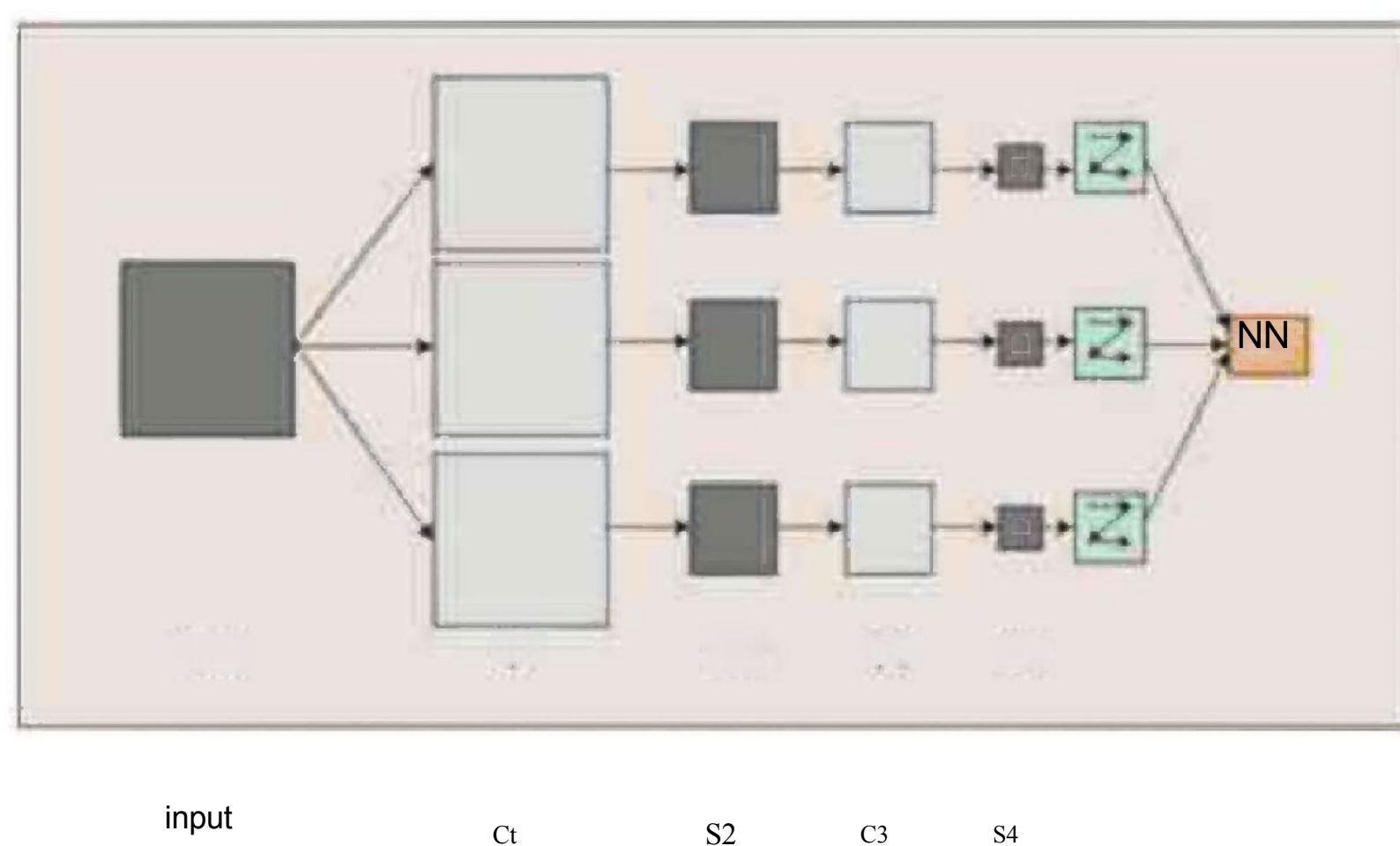


图2-1 卷积神经网络的概念示范

Fig 2-1 Convolutional Neural Networks (CNN)

图2-1 展示了卷积神经网络概念示范：输入图像通过和三个可训练的滤波器和可加偏置进行卷积，卷积后在C1 层产生三个特征映射图，然后特征映射图中每组的四个像素再进行求和、加权值、加偏置，通过一个 Sigmoid 函数得到三个 S2 层的特征映射图。这些映射图再经过滤波得到C3 层。这个层级结构再和 S2 一样产生 S4。最终，这些像素值被光栅化，并连接成一个向量输入到传统的神经网络，得到输出。

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^{l-1} + b_j^l\right)$$

### 2.1.1 卷积层

神经网络中的每个层有多个特征映射，每个特征映射通过一种卷积滤波器提取输入的一种特征，然后每个特征映射有多个神经元。卷积运算一个重要的特点就是，通过卷积运算，可以使原信号特征增强，并且降低噪音。假设第1层为卷积层，则第1层第j个特征映射的计算公式如(2.1)所示。

(2.1)

上面的\*号实质是让卷积核k在第1-1层所有关联的特征映射上做卷积运算，然后求和，再加上一个偏置参数，取 Sigmoid 得到最终激励值的过程。

### 2.1.2 子采样层

对于子采样层来说，有N 个输入特征映射，就有N 个输出特征映射，只是每个输出特征映射都变小了。采样公式如(2.2)所示。

$$x/=f(\beta' \text{down}(x/-)+b/) \quad (2.2)$$

down(.) 表示一个下采样函数。典型的操作一般是在输入特征矩阵中不同  $n \times n$  的块中取均值或最大值。这样输出图像在两个维度上都缩小了n 倍。每个输出 map 都对应一个属于自己的乘性偏置  $\beta$  和一个加性偏置b。

### 2.1.3 稀疏连接

卷积网络通过在相邻两层之间强制使用局部连接模式来利用图像的空间局部特性，在第m层的隐层单元只与第m-1 层的输入单元的局部区域有连接，第m-1 层的这些局部区域被称为空间连续的接受域。可将这种结构描述如下：

设第m-1 层为视网膜输入层，第m层的接受域的宽度为3, 也就是说该层的每个单元仅与输入层的3个相邻的神经元相连，第m层与第m+1 层具有类似的链接规则，如图2-2所示。

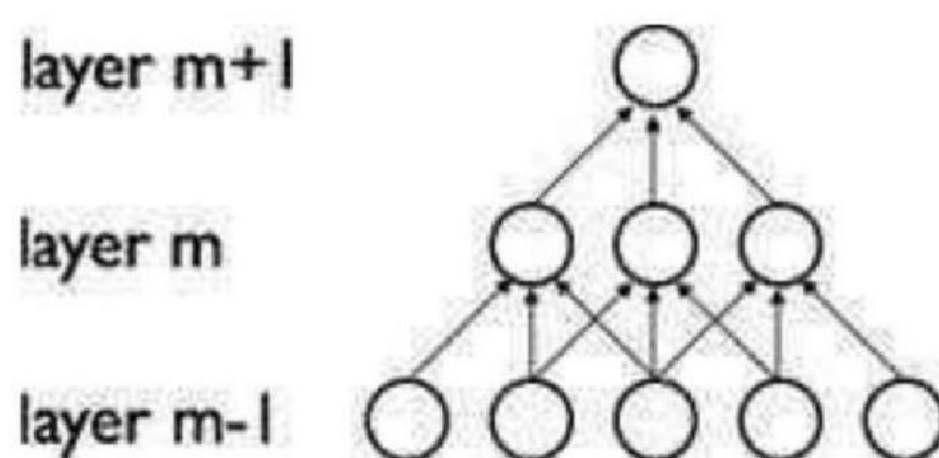


图2-2 稀疏连接示意图

---

可以看到 $m+1$  层的神经元相对于第 $m$  层的接受域的宽度也为3, 但相对于输入层的接受域为5, 这种结构将学习到的过滤器限制在局部空间模式。从图2-2也可以看出, 多个这样的层堆叠起来后, 会使得过滤器逐渐成为全局的。例如上图中第 $m+1$  层的神经元可以对宽度为5的输入进行一个非线性的特征编码。

#### 2.1.4 权值共享

在卷积网络中，每个稀疏过滤器通过共享权值都会覆盖整个可视域，这些共享权值的单元构成一个特征映射，如图2-3所示。

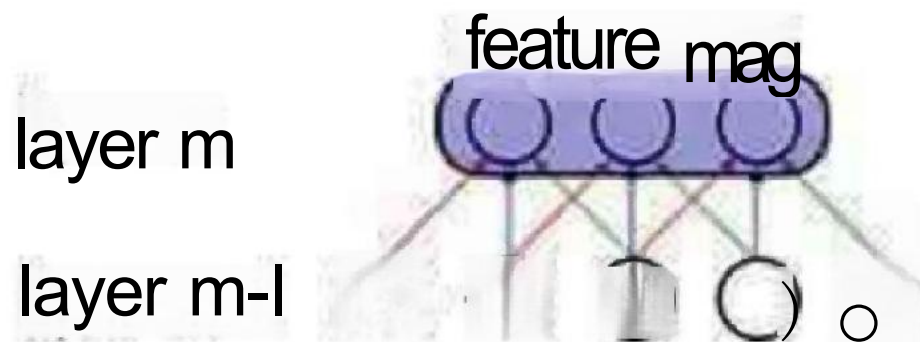


图2-3 特征映射

Fig 2-3 Feature map

在图2-3 中，有3个隐层单元，他们属于同一个特征映射。同种颜色的链接的权值是相同的，可以使用梯度下降的方法来学习这些权值，只需要对原始算法做一些小的改动，这里共享权值的梯度是所有共享参数的梯度的总和。一方面，重复单元能够对特征进行识别，而不考虑它在可视域中的位置。另一方面，权值共享能更有效的进行特征抽取，因为它极大的减少了需要学习的自由变量的个数。通过控制模型的规模，卷积网络对视觉问题可以具有很好的泛化能力。

#### 2.1.5 Softmax 回归

Softmax回归其实就相当于多类别情况下的逻辑回归，对比如下。

(1) 逻辑回归的假设函数 (hypothesis)

(2.3)

公式(2.3)给出了逻辑回归的假设函数。整个逻辑回归模型的参数就是  $\theta$ ， $h(\cdot)$ 是Sigmoid 函数，输出在0~1 之间，一般作为二分类算法。对于具体的问题，找出最合适的  $\theta$  便是最重要的步骤，这是最优化问题。一般通过定义代价函数，然后最小化代价函数来求解，逻辑回归的代价函数如公式(2.4)所示。

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m y^{(i)} \log h_{\theta}(x^{(i)}) + (1 - y^{(i)}) \log(1 - h_{\theta}(x^{(i)})) \right] \quad (2.4)$$

最小化 $J(\theta)$ , 一般采用梯度下降算法, 迭代计算梯度并更新  $\theta$ 。

(2) Softmax 的假设函数

$$h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (2.5)$$

公式(2.5)给出了 Softmax 的假设函数。逻辑回归里将  $-O \cdot x$  作为 sigmoid 函数的输入，得到的是0或者1, 两个类别。而 Softmax有  $k$  个类别，并且将  $-O \cdot x$  作为指数的系数，所以就有  $e^*$  至  $e^*$  共  $k$  项，然后除以它们的累加和，这样做就实现了归一化，使得输出的  $k$  个数的和为1, 而每一个数就代表那个类别出现的概率。因此，Softmax 的假设函数输出的是一个  $k$  维列向量，每一个维度的数就代表那个类别出现的概率。Softmax 的代价函数如公式(2.6)所示：

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=1}^n \theta_{ij}^2 \quad (2.6)$$

本质上跟逻辑回归是一样的，采用代价函数NLL，如果加上权重衰减项(正则化项)，则为：

(2.7)

最小化代价函数，同样可以采用简单而有效的梯度下降，需要提到的是，在程序实现中，一般采用批量随机梯度下降，即 minibatch Stochastic Gradient Descent (MSGD)，简单来说，就是每遍历完一个batch 的样本才计算梯度和更新参数，一个 batch 一般有几十到几百的单个样本。

## 2.2 Caffe 架构

Caffe[51]是一个清晰而高效的深度学习框架，在处理图像方面有着独具一格的优势，其作者是博士毕业于UC Berkeley的贾扬清，目前在 Google 工作。Caffe[40]是纯粹的C++/CUDA 架构，支持命令行、Python 和 MATLAB 接口；可以在CPU 和 GPU 直接无缝切换。

Caffe 具有如下几个优势:

(1)上手快: 模型与相应优化都是以文本形式而非代码形式给出。Caffe<sup>[40]</sup>给出了模型的定义、最优化设置以及预训练的权重, 方便立即上手。

(2)速度快: 能够运行最棒的模型与海量的数据。Caffe 与 cuDNN 结合使用, 测试AlexNet 模型, 在K40 上处理每张图片只需要1.17ms。

(3)模块化: 方便扩展到新的任务和设置上。可以使用Cafe 提供的各层类型来定义自己的模型。

(4)开放性: 公开的代码和参考模型用于再现。

(5)社会好: 可以通过BSD-2 参与开发与讨论。

Caffe<sup>[40]</sup>主要分为四个大类: Blob,Layer,Net,Solver。这四个类复杂性从低到高, 贯穿了整个Caffe<sup>[40]</sup>。

(1)Blob: 是基础的数据结构, 是用来保存学习到的参数以及网络传输过程中产生数据的类。

(2)Layer: 是网络的基本单元, 由此派生出了各种层类。修改这部分的人主要是研究特征表达方向的。

(3)Net: 是网络的搭建, 将Layer 所派生出层类组合成网络。

(4)Solver: 是 Net 的求解, 修改这部分人主要会是研究DL 求解方向的。

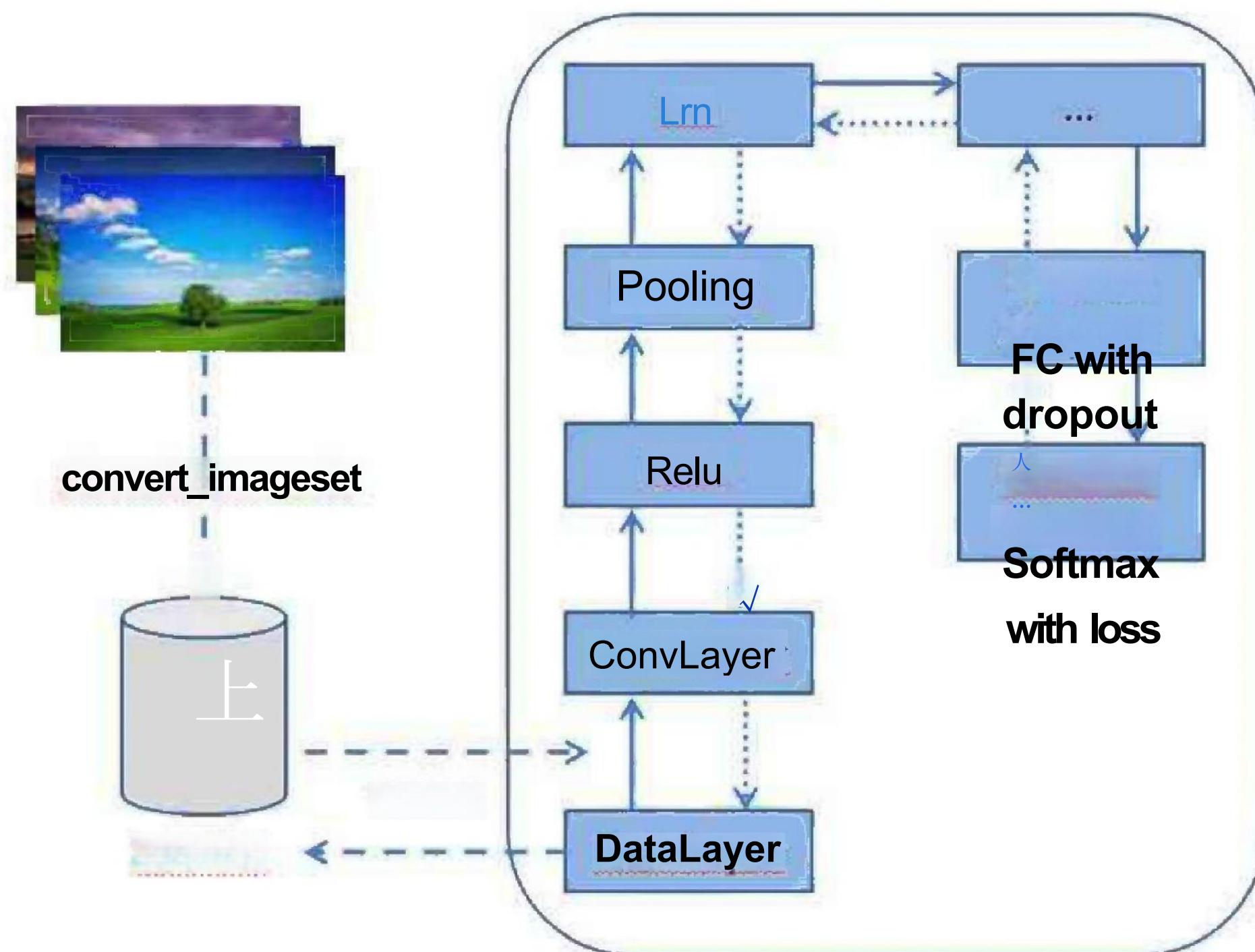


图2-4 Caffe[40]工作流程

Fig 2-4 Flowchart of Caffe

图2-4展示了 Caffe[40]工作的流程。基本上，Caffe[40]遵循了神经网络的一个简单假设——所有的计算都是以层的形式表示的，每层做的事情就是获取一些数据，然后输出一些计算以后的结果。比如说卷积，就是输入一个图像，然后和这一层的参数过滤器做卷积，然后输出卷积的结果。每一层需要做两个计算：前向反馈是从输入计算输出，然后反向反馈是从上面给的梯度来计算相对于输入的梯度。只要这两个函数实现了以后，就可以把很多层连接成一个网络，这个网络做的事情就是输入数据（图像或者语音或者其它的），然后来计算需要的输出（比如说识别的标签）。在训练的时候，可以根据已有的标签来计算损失和梯度，然后用梯度值来更新网络的参数。这个就是Caffe[40]的一个基本流程。

### 2.3 本章小结

卷积神经网络是一种深层神经网络模型，已经是现代智能识别语音和图像等信息的研究热点。本章首先简要介绍了卷积神经网络CNN 相关理论，然后介绍

了本文所采用的深度学习工具Caffe[40]的基本框架。

## 第3章基于级联卷积神经网络的人脸遮挡检测方法

现有的人脸遮挡检测算法通常需要事先设计人头检测或拟合算法来定位人头，然后利用各种特征提取方法提取特征，将空间域像素信息映射到高层次，以更抽象的形式进行表征，然后利用不同的机器学习算法训练分类器，最终利用训练好的分类器判断输入的图像是否包含遮挡的人脸。在这个过程中人脸遮挡检测算法的精度通常取决于所设计的人头定位算法的精度、所抽取的特征对遮挡检测的有效性。

深度学习本质上是一种特征学习方法，把原始数据通过一些简单的、非线性的模型转变成为更高层次的、更加抽象的表达，进而强化输入数据的判别能力，同时削弱不相关因素。此外，深度学习另一个优点是各层的特征都不是利用人工工程来设计的，而是使用一种通用的学习过程从数据中学到的，从而避免了在人脸遮挡检测过程进行特征抽取算法的选择。因此，利用深度学习技术来提高人脸遮挡检测的精度是一种可行的思路。然而相关研究工作还没有引起关注，在本文写作过程中才检索到一篇基于多任务卷积神经网络进行人脸遮挡检测的论文[39]。

### 3.1 算法框架

脸部局部特征中比较明显的部位有眼睛、鼻子和嘴，由于这些特征的形状、大小和结构上的差异使得人脸千差万别。这些局部特征可作为人脸检测的重要特征。而眼睛和嘴巴是最经常被遮挡的器官，因此，本文主要考虑检测人脸部眼睛和嘴巴的遮挡。

本章提出通过级联卷积神经网络来定位人头，在此基础上分别利用眼睛和嘴巴的级联卷积神经网络来检测人脸遮挡情况，有效利用深度学习结构的自主学习能力，克服单一低层特征判别力较差的缺点。

本章基于 Caffel40]框架的人脸遮挡检测采集静态二维图像进行级联网络的人头检测，同时以卷积神经网络自主学习眼睛和嘴巴特征，分别进行眼睛检测和嘴巴检测。为了加快检测速度和检测准确率，本章各个部分的检测网络均有两个

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：  
<https://d.book118.com/478105104126006055>