

西安电子科技大学

# 硕士学位论文



基于Faster R-CNN的人脸检测与识别算法研究  
与实现

作者姓名\_\_\_\_\_尉冰\_\_\_\_\_

学校导师姓名、职称\_\_\_\_\_苗启广\_\_\_\_\_教授

企业导师姓名、职称\_\_\_\_\_钟升\_\_\_\_\_研究员

申请学位类别\_\_\_\_\_工程硕士\_\_\_\_\_

## 西安电子科技大学 学位论文独创性（或创新性）声明

秉承学校严谨的学风和优良的科学道德，本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢中所罗列的内容以外，论文中不包含其他人已经发表或撰写过的研究成果；也不包含为获得西安电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同事对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文若有不实之处，本人承担一切法律责任。

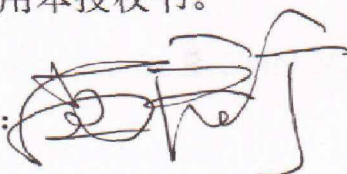
本人签名：尉冰 日期：2017.6.15

## 西安电子科技大学 关于论文使用授权的说明

本人完全了解西安电子科技大学有关保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权属于西安电子科技大学。学校有权保留送交论文的复印件，允许查阅、借阅论文；学校可以公布论文的全部或部分内容，允许采用影印、缩印或其它复制手段保存论文。同时本人保证，结合学位论文研究成果完成的论文、发明专利等成果，署名单位为西安电子科技大学。

保密的学位论文在年解密后适用本授权书。

本人签名：尉冰 导师签名：



日期：2017.6.15 日期：2017.6.15

学校代码 10701

分 类 号 TP311

学 号 1403121768

密 级 公开

# 西安电子科技大学

## 硕士学位论文

### 基于Faster R-CNN的人脸检测与识别算法研究 与实现

作者姓名：尉冰

领 域：计算机技术

学位类别：工程硕士

学校导师姓名、职称：苗启广 教授

企业导师姓名、职称：钟升 研究员

学 院：计算机学院

提交日期：2017 年 5 月

# **Research and Implementation of Face Detection and Recognition based on Faster R-CNN**

A thesis submitted to  
XIDIAN UNIVERSITY  
in partial fulfillment of the requirements  
for the degree of Master  
in Computer Technology

By

Wei Bing

Supervisor: Miao Qiguang    Professor

May 2017

## 摘要

随着计算机、互联网、海量存储等信息技术的飞速发展，身份鉴定技术在现在社会中的地位更加凸显，信息安全的作用也越来越重要，如何快速有效地进行身份鉴定，成为深度学习研究的一个重要方向，而人脸检测与人脸识别是身份鉴定技术中重要的方式之一。

Hinton 教授提出深度学习理论之后，得到了学术界的普遍关注，越来越多的学者利用深度学习去解决人脸检测与识别的问题，深度学习中的卷积神经网络模型是最常用于人脸检测与识别的模型。本文通过分析 CNN 模型对图像特征的表达特性，构造了一个 CNN 网络模型，该模型在传统的 VGG 模型上加以改进，本文的工作主要有：

1. 通过对 Faster R-CNN 算法的深入研究，在 RPN 网络模型的基础上，采用固定尺寸分割策略代替滑动窗口，给出了一种改进的 RPN 网络模型。RPN 网络模型通过滑动窗口的方式在最后一层卷积特征图上进行穷举，由于滑动窗口的大小是固定的，且会产生较多的窗口冗余，所以本文采用固定尺寸分割策略代替滑动窗口，从而可以产生更多尺寸的目标估计，对极端情况的人脸图像有很好的检测结果。

2. 采用空间金字塔池化技术(SPP)解决了 VGG 网络模型输入图像尺寸限制的问题。在对 VGG 网络模型训练前，如果输入的人脸图像样本不符合 VGG 网络模型的要求，就需要对人脸图像进行缩放，而缩放处理会造成一定程度上的图像形变，可能会导致图像空间信息的损失。针对这个问题进行研究发现，是因为全连接层要求输入的维度必须是固定不变的，从而在 VGG 网络里引入了 SPP，将卷积层和全连接层连接起来，从而不再限制输入的图像大小。SPP 池化技术可以提高图像提取的特征的表达能力，进而可以提高 VGG 模型对人脸识别的准确率。

3. 设计并实现了一套基于 Faster R-CNN 算法的人脸检测与识别软件，该软件测试结果表明，可以达到较高的检测率和识别率，而且对于不同姿态和表情的人脸检测与识别，具有较强的鲁棒性。

实验测试结果表明，基于改进的 VGG 模型对于人脸识别的准确率达到 99%。

**关键词：**人脸检测， 人脸识别， Faster R-CNN 算法， VGG 网络模型

## ABSTRACT

With the rapid development of information technology such as computer, Internet, mass storage and so on, the status of identity authentication technology is more and more prominent in the present society. The role of information security is becoming more and more important, especially carry out identity authentication quickly and effectively. Face detection and recognition is one of the important ways of identification technology.

The method of deep learning advanced by professor Hinton has been paid great attention by the academic community, and more and more scholars have used the depth learning to solve the problem of face detection and recognition. The convolution neural network model in depth learning is the most commonly used model in human face detection and recognition. In this thesis, a CNN model is constructed by analyzing the CNN model's image characteristics. The model is improved on the basis of traditional VGG model. The three aspects of innovation of this thesis is as follows.

Through the intensive study of the Fast R-CNN algorithm, it is found that in the RPN network model, exhaustion has been done on the last layer convolution feature by sliding the window, but it produces large redundancy because the size of the sliding window is fixed. Therefore, in this thesis, a modified RPN network (IRPN) is proposed by using the fixed size segmentation strategy instead of the sliding window, and the detection performance of the face has been improved obviously.

VGG network requires fixed size of the input image size. If the size of the face image does not meet the requirements of convolution neural network, you need to scale the face image, but scaling processing will cause the loss of image space information in a certain degree. After studying this problem, the reason has been found: it is because the dimension of the fully connection layer must be fixed. So in this thesis, the space pyramid pooling technology is introduced in the VGG network to improve the accuracy of the VGG model for face recognition.

In this thesis, a set of face detection and recognition software based on Faster R-CNN been designed and implemented. The testing of the software shows, the rate of face detection

and recognition can be more higher, at the same time, the software has strong robustness for faces with different gestures and expressions.

Experimental results show that the accuracy rate of face recognition based on improved VGG model is 99%.

**Keywords:** Face detection, face recognition, Faster R-CNN algorithmn, VGG network model

## 插图索引

图 1.1 自动人脸识别流程图 .....	3
图 2.1 R-CNN 算法流程示意图 .....	12
图 2.2 Fast R-CNN 算法流程图 .....	12
图 2.3 RPN 结构示意图 .....	13
图 2.4 合并的网络结构示意图 .....	14
图 2.5 前向传输图 .....	15
图 2.6 Faster R-CNN 算法流程图 .....	16
图 2.7 全连接和局部连接图 .....	17
图 2.8 局部连接图和卷积图 .....	18
图 2.9 经典神经网络结构图 .....	19
图 2.10 卷积层和下采样层 .....	20
图 2.11 全连接层示意图 .....	20
图 2.12 线性分类超平面 .....	21
图 3.1 IRPN 网络示意图 .....	23
图 3.2 SPP 示意图 .....	26
图 3.3 改进的 VGG 网络模型 .....	27
图 3.4 模型训练前的准备 .....	29
图 3.5 基于 Caffe 平台的 CNN 模型流程图 .....	31
图 3.6 Fddb 部分数据集示例图 .....	32
图 3.7 WIDER 部分数据集示例图 .....	32
图 3.8 LFW 部分数据集示例图 .....	33
图 3.9 CelebA 部分数据集示例图 .....	33
图 3.10 模型的部分识别结果示意图 .....	34
图 3.11 改进的 VGG 模型对人脸识别数据集分类结果图 .....	34
图 4.1 基于 Faster R-CNN 的人脸检测与识别软件框架 .....	38
图 4.2 人脸检测与识别系统的工作流程图 .....	39
图 4.3 人脸检测类图 .....	39
图 4.4 人脸识别类图 .....	40
图 4.5 人脸检测与识别算法功能模块设计 .....	40
图 4.6 人脸检测算法流程图 .....	41
图 4.7 人脸识别算法流程图 .....	42



图 4.8 软件主界面运行图.....	43
图 4.9 人脸检测结果示意图.....	43
图 4.10 人脸识别效果示意图.....	44

## 表格索引

表 3.1	分类器易混淆名词矩阵.....	29
表 3.2	几种典型的人脸识别算法结果统计表.....	35

## 符号对照表

符号	符号名称
$W$	权值
$AP$	分值正确率
$X$	样本集

## 缩略语对照表

缩略语	英文全称	中文对照
SVM	Support Vector Machine	支持向量机
BP	Back Propagation	反向传播算法
CNN	Convolutio Nerual Network	卷积神经网络
HOG	Histogram of Oriented Gradient	方向梯度直方图
LBP	Local Binary Pattern	局部二值模式
SIFT	Scale-invariant Feature Transform	尺度不变特征变换

# 目录

摘要 .....	I
ABSTRACT .....	III
插图索引 .....	V
表格索引 .....	VII
符号对照表 .....	IX
缩略语对照表 .....	XI
<b>第一章 绪论</b> .....	<b>1</b>
1.1 本课题研究背景和意义 .....	1
1.2 国内外研究现状 .....	2
1.2.1 人脸检测的研究现状 .....	2
1.2.2 人脸识别的研究现状 .....	3
1.3 论文主要研究工作 .....	5
1.4 本文章节安排 .....	5
<b>第二章 相关理论基础</b> .....	<b>7</b>
2.1 人脸检测与人脸识别理论 .....	7
2.1.1 基于显式特征的人脸检测方法 .....	7
2.1.2 基于隐式特征的人脸检测方法 .....	7
2.1.3 人脸识别理论 .....	8
2.2 R-CNN 算法介绍 .....	10
2.3 VGG 网络模型 .....	16
2.3.1 卷积神经网络结构 .....	16
2.3.2 VGG 网络模型介绍 .....	19
2.3.3 分类器 .....	20
2.3.4 卷积神经网络的训练过程 .....	21
2.4 本章小结 .....	22
<b>第三章 基于 Faster R-CNN 算法的人脸检测与识别</b> .....	<b>23</b>
3.1 Faster R-CNN 人脸检测算法的改进 .....	23
3.1.1 RPN 网络的改进 .....	23
3.1.2 非极大抑制 .....	24
3.1.3 构建损失函数 .....	24
3.1.4 模型交替训练策略 .....	24

3.1.5 ReLU 函数的改进 .....	25
3.2 空间金字塔池化 .....	26
3.3 基于改进的 VGG 网络模型人脸识别算法.....	27
3.4 数据集预处理 .....	28
3.5 模型的训练 .....	28
3.6 人脸检测与识别算法评价标准 .....	29
3.7 实验结果与分析 .....	30
3.7.1 实验环境.....	30
3.7.2 实验的数据集.....	31
3.7.3 改进的 VGG 模型的参数选择.....	33
3.7.4 实验结果与分析.....	35
3.8 本章小结 .....	35
第四章 基于 Faster R-CNN 的人脸检测与识别软件设计 .....	37
4.1 软件设计目标 .....	37
4.2 软件系统设计 .....	37
4.2.1 系统结构设计.....	37
4.2.2 人脸检测与识别算法功能模块设计与实现.....	38
4.3 软件检测与识别结果 .....	42
4.4 本章小结 .....	44
第五章 总结与展望 .....	45
5.1 总结 .....	45
5.2 展望 .....	46
参考文献.....	49
致谢.....	53
作者简介.....	55

## 第一章 绪论

### 1.1 本课题研究背景和意义

近些年来,随着摄像机、照相机和多媒体技术等电子设备的普及和互联网技术的飞速发展,信息安全的重要性更加突出,而身份鉴定技术是保证信息安全的重要手段之一,在现代社会中的地位也越来越凸显。目前,个人的身份鉴定主要通过验证 ID 卡(如护照、工作证、身份证、银行卡等)和输入密码两种手段,然而 ID 卡存在携带不方便、容易丢失、使用时间过长或使用方式不当容易造成损坏的问题,而密码则存在容易被遗忘或者被破解等问题。生物识别技术可以将传统的信息技术和生物技术结合在一起,具有更高的安全性。

人脸识别<sup>[1]</sup>作为一种重要的生物识别技术,因为具有不易被复制性、非侵入性、唯一性、便捷性、安全性等特性,拥有着广泛的应用前景和科研价值。同时,IBM 提出了“智慧地球”的概念,“智慧地球”的一个重要方面就是人工智能,而人脸识别是人工智能的一个重要分支。在过去的几十年里,随着人工智能的发展,人脸识别技术也取得了长足的进步,在特定的理想环境下,人脸识别算法已经取得了很好的结果。人脸识别在现实生活中的很多方面,都可以给人们提供便利,例如以下几个方面:

1) 可以对认证和注册的流程进行简化。假设你想要开一个微店,通过互联网方便地对自己家农特产进行出售,或者注册一个微信公众号对自己的一些想法进行宣传,出于安全和信用两个方面的考虑,你可能需要通过手持身份证拍照的方式对自己的身份进行确认。但人脸识别技术却能通过以人脸为独立验证 ID (IDentity) 的方式简化注册和验证的流程。

2) 在安全领域方面,和通过验证由普通字符、字串组成的密码方法相比,成熟完善的人脸识别技术肯定更加便捷和安全。当用户需要进行支付时,用人脸验证的方式代替复杂的密码输入操作,可以在很大程度上减少用户用于支付的时间;当用户出现密码被盗的情况时,不仅可以通过人脸 ID 迅速找回,还能通过后台保存的用户人脸信息避免密码被轻易修改。

3) 人脸识别在一定程度上,也可以帮助警察抓捕嫌犯。腾讯的优图人脸识别就和公安机构在人脸识别方面达成了合作。

但是,当所处环境发生剧烈变化(如姿态的变换、各种夸张的表情、画质不清晰、脸部部分被遮挡)时,进行人脸识别的难度就会大幅增加,与此同时,识别的效果也会急剧变差。人脸识别的难点在于人脸图像具有较大的类内变化和较小的类间差异的特性,并且由于人脸图像容易受到自然环境的影响,例如光照的强度、拍摄的角度、

拍摄时的表情、年龄的增长、是否化妆、是否存在遮挡和图像质量的变化等，在不同的自然环境下，同一个人的人脸图像的差异性会比较大。

随着科技和社会的不断进步，医疗水平的不断提升，人类社会逐渐走向“低增长，低死亡”的老龄化趋势，造成人类社会劳动力不足，在很多领域需要机器人代替人类去工作，而代替人类工作的前提，首先需要做到使机器人和人类一样可以用眼睛去“看”。一般来说，人脸的识别过程主要包括输入图像、人脸检测、特征点的定位、特征提取与分类器的设计四个步骤。其中，人脸检测是指通过输入的图像，对所有人脸（如果存在）的位置、尺寸大小和人脸表情姿态进行确定的过程。人脸检测作为人脸识别的关键步骤，对特征点的定位以及特征提取的合适与否有着关键性的影响，而人脸识别的准确率与提取到的特征有着密切的关系，所以对于人脸识别的研究和人脸检测的研究几乎是密不可分的。

深度学习<sup>[2]</sup>可以从样本数据中自动地提取到最有效和最本质的特征，但是需要大量的训练样本对神经网络结构进行优化和训练，并且一般来说，神经网络提取出来的特征往往具有一定的语义，非常适合用来对图像进行分类和识别。深度学习算法对复杂客观的事物具有较强的表达能力，这是深度学习算法的最大优点之一。因此使用深度学习的方法进行人脸的检测与识别是非常有意义的。

综上所述，研究人脸的检测与识别以及与此相关的技术对推进身份鉴定、智能视频的监控和人机交互等的发展具有重要的意义，而且 also 具有重要的科学研究价值和广阔的应用前景。

## 1.2 国内外研究现状

### 1.2.1 人脸检测的研究现状

人脸检测<sup>[3]</sup>问题最初来源于对人脸识别的研究，只有先从复杂的图像中检测到人脸的位置，才可以再进行人脸识别。人脸检测既是人脸识别技术中的第一步，也是人脸识别技术中关键性的一步。该步骤主要完成的工作是在一幅图像或视频的某一帧中寻找人脸的大致区域进而再确定人脸的精确位置和大小。拍摄的照片或视频都包含大量的背景信息，但是对于人脸识别来说，这些背景信息是毫无意义的，所以需要对面脸图像进行处理，进而找到背景信息尽可能少的人脸图像，再对图像进行检测以及定位：

$$F_j = S(f) \quad (1-1)$$

其中  $F_j$  表示对经过定位对齐的人脸图像进行提取， $f$  是被提取的原图像，包含大量的背景信息， $S(\cdot)$  表示对人脸图像进行提取，并对人脸图像实现尺寸归一化的操作。



早在二十世纪六十年代,就有学者对人脸检测的问题进行了研究,早期的研究主要是基于简单的启发式和人体测量技术的方法,这些方法假设人脸图像背景单一或者没有背景等条件,具有很大的局限性,当时只应用在证件照上;到了 20 世纪 90 年代,随着新的人脸识别系统的提出,人脸检测算法得到了迅速的发展,具有更强鲁棒性的算法被提出来,例如,基于模板匹配的人脸检测、基于子空间的人脸检测、基于变形模板匹配的人脸检测等算法;基于数据驱动的学习方法是近期人脸检测的主要研究方向,例如,基于卷积神经网络的学习方法、基于支持向量机(Support Vector Machine, SVM)<sup>[4-9]</sup>的方法和基于统计模型的方法等。

目前,国内外计算机视觉领域的研究人员已经对人脸检测进行了大量的深入研究。国外的起步研究相对较早,比较著名的研究机构有 CMU(Carnegie Mellon University, 卡内基梅隆大学)的机器人研究所、ISU(Illinois State University, 伊利诺伊州立大学)的贝克曼研究所和 MIT(Massachusetts Institute of Technology, 麻省理工大学)的多媒体和人工智能实验室等;国内比较著名的研究机构有中科院计算所和自动化所、清华大学、南京理工大学等。随着深入研究人脸检测问题,相关论文在国际上的发表数量也出现了大幅的增长,每年都有很多关于人脸检测的论文出现在 FG、CVPR 和 ICCV 等国际会议上。

可见,现阶段的人脸检测已经取得了很多研究成果,但仍然有很大的改进空间。对于光照的强度、拍摄的角度、拍摄时的表情、年龄的增长、是否化妆、是否存在遮挡和图像质量的变化等人脸图像,如何有效地提高人脸检测的准确率和速率有待于进一步研究。

## 1.2.2 人脸识别的研究现状

通过计算机提取到合适的人脸特征,并根据提取到的合适特征进行身份识别的技术叫做人脸识别。一般地,自动人脸识别的流程如图 1.1 所示。

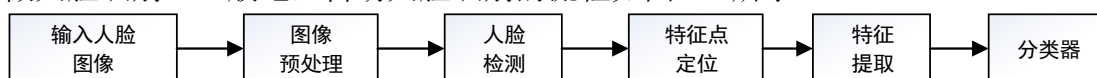


图1.1 自动人脸识别流程图

其中,特征点定位又称人脸对齐,一般来说,狭义的人脸识别只包括自动人脸识别流程图的最后两个部分,也就是特征提取和分类器的算法研究。人脸识别的研究历史可以分为以下四个阶段:

1. 20 世界 60 年代,是人脸识别研究的萌芽阶段,人类才开始研究人脸识别的问题<sup>[10]</sup>。在这个阶段,对现实场景中的人脸图像识别效果不好,主要因为人脸识别方法有两大特点:

- 1) 人脸识别被当作模式识别方面的问题,主要是基于人脸的几何结构特征,提

取到的特征是人脸的五官特征信息和各个器官之间的几何关系。这种识别方法比较简单，很容易丢失关于人脸的关键信息，因而当视角，表情，姿态等发生变化时，对人脸的识别能力较差。

2) 主要针对在较强约束条件下的人脸图像的识别，识别的是背景单一或者没有背景的人脸图像，由于背景单一或者没有背景，那么人脸的位置就相对容易获得或者人脸的位置直接是已知的。

2.到了 20 世纪 90 年代，计算机软件 and 硬件性能幅度的提高，以及对人脸图像识别能力更高的要求，人脸识别算法得到了快速的发展，鲁棒性更强的人脸识别方法出现了，例如，基于弹性图匹配的识别方法和基于特征脸的识别方法，此类识别方法是通过对人脸表观建模的整体识别方法。

3.在 2000 年后，随着不断深入研究人脸识别问题，面向真实环境下的人脸识别吸引了众多研究者的目光，主要研究了以下四个方面：1)不同人脸空间模型的设计，例如，基于以线性判别分析为代表的线性建模方法的人脸识别，基于以 Kernel 方法为代表的非线性建模方法的人脸识别和基于人脸 3D 信息的 3D 识别方法。2)深入分析和研究影响人脸识别的因素，包括基于不变姿态条件下的人脸识别方法、基于不变表情条件下的人脸识别、和基于不变光照条件下的人脸识别方法等。3)引入了新的特征表示方法，例如深度学习方法和 Gabor Face 因子方法和 LBP Face<sup>[11-13]</sup>因子方法等。4)例如素描的人脸识别方法、视频的人脸识别方法和近红外图像的人脸识别方法。

4.自 2014 年以后，主要通过深度学习的方法进行人脸识别。

Hinton<sup>[14-16]</sup>等人在 2006 年提出了深度学习的相关理论知识，主要通过模拟人脑分析学习的方式建立网络模型。相比于传统的人脸识别算法，深度学习以无监督学习方式和多层网络结构能更好的抽象数据，更适合目标的检测与分类问题。在各式各样的传统人脸识别算法中，以主成分分析 (Principal components analysis, PCA) 算法<sup>[17]</sup>最为突出。但是在光照姿态变化等条件下人脸识别的效率大幅度降低，这也是所有人脸识别算法受到限制的最大原因。现在很多研究机构科研人员将深度学习方法引入人脸识别中，并有了很好表现。

国外对基于深度学习的人脸识别<sup>[18-21]</sup>研究相对较早，勒尼德·米勒研究小组早在 2012 年就将深度学习应用于 LFW 人脸数据库，并且取得了 87% 的识别率。另外，英国谢菲尔德大学的 Neil Lawrence 研究组和加拿大蒙特利尔大学也对基于深度学习的人脸识别进行了深入研究。相比于国外，国内的对于基于深度学习的人脸识别算法研究较晚，但也取得了大量的研究成果。

在国家自然基金的资助下，清华大学的计算机研究所、电子研究所和自动化研究所，哈尔滨工业大学的计算机研究所，中科院的计算机研究所，南京理工大学的计算

机研究所,上海交通大学的研究所等都对人脸识别进行了深入的研究,其中,香港中文大学汤晓鸥教授团队使用的卷积神经网络,DeepID<sup>[22-24]</sup>经过三次改进,在人脸识别的准确率已经超过了 99%。国内比较著名的研究机构还有清华大学的苏光大教授,随着学术界对人脸识别的深入,工业界对人脸识别也取得了很大的成功。云从科技,由计算机视觉之父,中科院研究所创立团队所创立,提出了首个刷脸支付系统原型,也是唯一一家参与人脸识别国际、部级、行标制定的研发企业。

中科院的自动所的李子青教授,李子青教授的中科奥森公司首先提出了基于近红外线的人脸识别技术,并将该项技术成功应用于在北京举办的奥运会。清华大学丁小青带领的研究团队,是唯一一个完成大规模 3D 人脸识别性能测试的参赛团队,他们的算法在 3D 领域排名第一。浙江大学何晓飞带领的课题组等。随着对人脸识别问题的深入研究,每年都有大量的关于人脸识别的论文被发表在 FG、CVPR 和 ICCV 等国际会议上。

深度学习通过大量的数据进行训练,从而可以自己从大量数据中学习到对于光照、尺度、表情角度等的不变性,目前,基于深度学习的人脸识别算法在 LFW 数据库上已经达到了 99.63 的准确率,超过了人眼对物体的识别结果。

### 1.3 论文主要研究工作

本文首先阐述了人脸检测与识别的研究背景和意义,对人脸检测与识别技术的研究现状进行了描述与分析,重点研究了基于卷积神经网络的人脸检测与识别算法,对卷积神经网络不同层提取的特征进行可视化分析,更深一步了解高层特征是低层特征的抽象;其次对 Faster R-CNN 算法进行深入研究,改进了 Faster R-CNN 算法中的 RPN 网络模型,用固定尺寸分割策略代替滑动窗口的策略,将空间金字塔池化技术引入卷积神经网络模型 VGG,构建了一个新的 VGG 模型,最后通过对该模型的监督优化训练,使得该模型达到了较高的检测率和识别率同时,也对人脸的检测与识别具有较好的鲁棒性。在本文的最后,基于改进的网络模型,基于 Caffe 开源平台实现了一个人脸检测与识别软件,并对该软件进行了测试,实验测试结果表明,该系统达到了较高的检测率与识别率。

### 1.4 本文章节安排

本论文总共分为五章,各章内容安排如下:

第一章:绪论。

本章首先对人脸检测与识别的背景和意义进行了阐述,然后分别介绍了目前人脸检测与人脸识别的发展历史以及国内外研究现状,最后阐明了本文的主要研究工作和

章节安排。

第二章：人脸检测与识别相关理论。

本章首先介绍了人脸检测与人脸识别的基础理论知识,接着阐述了 R-CNN 算法、卷积神经网络的基础知识以及 VGG 网络模型,对本文涉及到的基础理论知识进行了详细的说明。

第三章：基于 Faster R-CNN 算法的人脸检测与人脸识别。

本章在 Faster R-CNN 算法的基础上,改进了 RPN 卷积神经网络网络,提升了人脸检测的准确率,将 SPP 引入 VGG 模型,对人脸识别的准确率有了一定的提升。最后介绍了如何交替训练 RPN 和 Fast R-CNN,用到的数据集和数据集的预处理工作等。

第四章：基于 Faster R-CNN 开源平台的人脸检测与识别软件设计。

主要是根据第三章提出的改进 RPN 网络和改进的 VGG 网络,来进行软件系统的设计和集成开发,最后介绍了系统集成所需要的环境,如何实现该软件和详细的操作流程等。

第五章：总结与展望。

本章节简要概括了本文的研究工作和贡献,并在当前工作的基础上,对人脸检测与识别的下一步工作提出了相应的参考方向和建议。

## 第二章 相关理论基础

### 2.1 人脸检测与人脸识别理论

人脸检测就是对图像或视频中是否存在人脸进行判断,在存在人脸的情况下返回人脸的位置、大小等信息。一般的,可以将人脸检测算法分为两种<sup>[25]</sup>: 1) 基于显式特征的人脸检测。主要是根据人脸的外观信息,例如,轮廓信息、几何特征和肤色特征等; 2) 基于隐式特征的人脸检测。该方法需要对大量的正负人脸图像样本进行训练,构造分类器,进而对图像中的人脸区域进行判断。相比较而言,第一种方法检测速度快,但是检测效果不好;第二种方法需要提前训练好分类器,所以检测效果会提升,而且适用范围比较大。

#### 2.1.1 基于显式特征的人脸检测方法

显式特征是指对人类来说可以直接观察到的特征,例如,人脸的肤色、脸部轮廓信息和脸部结构等。基于显式特征的人脸检测方法主要方法有:(1) 基于特征的人脸检测方法;(2) 基于模板的人脸检测方法<sup>[26-27]</sup>。

##### 1. 基于特征的人脸检测方法

人脸具有鲜明的几何特征,例如:人的两个眼睛在人脸的上半部分对称分布,两个眼睛之间的连线和鼻子嘴唇的中间连线是大体垂直的,人的鼻子大致在嘴唇的中间,也在眼睛的中间等鲜明的特征,通过利用这些特征,可以大致地对人脸局部特征之间的相对距离和位置关系进行确定。研究人员通过长期的生活观察可以获得此类特征,没有复杂的学习过程,简单快速,但是由于描述人脸特征关系的规则不容易设计,设计的过高或过低,会造成人脸的拒绝识别或人脸的错误识别,进而导致人脸检测的准确率过低,并且很难适应自然场景下的不同姿态的人脸检测,因此,此方法的鲁棒性和稳定性都比较差。

##### 2. 基于模板的方法

此方法可以有效地对图像或视频中的人脸进行检测,是一种经典的模式识别方法。首先需要建立一个标准的人脸模板,在进行检测时,将待检测的人脸图像模板和标准的人脸模板进行比较,进而求得两者的相似程度。

#### 2.1.2 基于隐式特征的人脸检测方法

基于隐式特征的方法,需要提前使用大量的负类样本,正类样本进行训练,构造分类器,再用得到的特征对人脸图像进行分类识别。基于隐式特征的主要方法有:(1)

基于 Adaboost 的人脸检测方法；(2) 基于 PCA 的人脸检测方法；(3) 基于可变形组件模型 (Deformable parts models, DPM)<sup>[28]</sup> 的人脸检测方法；(4) 基于人工神经网络 (Artificial Neural Network, ANN)<sup>[29]</sup> 的人脸检测方法。

### 1) 基于 Adaboost 的人脸检测方法

Adaboost 算法主要是通过反复训练样本数据集中的正负样本,不断地调整权重值,进而得到多个不同的弱分类器,接着将得到的多个弱分类器进行加权叠加,最后得到一个强分类器,用来进行人脸检测。理论研究证明:只要弱分类器检测的准确率高于 50%,通过对弱分类器进行级联组成强分类器,当级联的个数趋向于无穷多个时,强分类器的检测准确率将趋于 1。但是,复杂环境会对 Adaboost 人脸检测算法产生较大的影响,从而导致人脸检测结果不稳定,误检率较高。

### 2) 基于 PCA (Principal Component Analysis, 主成分分析) 的人脸检测方法

PCA 是模式识别中的有效方法之一,通过特定的方法,将样本线性映射到高维空间,使样本的类内散布程度达到最小的同时,而类间散布程度达到最大,进而提取图像数据的主要特征。

一般来说,PCA 可以将以下两个参数构成检测特征向量:1) 人脸基准点的相对比率,2) 其他描述人脸部特征的轮廓参数或类别参数等,通过这种方式,使基于人脸整体的检测不仅对人脸不同部件之间的相对位置信息进行了保留,也对各部件本身的信息进行了保留。

### 3) 基于可变形组件模型的人脸检测方法

DPM 是传统检测的主流方法,是一种基于组件的检测方法,通过使用一组混合的 DPM 获取人脸在不同姿势、表情下的参数,进而得到人脸位置和关键点信息。此方法可以有效地使用较少的数据集进行训练,并且因为没有模型的变形,具有较好的泛化性能,对扭曲、多姿态和多角度的人脸有较好的检测效果。但是该模型过于复杂,判断时计算复杂,很难满足实时性的要求。

### 4) 基于人工神经网络的人脸检测方法

通过对生物神经网络的结构和功能进行模仿,构建的神经网络模型叫做人工神经网络。此方法一般需要大量的数据样本进行训练学习,在训练的过程中,对各个层的权重进行校正而创建模型的过程,经常通过反向传播算法对网络的权重值进行验证,最终对模型进行优化,达到较高的检测率。

## 2.1.3 人脸识别理论

人脸识别的本质就是通过计算机分析人脸的面部图像,然后采用不同的方法来提取人脸中最有效、最本质的特征,是一种可以用来鉴定身份的识别技术。人脸识别主

要包括人脸图像的预处理、人脸图像中人脸的检测、人脸图像特征的提取和人脸图像的分类识别这四个步骤。也可以主要分为人脸检测和人脸识别两个过程，其中，人脸检测是对输入的人脸图像，通过一定的方式，判断出人脸所在的位置、尺寸大小等信息，为后面的人脸识别提供良好的材料，人脸识别是将现实生活中的人脸图像映射到计算机的空间，然后通过某些算法，尽可能完整而准确地描述出人脸，接着再将待识别的人脸与已知的人脸进行对比，通过比较两者之间的相似度，进行人脸的身份判断。

人脸识别是对人脸图像进行检测之后，再对人脸图像中的人脸进行特征提取，对比提取到的人脸特征和人脸数据库中的人脸图像的特征，对待测人脸的身份进行确认和识别的过程。人脸的识别又成为人脸的分类，人脸检测与人脸识别的区别：人脸的识别不需要对人脸图像中的人脸进行定位，而人脸检测需要定位，也就是需要把人脸的 bbox (bounding box) 标记出来，并且人脸检测需要将人脸图像中的所有人脸都标记出来。

人脸识别的主要研究方法有：1) 基于 PCA (Principal Component Analysis, 主成分分析) 特征的方法；2) 基于几何特征的方法；3) 基于弹性图匹配 (Elastic graph matching) 的方法；4) 基于深度学习的方法。

#### 1) PCA 特征人脸识别

PCA 是一种无监督的学习方法，指向数据能量分布最大的轴线方向就是主分量，因此，对于数据的最优的表达，需要进行最小均方误差。通过主分量分析得到的特征，对于分类任务来说，并不能保证可以区分开各个类别。PCA 人脸识别方法的流程主要如下：

1. 将人脸图像视为  $R^N$  的空间向量  $X$

2. 将人脸图像集合视为线性子空间，即

$$x \rightarrow ((x - \bar{x}) * v_1, (x - \bar{x}) * v_2, \dots, (x - \bar{x}) * v_n)$$

3. 选取若干主方向  $v_i$  作为子空间的基

$$x \rightarrow (\underbrace{(x - \bar{x}) * v_1}_{A_1}, \underbrace{(x - \bar{x}) * v_2}_{A_2}, \dots, \underbrace{(x - \bar{x}) * v_n}_{A_n})$$

4. 人脸图像在主方向上投影系数  $A_i$  为特征

#### 2) 基于几何特征的人脸识别方法

一般来说，人的眼睛、鼻子、嘴巴、下巴等五个部件构成了人脸，世界上的每个人脸互不相同就是因为这五个部件的形状、大小和结构上的各种差异，所以人脸识别的特征可以通过对于这五个部件的几何形状和结构关系的描述。最早用于人脸侧面轮廓的描述与识别的特征就是几何特征，首先需要对若干显著点进行确定，通过侧面轮廓曲线的方式进行确定，接着一组用于识别的特征度量如距离、角度等通过这些显著点导出。一般通过对人的眼睛、嘴巴、鼻子等重要特征点的位置和重要器官的几何

形状进行特征提取，提取的特征将作为采用几何特征进行正面人脸识别分类的特征。

### 3) 基于弹性图匹配的人脸识别

基于弹性图匹配的人脸识别方法是由 Lades<sup>[30,31]</sup>等人提出来的。在该算法中，人脸图像通过属性图的方式来进行表征。由顶点和边组成了属性图，其中，特定图像特征点处不同尺寸的 Gabor 特征用顶点的属性进行表示；各节点之间的连接关系通过边的属性表示。即使人脸图像发生了一定程度上的局部变形，弹性图匹配算法对该人脸图像识别的准确率也比较好，但是由于该算法的计算量比较大，造成该算法的识别效率不高。

### 4) 基于深度学习的人脸识别

与传统的人脸识别算法相比，深度学习方法以无监督学习方式和多层网络结构，能更好地抽象数据，更适合目标分类问题。在各式各样的人脸识别算法中，以 PCA(主成分分析) 算法最为突出。但是在光照姿态变化等条件下人脸识别的效率大幅度降低，这也是所有人脸识别算法受到限制的最大原因。现在很多研究机构科研人员将深度学习方法引入人脸识别中，并有了很好表现。

近些年来，得益于更强大的计算机、更大的数据集和能够训练更深网络的技术，深度学习的普及性和实用性都有了极大的发展。深度学习的实质就是学习到多层非线性的函数关系，这种多层的非线性函数关系使得人们能更好地对视觉信息进行建模，从而更好地理解图像和视频，更好地处理视频人脸目标和目标识别这类复杂的问题。深度学习的优点在于其能够逐层地学习原始数据的多种表达。每一层都以底一层的表达为基础，但往往更抽象，更加适合复杂的分类等任务。

对于人脸图像来说，在最底层的特征基本是类似的，就是各种边缘的信息，越往上，越能提取出人脸的一些特征，例如眼睛，鼻子，嘴巴等等，到最上层，不同的高级特征最终组合成相应的人脸图像。

在深度学习的框架下，深度学习算法可以直接从原始图像学习具有判别性的人脸特征。并且在海量人脸数据的基础上，基于深度学习的人脸识别在速度和精度方面已经远远超过人类。

## 2.2 R-CNN 算法介绍

传统的人脸检测算法主要依赖研究者手工设计特征，随着卷积神经网络的提出，人们发现可以利用卷积神经网络自动提取特征，并且，提取到的特征不仅包含了人脸的类别信息，还包含着人脸的位置信息，而且提取到的特征具有位移、尺度、平移和形变等不变性，R-CNN(Regions with Convolutional Neural Network Feature)<sup>[32-34]</sup>算法是卷积神经网络在物体检测的开山之作，主要分为以下四个步骤：



1. 生成候选区域 (region proposal): 对于输入的每一张图像, 使用选择性搜索 (Selective Search, SS) 方法生成 1K-2K 个候选区域。

其中, SS 算法的主要思想如下:

[1] 使用一种过分割的方法, 将输入的图像分割为小区域, 约 1K-2K 个。

[2] 根据分割而成的小区域, 按照一定的合并规则 (根据人脸图像的纹理, 可以找到人脸的位置, 最后根据人脸图像的颜色, 找到不同人物的人脸), 对合并可能性最高的相邻两个区域进行合并, 不断重复此过程, 直到整张图像合并成一个整体区域。

[3] 输出所有合并出的区域, 也就是候选区域。

2. 特征提取 (feature extraction): 对于生成的候选区域, 使用深度卷积神经网络提取特征, 也就是神经网络全连接层输出的 4096 维特征向量。在进行特征提取前, 需要对输入的候选框进行尺寸归一化, 归化成统一尺寸大小为  $227 \times 227$ 。

3. 类别判断 (classification): 使用线性 SVM (Support Vector Machine) 分类器对提取到的特征进行判别, 判别是否属于该类别。

采用卷积神经网络训练时, 需要对 bounding box (具有包围盒) 的目标进行识别分类训练, 最后一层是 Softmax 分类器, 但是在原文中采用 SVM 分类器进行分类识别, 具体原因如下:

SVM 训练和 CNN 训练对正负样本的定义方式是不同的, 在 CNN 进行训练时, 需要对训练数据进行阈值较低的标注, 即 IoU (重叠度) 大于 0.5 就可以标注为正样本, 因为 CNN 在训练过程中容易过拟合, 需要大量的训练数据, 如果阈值标注过高, 就会导致 CNN 训练样本数很少。而 SVM 训练只需要少量的样本, 所以需要 IoU=1 才可以标注为正样本。

4. 位置精修 (rect refine): 对于每一个分类, 使用一个回归器 (regressor) 根据人脸检测算法的评价标准——IoU (重叠度) 对候选框的位置进行精修, 生成预测窗口的坐标。

R-CNN 算法的具体流程如图 2.1 所示。

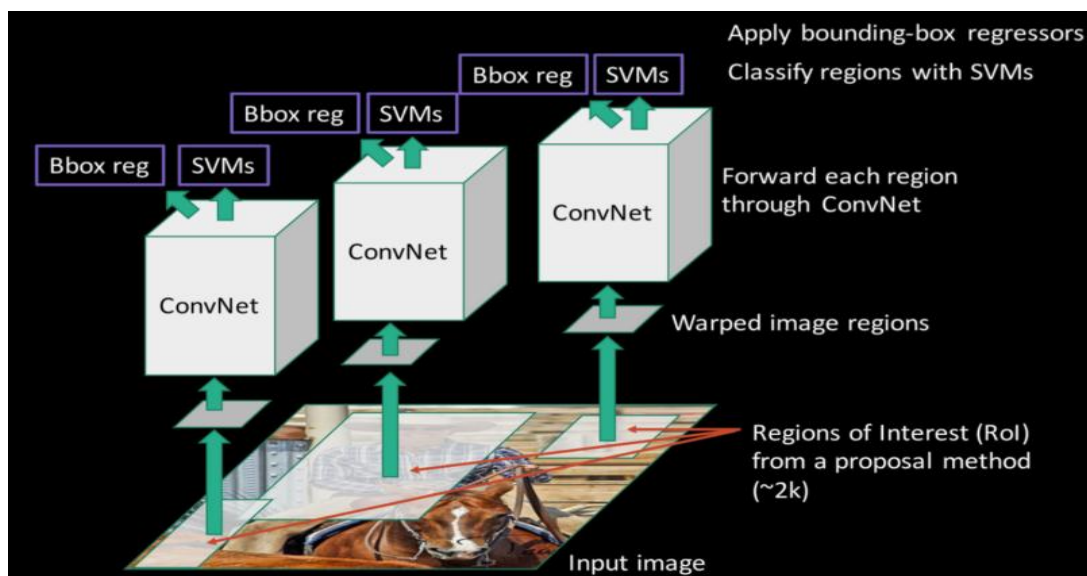


图2.1 R-CNN 算法流程示意图

随着 R-CNN 的提出和使用，发现 R-CNN 具有以下缺点：

1. 训练和测试速度慢。对于输入的每张图像，生成的候选框之间存在很大的重叠，重叠部分会被多次重复提取特征，从而第二步对特征的提取也就存在大量的冗余操作。
2. 训练占用内存大。对于每一类分类器和回归器，都需要大量的特征作为训练样本。

针对 R-CNN 的以上缺点，提出了 Fast R-CNN 算法，对于测试速度慢，采用将整张图像输入网络之后，在最后卷积层输出的特征向量上进行提取候选区域的操作，使之前的 CNN 运算可以共享，避免了候选区域前几层特征的重复计算。对于训练速度慢的问题，选择将整张图像归一化（同一尺寸大小为  $227 \times 227$ ）后直接输入网络，在邻接时，加入候选框的信息。针对训练占用内存大的问题，选择使用神经网络进行分类和位置精修，不再需要额外的存储空间。Fast R-CNN 的算法流程图如图 2.2 所示。

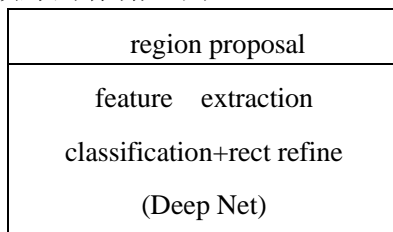


图2.2 Fast R-CNN 算法流程图

而对于 Faster R-CNN，可以简单理解为 RPN(Region Proposal Network)和 Fast R-CNN 的结合，通过 RPN 代替了 SS 提取候选区域，加快了运行速度。

在原文中，RPN 网络是一个全卷积网络（Fully Convolutional Network, FCN），

主要是用于生成候选区域，也就是估计目标物体的位置和大小。传统的图像分割方法是以像素点为基础的，例如，SS 通过聚类的方式在低维像素特征上进行分割，速度很慢，花费的时间很长；边缘检测算法和 SS 算法的速度相比，得到了很大程度上的提升，但是和卷积特征提取和分类的速率相比，还是太过缓慢。而 RPN 利用全卷积网络的特性，将分割操作放在了卷积网络提取特征之后，因为提取的高维特征是高度抽象的，需要搜索的范围会变小，那么搜索花费的时间也会变少，效率也会相应的提高。RPN 会将置信率高的目标输出到下一层网络。

在 CNN 网络模型的最后一层卷积特征图上构建 RPN 网络，RPN 结构如图 2.3 所示。

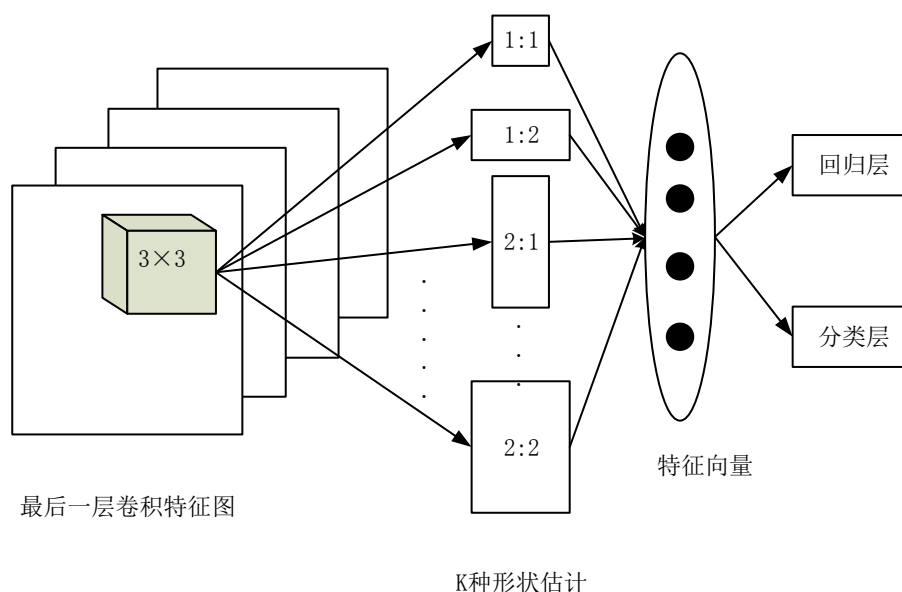


图2.3 RPN 结构示意图

如图 2.3 所示，RPN 在共享网络的最后一层卷积特征图中，找到一个点，通过感受野，映射到原始图像，其次，回归层和分类层是两个独立平行的全连接层，回归层的主要是对原始目标的位置进行估计操作，分类层主要是对目标进行分类，得到目标属于某个类别的概率。对于原图像的每一个位置，形成了 9 个不同大小的候选窗口：三种面积  $\{128^2, 256^2, 512^2\}$  和三种比例  $\{1:1, 1:2, 2:1\}$ ，这 9 个候选框又称为 anchors，是使用  $3 \times 3$  的滑动窗口在卷积网络最后提取的特征上滑动扫描得到的。

而我们知道，如果对完成两种不同任务的网络模型分别进行训练，即使它们的网络结构、参数和数据集等完全一致，但是各自的网络层的卷积核也会向着不同的方向改变，最后的网络权重是无法共享的。对于 RPN 和 Fast R-CNN 来说，都需要一个原始特征提取网络，如何训练这个网络使得提取的特征满足 RPN 和 VGG 的不同需求，一共有三种不同的模型交替训练策略：

### 1. 轮流训练

- 1) 首先通过使用 ImageNet 数据集训练得到初始权重值  $W_0$ ;
- 2) 用  $W_0$  初始化 RPN 后再进行训练;
- 3) 用  $W_0$  初始化 VGG 网络后再进行训练, 训练参数标记为  $W_1$ ;
- 4) 接着再用  $W_1$  初始化 RPN 后再进行训练;
- 5) 最后用  $W_1$  初始化 VGG 网络, 用 RPN 输出的候选区域对训练 VGG 网络。

## 2. 近似联合训练

与轮流训练方法不同, 不再是分别串行训练 RPN 网络和 VGG 网络, 而是把两者合并到一个网络里面, 网络结构如图 2.4 所示。

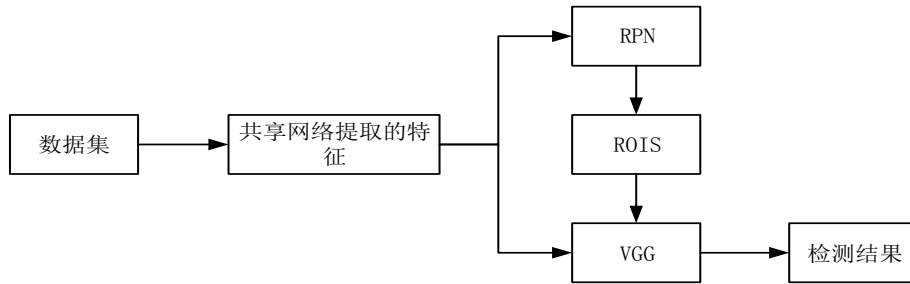


图2.4 合并的网络结构示意图

从图中可以看到, RPN 会输出候选区域, 而不需要从网络外部获得。RPN 使用反向传播算法 (BP<sup>[35]</sup>传播算法) 计算梯度, 将提取到的 RIO 区域作为一个固定值, 在更新权值时, 需要考虑将 RPN 和 Fast R-CNN 的增量合并输入特征提取层。

计算梯度的 BP 算法主要包括两个部分: 正向传递的信息与反向传播的误差。在正向传播的过程中, 信息经过隐含层的逐层计算, 从输入层传向输出层, 只有相邻层的神经元状态会互相影响, 前一层的状态只会对后一层神经元的状态产生影响。如果输出层的输出不是期望的值, 则需要计算输出层的误差变化值, 然后从输出层传向输入层, 沿着网络结构, 将误差信号沿原来的连接通路反传回输入层, 通过对各层神经元权值的修改来达到期望目标。

正向传播的主要执行过程:

在对网络训练之前, 我们需要对该网络权重进行随机初始化和偏置, 权重是  $[-1, 1]$  的一个随机实数, 偏置是  $[0, 1]$  的一个随机实数, 在完成权重和偏置的初始化之后, 开始进行前向传输。通过多次迭代可以完成神经网络的训练, 使用训练集的所有数据进行每一次迭代, 只有一条数据用于每一次的网络训练, 首先设置输入层的输出值, 假设属性的个数为 100, 那我们就设置输入层的神经单元个数为 100, 输入层的结点  $N_i$  为记录第  $i$  维上的属性值  $X_i$ 。对输入层的操作比较简单, 之后的其他层就要略微复杂一些了, 除输入层外, 其他各层的输入值是上一层输入值按权重累加的结果值加上偏置, 每个结点的输出值等该结点的输入值作变换。如图 2.5 所示。

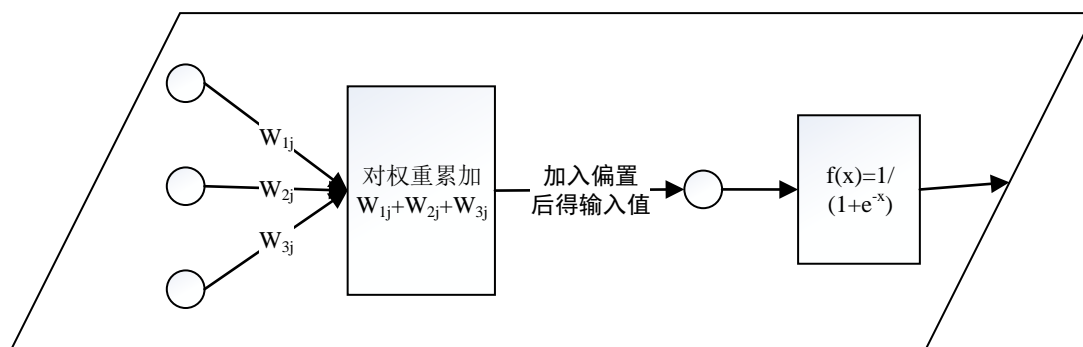


图2.5 前向传输图

反向传播的主要执行过程：

从最后一层即输出层开始进行逆向反馈，最后一层的输出可以描述该数据样本的类别是我们训练神经网络作进行分类的最终目的，例如对于一个二分类的问题，输出层会使用两个神经单元，通过对两个神经单元进行比较进行分类，若第一个神经单元数值比较大，那么，这个样本数据应该属于第一类，否则属于第二类。

反向传播算法是基于梯度下降的算法，对全局误差的计算，是为了神经元的权重进行调整，且需要将权重误差向误差减少的方向进行调整。

反向算法的执行步骤如下：

- 1) 对权值进行初始化；
- 2) 输入一个样本，对其相应的期望输出结果；
- 3) 计算各层的输出，求出每一层对应的某一个神经元的输出；
- 4) 求出各层输出单元的学习误差
- 5) 再反向计算隐藏层的误差。
- 6) 更新每个神经单元的权重和阈值

7) 当得到各层的权重之后，可以根据指定的判定标准判断求得的权重是否满足要求。如果满足，则算法结束，如果不满足，则前往步骤（3）进行执行，直到满足要求为止。

### 3. 联合训练

该训练流程与近似联合训练相似，但是需要考虑 ROI 区域的变化对 RPN 和 VGG 网络权重的影响。

如果对 RPN 网络和 Fast R-CNN 网络单独训练，会导致不同的卷积层参数，故需要采用一种模型交替训练的策略，从而达到两者共享卷积层参数的目的。根据以上三种联合训练的方法，Faster R-CNN 算法采用了轮流训练策略，使生成候选区域的网络 RPN 和 Fast R-CNN 实现了共享，Faster R-CNN 算法流程图如图 2.6 所示，主要解决了三个问题：

1. RPN 的设计;
2. RPN 的训练;
3. RPN 和 Fast R-CNN 网络共享特征提取网络。

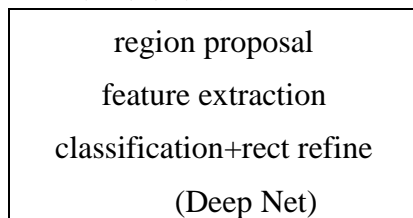


图2.6 Faster R-CNN 算法流程图

## 2.3 VGG 网络模型

VGG 网络模型是由牛津大学视觉几何组开发的卷积神经网络结构，是经典的 CNN 模型之一，因为本文使用 VGG 模型对人脸图像进行识别，所以接下来首先对卷积神经网络进行介绍。

### 2.3.1 卷积神经网络结构

CNN<sup>[36-39]</sup>是深度学习在视觉领域中最具有代表性的网络之一，主要用于研究语音分析和图像识别领域。它的网络结构具有权值共享的特性，使之类似于生物的神经网络，CNN 可以自动提取图像的特征，进而避免了传统识别算法中复杂的特征提取。CNN 提取的高层特征是低层特征的组合，也是低层特征的进一步抽象，抽象的层次越高，存在的可能性猜测就会越少，进而更有利于图像的分类识别。

以人脸学习为例，卷积神经网络可以通过模拟人脑认识的过程，针对人脸图像的分层特征表达进行；最底层从图像的原始像素开始学习滤波，刻画局部的边缘和纹理特征；中层滤波器通过将各种边缘滤波器进行组合，描述不同类型的人脸器官；最高层描述的是整个人脸的全局特征。

卷积神经网络来源于普通的神经元网络，多层感知器（MLP）<sup>[40]</sup>。是一种深度监督学习下的网络学习模型，具有很强的适应能力，可以挖掘数据的局部特征，进而提取到全局训练特征和进行分类，它的权值共享网络结构使之类似于生物神经网络，在模式识别的各个领域都得到了成功应用。CNN 擅长处理图像特别是大图像的相关机器学习问题的多层神经网络。其中，卷积是两个函数之间的相互关系，是在连续空间做积分计算和离散空间内求和的过程。在计算机视觉领域，卷积是一个抽象的过程，可以把小区域内的信息统计抽象出来。

CNN 通过结合人脸图像的局部感知区域、权值共享、在空间或者时间上的降采样等特性，优化网络模型结构，保证了一定程度上的位移、尺度、平移和形变等不变

性。

### 1. 局部感受野

人的眼睛对事物的观察总是聚焦在一个相对较小的局部，并且在人脸图像空间，也是局部区域内的像素联系比较紧密，反之，联系则较弱。一般的，对于多层感知机来说，需要将隐层节点与图像的每一个像素点全部连接，在 CNN 中，每个隐层节点只需要连接到图像某个足够小的局部像素点上，获得局部特征，然后在更高一层通过将局部特征抽象起来得到全局特征，从而在很大程度上减少了需要训练的权值参数。

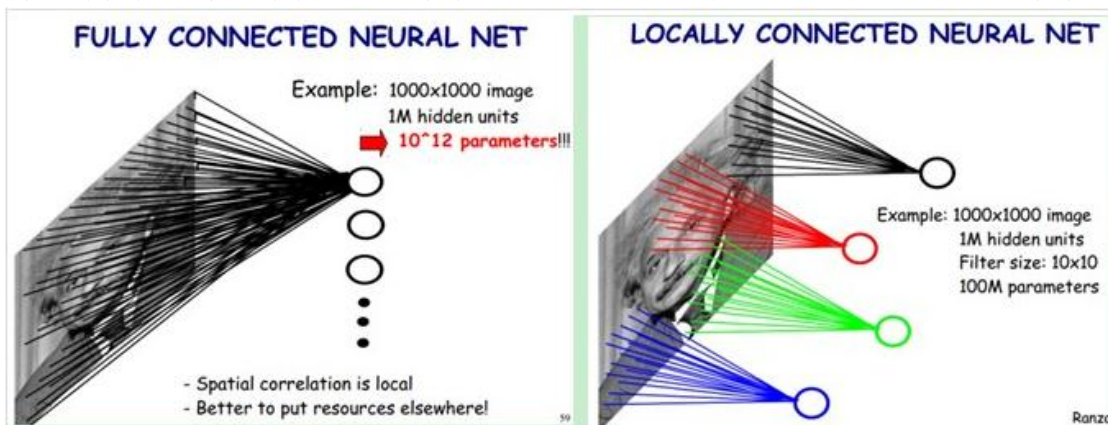


图2.7 全连接和局部连接图

如图 2.7 所示，左图是全连接图，假设该图像的像素是 $1000 \times 1000$ ，神经元数目是 $10^6$ 个的隐藏层。如果使用左图的全连接方式，则参数数量会达到 $1000 \times 1000 \times 10^6$ 个，这样需要训练的参数将会有 $10^{12}$ 个，面对如此众多的参数，将会导致该网络模型无法训练。如果使用图 2.1 右图中的局部连接方式，大小为 $10 \times 10$ 的局部感受野，那么需要训练的参数数量为 $10 \times 10 \times 10^6$ 个，将训练的参数减少了 4 个数量级。

### 2. 权值共享

在人的整个神经系统中，某个神经中枢的神经细胞的结构、功能应该是相同的，甚至是可以相互替代的。在 CNN 中，对于位于同一个卷积核内的所有神经元来说，其权值应该是相同的，进而可以将需要训练的参数减少。



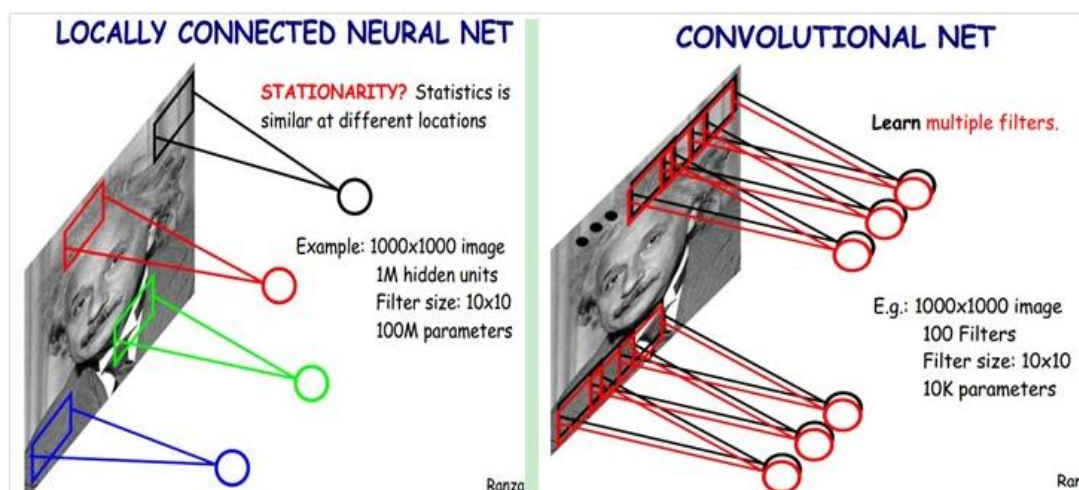


图2.8 局部连接图和卷积图

如图 2.8 所示，根据局部感受野的方式，将训练的参数减少了四个数量级，但是参数还是很多，可以通过权值共享进一步减少参数。如图 2.2 中右图所示，在局部连接中隐藏层的每一个神经元连接的都是  $10 \times 10$  的感受野，可以通过权值共享将  $10 \times 10$  的权值参数共享给其他的神经元，最后需要训练的参数个数  $10 \times 10$ 。因为不同的卷积核可以得到图像在不同映射下的特征，如果需要多提取特征，可以通过增加卷积核的方式。假设有 100 个卷积核，由于权值共享，也只需要训练  $10 \times 10 \times 100$  个参数。

### 3. 池化

人对看过的事物，回想起来一般不会记得细节。在 CNN 中，不需要对原图进行处理，这就是池化。每次对图像进行卷积操作后，通过下采样过程，减少需要训练的参数个数。

### 4. 网络结构

人的大脑可以处理各种复杂的数据，例如声音、图像和文本等，CNN 是一种通过模拟人脑建立的分析学习的神经网络，卷积神经网络可以通过抽象低层特征形成高层特征，具有更好的表示能力，所以，卷积神经网络可以用来做人脸的检测与识别。

由多层神经网络组成卷积神经网络，包括位于底层的输入层、卷积层、下采样层和全连接层，每一层神经网络有多个特征映射图，每个特征映射图有多个神经元，特征可以被每个特征映射图通过神经元进行提取，并且同一个特征映射图上的多个神经元之间的权值是共享的。大部分的卷积神经结构都是类似的，因此本文将详细介绍典型神经网络的各层具体操作。



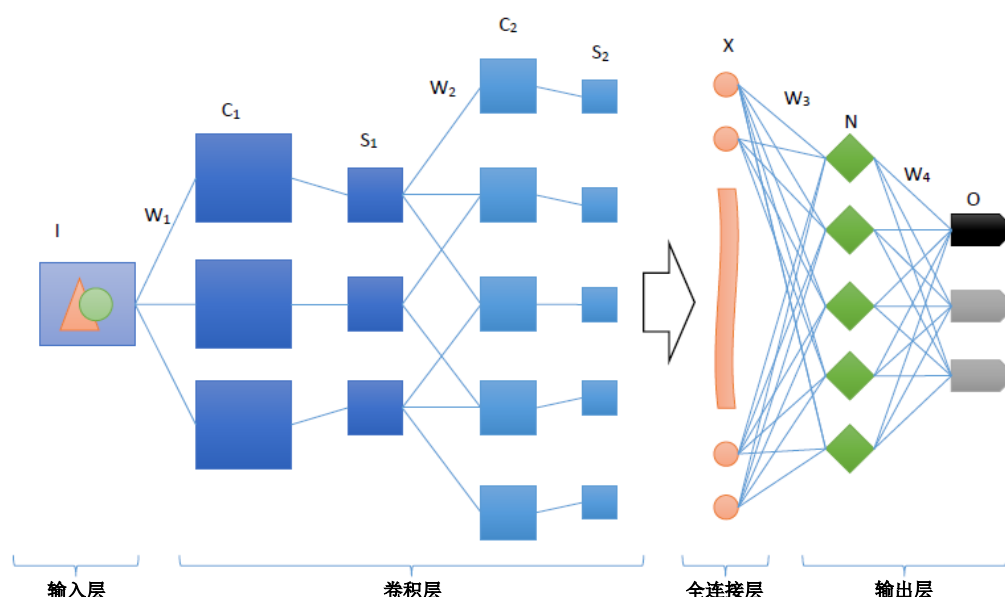


图2.9 经典神经网络结构图

如图 2.9 所示，最前面的是输入层，用于对输入图像数据进行处理。数据层位于网络的最底层，是模型的入口，不仅提供数据的输入，也提供数据从 Blobs 转换成别的格式进行保存输出。主要指的是对图像数据的处理，将数据转换为 Caffe 平台中识别的数据类型。

卷积层的卷积运算可以使原图像信号特征增强，并降低噪音。卷积层相当于对图像进行滤波，抽象出来图像的局部信息，局部信息是通过较小的卷积核在图像不同的局部位置上扫描得到的，在实际训练过程中，卷积核的值是在学习的过程中学到的。

池化可以根据图像局部相关性的原理，对图像进行子采样减少计算量，同时保持图像旋转的不变性。主要作用包括降低特征图的分辨率，从而减少计算量，以及增强网络的鲁棒性，降低数据维度。池化的方式一般有最大值池化和均值池化，而空间金字塔池化，使得任意大小的特征图都可以转换为固定大小的特征向量。如果希望金字塔的某一层输出  $n \times n$  个特征，则使用的 windows size 大小为  $(w/n, h/n)$  进行池化。

下采样层主要是为了减少过拟合的问题，减少不同参数之间的耦合性。从某种程度上来说，池化和下采样是被包含与包含的关系，下采样包含的范围更大，而池化仅仅是其中的一种方法而已。

而全连接层在一定程度上泛化了 Dropout，也是一种缓解拟合的技术。而 Dropout 只在全连接层使用，随机的将全连接层的某些神经元的输出置为 0。

最后采用 Softmax 全连接，得到的激活值就是卷积神经网络提取到的图像特征。

### 2.3.2 VGG 网络模型介绍

VGG<sup>[41]</sup>网络模型一共有 19 层，其中，有 13 个卷积层，5 个池化层，3 个全连接

层，分别是两个图像特征层和一个分类特征层。因为较小的卷积核能够减少需要训练的参数，减少运算开销，与此同时，也可以达到较高的精度，所以，VGG 网络和其他卷积神经网络相比，具有较小的卷积核和较小的跨步。

本文采用 VGG 神经网络进行人脸识别。VGG16 网络结构模型的部分示意图如下图所示，图 2.10 是卷积层和下采样示意图，图 2.11 是全连接层示意图。

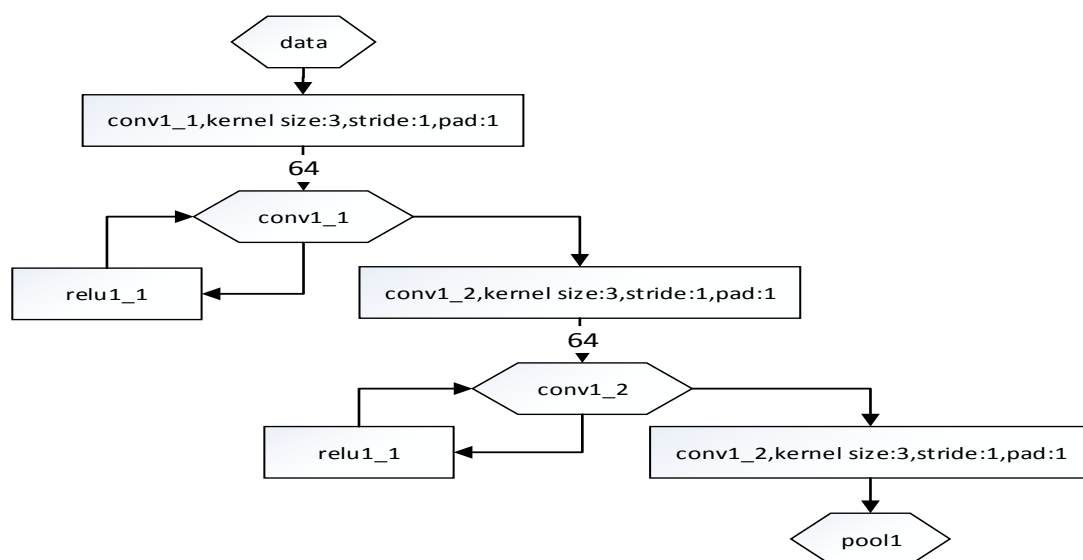


图2.10 卷积层和下采样层

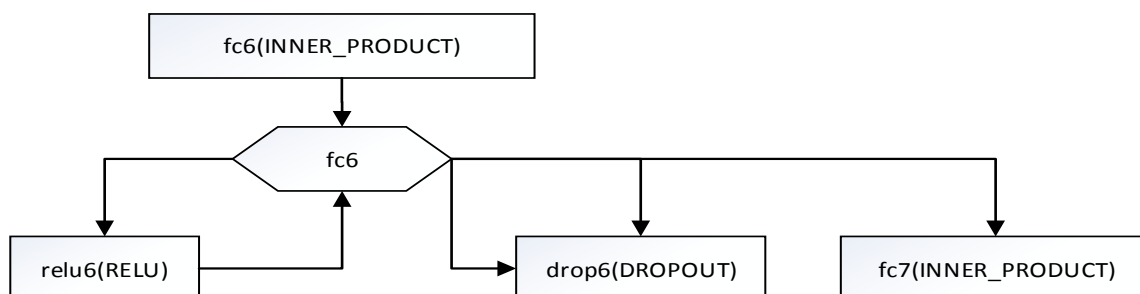


图2.11 全连接层示意图

### 2.3.3 分类器

分类器主要是对输入的数据进行类别的判定。在深度学习理论中，有很多经典的分类器，例如 Softmax<sup>[42]</sup>分类器；逻辑回归(Logistic Regression, LR)和 SVM 分类器等。

1) Softmax 分类器主要是针对多分类问题，类别之间是相互排斥的，即每一个输入样本只能被归为一类。

2) LR 分类器主要用于二分类问题，多个 LR 分类器可以进行多分类，但是输出

的类别之间不是相互排斥的。

3) **SVM** 是机器学习中经典的分类算法之一，也可以作为分类器使用。

**SVM** 是一种二分类模型，通过确定一个分类的超平面，来使特征空间上的间隔最大化，即使用一个核函数将数据映射到高维空间，从而可以解决原始空间线性不可分的问题。只需要少数的样本信息就可以确定分类超平面

**SVM** 线性超平面的示意图如图 2.12 所示。

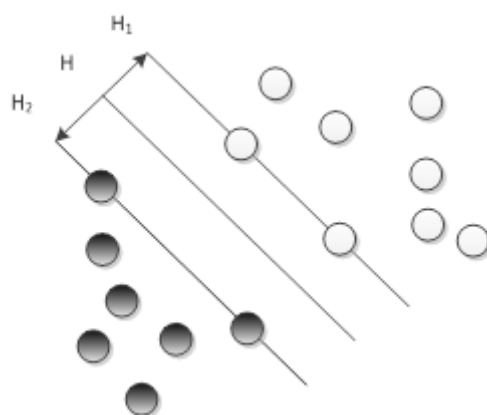


图2.12 线性分类超平面

通过观察图可知，在此二维空间中，样本具有线性可分性，即我们在此二维空间画一条直线，可以使该空间所有的正负样本完全分开。但是可以使正负样本完全分开的直线不止一条，**SVM** 算法需要找到使正负样本之间的几何间距最大的那条直线。通过使样本与分类超平面的距离越远，就会使类间距离增大，类内的距离减少，进而使分类的置信度就越高，分类效果就更好。

综上所述，**SVM** 的核心思想就是通过某种方法，使几何间距尽可能的大，使得该分类器对样本分类的置信度达到最大，即通过一定的方法，使置信度最小的样本尽可能达到最大的置信度。通过很少的训练样本对 **SVM** 分类器进行训练，就可以达到很好的识别率。

### 2.3.4 卷积神经网络的训练过程

卷积神经网络用于图像分类是有监督的学习过程，有监督学习的过程指的是在训练过程中，图像类别是已知的，图像经过网络得出的最终分类结果，得到的分类结果和实际的类别进行比较，若得到的分类结果和实际的类别相同，则结束，若不同，则计算两者之间的误差，然后对训练的网络权值进行更新，直到得到的分类类别与实际的类别之间的误差在设定的阈值内则停止训练。

卷积神经网络通过大量的输入，能够学习到与之对应的输出，它实际上是一种输入到输出的映射。在映射的过程中，卷积神经网络不需要任何表示输入到输出之间的精确表达关系式，只是用已知的模型对卷积神经网络进行训练，就能够得到输入与输出之间的映射关系。

卷积神经网络的训练过程与传统神经网络类似，也是参照了反向传播算法，主要步骤如下：

第一阶段，向前传播阶段：

- 1) 从样本集中取一个样本  $(X, Y_p)$ ，将  $X$  输入网络；
- 2) 计算相应的实际输出  $O_p$ 。

在此阶段，信息从输入层经过逐级的变换，传送到输出层。这个过程也是网络在完成训练后正常运行时执行的过程。在此过程中，网络执行的是计算（实际上就是输入与每层的权值矩阵相点乘，得到最后的输出结果）： $O_p = F_n(\dots(F_2(F_1(X_p W_1) W_2) \dots) W_n)$

第二阶段，向后传播阶段

- 1) 算实际输出  $O_p$  与相应的理想输出  $Y_p$  的差；
- 2) 按极小化误差的方法反向传播调整权矩阵。

## 2.4 本章小结

本章主要介绍了人脸检测与人脸识别的相关背景和理论知识，因为本文使用 VGG 网络进行人脸识别，而 VGG 网络是经典的卷积神经网络之一，所以本章节先对卷积神经网络进行了介绍，因为人脸识别的结果与分类器的好坏密切相关，所以也介绍了分类器的相关知识。最后阐述了空间金字塔池化算法，这些理论为人脸检测与识别提供了理论基础，给后续章节提供了很好的指导意义。

## 第三章 基于 Faster R-CNN 算法的人脸检测与识别

### 3.1 Faster R-CNN 人脸检测算法的改进

#### 3.1.1 RPN 网络的改进

本文第二章对 Faster R-CNN 算法进行了比较详细的介绍，通过对 Faster R-CNN 进行研究发现：RPN 在预测人脸的位置时，采用滑动窗口的方式在最后一层卷积特征图进行穷举。该方法有两个缺点：1) 由于滑动窗口的大小是固定的，对于真实场景的人脸预测会有很大的限制，因为不同图像中人脸的大小是一样的，有可能会非常大，也有可能非常小；2) 滑动窗口会产生较多的窗口冗余，导致计算量过大，时间复杂度过高。

本文将通过一个简单的策略对人脸的检测效率进行提高，将改进的 RPN 命名为 IRPN (Improved RPN)。这个策略就是用固定尺寸分割策略代替滑动窗口策略。对最后一层的共享卷积特征图，进行  $M$  种图像分割，再将划分出来的每一个分割窗口映射到  $K$  种形状估计，之后的过程不发生改变。也就相当于将原本固定大小为  $N \times N$  卷积核变为多种尺度大小的卷积核操作，也就相当于构建了多个 RPN 网络模型。改进的 RPN 网络如图 3.1 所示。

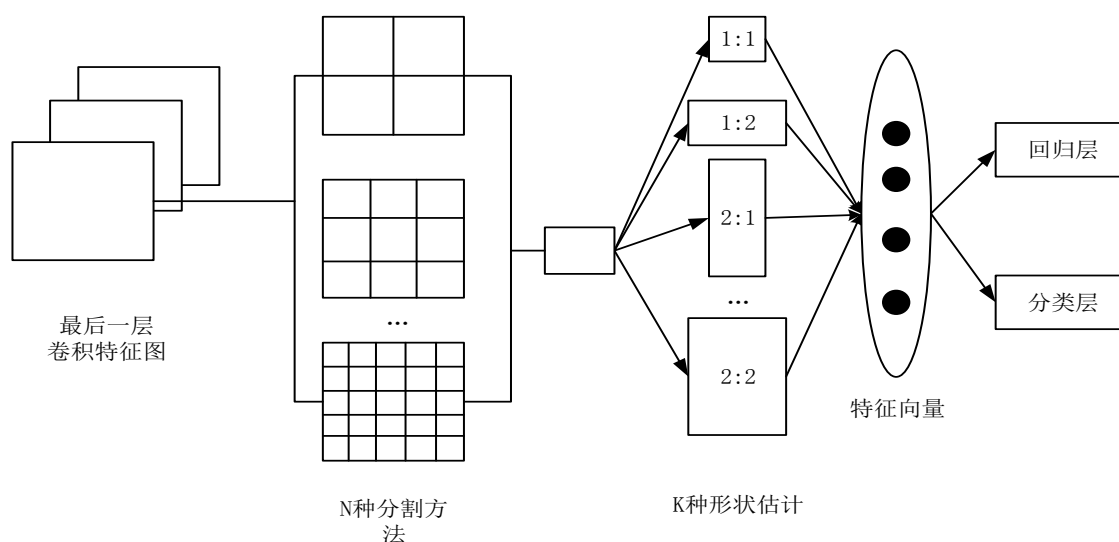


图3.1 IRPN 网络示意图

如图 3.1 所示，采用 3 种固定尺寸的分割方法，固定尺寸大小分别为  $2 \times 2$ ,  $3 \times 3$ ,  $5 \times 5$ ，则产生的候选框总个数是  $K \times (2 \times 2 + 3 \times 3 + 5 \times 5)$ ，但是包含了更多尺寸的目标估计，对极端情况的人脸图像有很好的识别效率，例如，人脸图像非常大或非常小的

情况。

对于同一个人脸区域，可能被包围在不同尺寸大小的目标包围盒中，所以需要把所有的目标包围盒进行筛选，过滤掉 IOU 比值较小的目标包围盒，选出最合适的目标包围盒。而如何对 IOU 比值较小的目标包围盒进行过滤，可以将其送入训练好的 VGG 网络进行判别，通过 VGG 网络的输出向量判断当前的目标包围盒是否是最佳的人脸区域，若是，则得到了人脸的最终区域。多个尺度和多个比例基准框，能够有效提高对极端人脸区域的检测，也就是对人脸区域特别大，或者人脸区域特别小的情况检测。

### 3.1.2 非极大抑制

由于同一个人脸可能被多个目标包围盒包含，所以我们需要对目标包围盒进行筛选，最终得到最合适的候选区域。本文采用非极大抑制（Non-maximum suppression, NMS）的方法对多个目标包围盒进行过滤，具体步骤有：

- 1) 按照一定的算法，得出所有目标包围盒的分数，并按降序排列，选出得分最高的目标包围盒。
- 2) 遍历剩余的所有目标包围盒，如果和当前得分最高的目标包围盒的重叠面积大于一定的阈值，则删除这个目标包围盒。
- 3) 迭代上述过程，最终得到分数最高的目标包围盒。

### 3.1.3 构建损失函数

IRPN 卷积神经网络的训练是一个有监督的过程，所以对于 IRPN 的训练需要构建一个损失函数。对于神经网络的训练来说，必须有正类样本和负类样本，从而需要挑选出合适的候选区域再进行训练。当候选区域的 IOU 值超过 0.7 时，将被定义为正类样本，低于 0.3 被定义为负类样本，候选区域样本则被丢弃。

因为此处的分类是两种，要么是正样本，要么是负类样本，而 SVM 是典型的二分类分类器，其损失函数 Hinge 损失函数的定义为：

$$\min_{w,b} \sum_i^N [1 - y_i(w * x_i + b)] + p \|w\|^2 \quad (3-1)$$

而此处的分类损失函数的公式可以直接引用 SVM 分类器的损失函数。

### 3.1.4 模型交替训练策略

改进的 Faster R-CNN 算法主要包括 IRPN 卷积神经网络和 Fast R-CNN 卷积神经网络，其中，IRPN 卷积神经网络由八个卷积层和一个 Softmax 层组成，Fast R-CNN

卷积神经网络由五个卷积层，一个 ROI pooling 层，四个全连接层和一个 Softmax 层组成。

IRPN 卷积神经网络和 Fast R-CNN 卷积神经网络的前五层都是卷积层，如果分别对改进的 RPN 网络和 Fast R-CNN 卷积神经网络进行单独训练，会导致卷积层参数不能共享，因此，需要采用一种模型交替训练策略，使得两者能够对卷积层的参数共享。根据 2.3 小节介绍的三种联合训练方法，本文采用轮流训练策略。训练过程如下：

1. 单独训练 IRPN 网络，使用大规模数据集训练好的模型对 IRPN 网络进行初始化。这一步操作是对 IRPN 进行微调，也就是用监督学习的方式去调整 IRPN 网络，进而实现卷积网络参数的共享。
2. 用 IRPN 生成的候选区域去训练 Fast R-CNN 卷积神经网络，而 Fast R-CNN 卷积神经网络也需要提前使用大规模数据集训练好的模型进行初始化。
3. 重新训练 IRPN 神经网络，将 IRPN 卷积神经网络前五层卷积层的学习率设置为 0，权重参数来自 Fast-rcnn 卷积神经网络模型，训练得到新的 IRPN 卷积神经网络模型。
4. 重新训练 Fast R-CNN 卷积神经网络，将 Fast R-CNN 卷积神经网络的前五层卷积层学习率设为 0，权重参数来自 IRPN 卷积神经网络模型，
5. 使用样本数据集和得到的人脸候选区域标注，对 Fast R-CNN 重新训练得到新的 Fast R-CNN 卷积神经网络模型。

### 3.1.5 ReLU 函数的改进

激活函数，用在各个卷积层和全连接层的输出位置。一般来说，激活函数是卷积神经网络非线性的主要来源。

一般来说，直线拟合的精确度与曲线拟合相比会差很多，为了在神经网络中达到曲线拟合的目的，我们经常使用非线性的激活函数，通常使用的激活函数是 sigmoid 函数，取值范围为(0,1)和 tanh 函数，取值范围为(-1,1)，该两种函数都是一种非线性函数，用下面的公式表示：

$$f(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (3-2)$$

$$f(x) = \text{sigmoid}(x) = \frac{1}{1 + e^{-x}} \quad (3-3)$$

但是，现阶段，ReLU 是最常用的激活函数，主要原因有：

1. ReLU 本质上是分段线性模型，前向计算非常简单，无需指数之类的操作；
2. ReLU 的偏导也很简单，反向传播梯度，无需指数或者除法之类操作；
3. ReLU 不容易发生梯度发散问题，Tanh 和 Logistic 激活函数在两端的时候导数容易趋近于零，多级连乘后梯度更加约等于 0；

4. ReLU 关闭了右边，从而会使得很多的隐层输出为 0，即使网络变得稀疏，起到了和正则化相似的作用，在一定程度上可以缓解过拟合问题；
5. RELU 是一个非线性操作，可以使得网络的非线性表达更加丰富。

ReLU 的公式为：

$$f(x) = \begin{cases} x & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (3-4)$$

在本文，将对 ReLU 函数进行一点改进，改进的 ReLU 函数避免了倒数为 0，无法传播的情况，也使整个网络的结构更加的平滑。

改进的 ReLU 公式为：

$$f(x) = \begin{cases} x & x \geq 0 \\ 0.33x & x < 0 \end{cases} \quad (3-5)$$

### 3.2 空间金字塔池化

空间金字塔池化<sup>[43]</sup> (Spatial pyramid pooling, SPP)，使得任意大小的特征图都可以转换成固定大小的特征向量。具体流程如下：

输入任意大小的一张图像，假设大小为 $(W, H)$ ，输出神经元的个数是 38 个，即我们希望特征向量是 38 维的，空间金字塔提取特征的过程如图 3.2 所示。

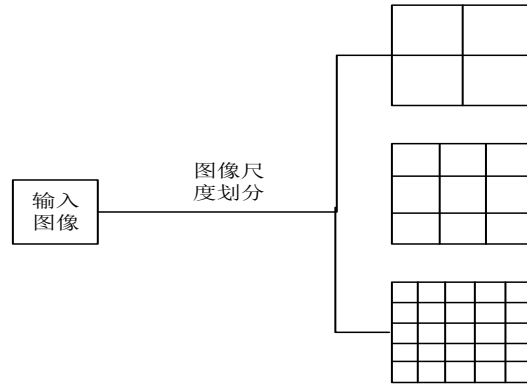


图3.2 SPP 示意图

如图 2.13 所示，对于输入的图像，我们采用大小不同的刻度对其进行划分，分别为 $(2 \times 2, 3 \times 3, 5 \times 5)$ ，从而可以得到 $4 + 9 + 25 = 38$ 个特征点，从这 38 个不同的特征点上抽取一个特征，从而就得到了一个 38 维的特征向量。

通过上述过程可知，从本质上来说，SPP 池化技术就是一种多尺寸的池化技术，如同当你通过多种角度解决一个问题，就会考虑的比较全面，从而，就一定程度上来说，SPP 池化技术可以提高图像提取的特征的表达能力，进而可以达到提高图像分类的准确率。



### 3.3 基于改进的 VGG 网络模型人脸识别算法

通过 2.4 小节对 VGG 网络模型的介绍可知，VGG 网络模型的最后三层是全连接层，而全连接层输入的维度必须是固定不变的，进而也就使输入图像的尺寸大小是固定的。所以我们对 VGG 网络模型训练前，需要对图像进行归一化处理，图像的归一化是指通过对图像进行一系列标准的处理变换，使之变换成固定标准形式的过程，一般对图像进行 crop 或 wrap 操作：

- 1) crop: 截取原图像的一个固定大小的 patch。(物体可能会产生截断)。
- 2) wrap: 将原图像的感兴趣区域 (Region of Interests, ROI) 缩放到一个固定大小的 patch (物体被拉伸，尤其是长宽比大的图像)。

而图像的归一化处理，会造成一定程度上的图像形变，而图像的形变可能会导致图像空间信息的损失。针对这个问题，我们可以在 VGG 网络的卷积层和全连接层之间加一层 SPP(空间金字塔池化)，使得不同尺寸的图像也可以产生固定的输出维度。

SPP 池化是一种多尺度的池化，可以从不同尺度反应图像的特征信息，在一定程度上来说，通过 SPP 可以提高了图像特征的表达能力，从而也就可以使人脸识别的准确率提升。

改进的 VGG 模型可以实现对多种尺度脸部特征的共享，增强了 VGG 模型分类识别的能力。改进的 VGG 网络模型层数不发生变化，具体框架图如图 3.3 所示。

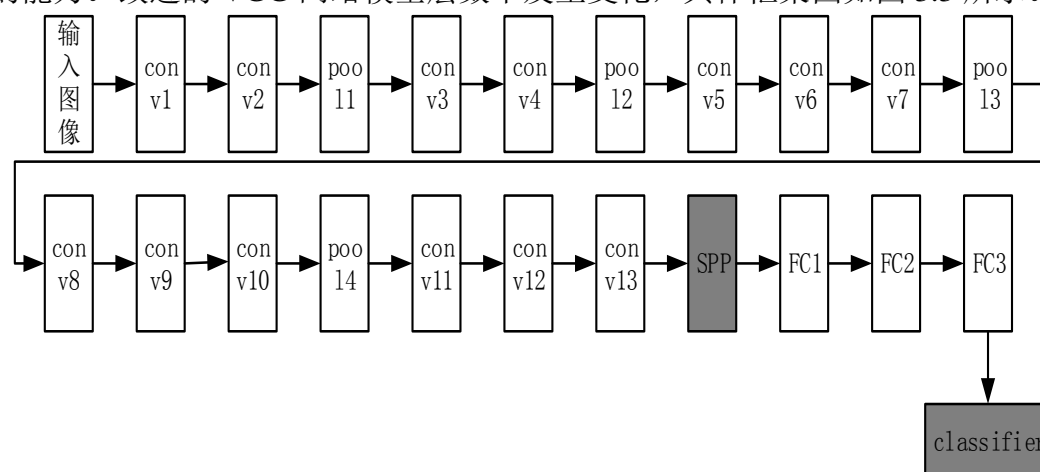


图3.3 改进的 VGG 网络模型

VGG 网络在最后的全连接层使用 Softmax 分类器进行图像的分类和识别，本文将卷积网络模型最后一层输出的全连接特征作为 SVM 分类器的训练样本，训练一个 SVM 分类器。将 SVM 分类器的识别结果与 Softmax 分类器进行对比。

### 3.4 数据集预处理

在样本数据训练之前，对样本数据进行的变换和准备工作叫做数据集的预处理。在本文中对所用到的图像预处理包括三个过程。

1. 对图像进行均值化处理，具体做法是通过统计每个像素点对应位置的均值，然后对所有的图像的对应位置减去对应的均值即可。这种简单的归一化策略好处就是减小输入样本的像素变量值，减轻噪声像素值的影响。Caffe 封装了图像均值计算工具。考虑到样本数据集特别大，利用工具计算时间开销特别大，故直接使用 Caffe 中已有的模型提供的均值数据。

2. 对用于训练的图像进行数据库封装。采用直接从内存读取图像的方式效率很低，Caffe 提供了 Lmdb 和 Leveldb 两种数据库接口，可以将图像集封装成数据库。这两种方式都是键值对嵌入式数据库管理系统编程库，lmdb 速度比 Leveldb 大约快 15%，但是，在内存消耗上，Lmdb 要略大于 Leveldb，默认情况下 Caffe 使用的是 Lmdb 数据库。本文也采用的是 Lmdb 数据库。

3. 对于改进的 Faster-RCNN 进行训练，需要将数据集格式修改为 VOC2007，具体步骤如下：

- 1) 需要对图像的名字进行格式化命名，可以通过运行脚本进行批量化操作；
- 2) 需要对图像进行标框，提取目标物体的包围框坐标（左上角和右下角）；
- 3) 将第二步做好的 TXT 文件转成 XML 文件；
- 4) 将保存好的 XML 放到命名为 Annotations 的文件夹；
- 5) 再将训练的图像放到命名为 JPEGImages 的文件夹；

当完成以上工作之后，就可以将得到的数据集进行训练，进而得到人脸的检测与人脸识别模型。

### 3.5 模型的训练

对于模型的训练来说，solver 文件的参数的设置对模型最终分类结果的好坏有决定性的影响，对 solver 文件中的主要参数进行说明：1) iteration: 数据进行一次正向-反向的训练；2) batchsize: 每次迭代训练图像的数量 3) epoch: 1 个 epoch 就是将所有的训练图像全部通过一次网络；4) 学习率和模型训练速度的快慢有较大的关系。

对本文改进的网络模型采用微调的模式进行训练。具体原因如下：

经过大量数据集训练出来的模型具有非常强大的泛化能力，而经过对模型各层特征的可视化证明：一般来说，网络模型的前几层结构的特征是泛化的，在网络的最后几层提取出来的特征才会具有训练样本数据集的特定性，也就是说前面的卷积层一般提取的是颜色和边缘这些特征，通过将模型前面几层的学习率设置为 0，从而加快模

型的训练过程，也不会对模型的检测和识别结果造成影响。

在进行模型训练之前，需要将数据集按照第四小节的格式进行修改。在训练前需要准备的文件有，以改进的 VGG 模型训练为说明，如图 3.4 所示。

snapshot	2016/7/7 10:16	文件夹
vgg_train_lmdb	2016/7/5 9:44	文件夹
vgg_val_lmdb	2016/7/5 9:44	文件夹
create_imagenet.sh	2016/7/5 9:43	SH 文件
face_mean_train.binaryproto	2016/7/7 9:29	BINARYPROTO
fine-tuning.sh	2016/7/5 10:24	SH 文件
make_mean.sh	2016/7/7 9:29	SH 文件
sloper_prototxt.prototxt	2016/7/7 9:23	PROTOTXT
VGG_FACE.caffemodel	2015/10/14 0:55	CAFFEMO
VGG_FACE.prototxt	2016/7/7 10:11	PROTOTXT

图3.4 模型训练前的准备

### 3.6 人脸检测与识别算法评价标准

对人脸检测与识别的结果进行评价是人脸识别作为身份鉴定的重要环节。一般包括人脸检测与识别的准确率、精确度、效率等几个方面。

#### 1) 准确率和精确率

表3.1 分类器易混淆名词矩阵

		真实类别	
		正类别	负类别
预测类	正类	真阳性(TP)	假阳性(FP)
	负类	假阴性(FN)	真阴性(TN)

其中，真阳性是指将属于正类的样本正确地预测为正类的个数，而假阳性则是将属于负类样本错误地预测为正类样本的个数。假阴性是指将属于正类的样本错误地预测为负类的个数，而真阴性则是将属于负类的样本预测为负类的个数。

假设，将样本中正类样本记为 T，则  $T=TP+FN$ ，负类样本记为 N，则  $N=FP+TN$

一般，物体的检测与识别结果，通过以下指标进行判断，其中，物体的检测通过指标 IoU(重叠度)进行判断：

重叠度 IoU：

$$IoU = \frac{\text{Region Proposal} \cap \text{Ground Truth}}{\text{Region Proposal} \cup \text{Ground Truth}} \quad (3-6)$$

通过 IoU 的值来判断目标物体是否被检测到，如果  $IoU < 0.5$ ，则认为物体没有被

检测到，反之，则认为物体已经被正常检测到。

精确率 Precision :

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3-7)$$

准确率 Accuracy :

$$\text{Accuracy} = \frac{TP+TN}{P+N} \quad (3-8)$$

错误率 Error :

$$\text{Error} = \frac{FP+FN}{P+N} \quad (3-9)$$

召回率 Recall :

$$\text{Recall} = \frac{TP}{P} \quad (3-10)$$

## 2) 识别的效率

识别的效率是人脸识别技术的一个重要性能指标。而对于识别效率的考量，一般主要从以下两个方面来考虑：一种是图像的人脸识别，也就是静态的人脸识别，这种人脸识别速度一般可以达到每秒处理 20 万张人脸图像；另一种是基于实时的人脸识别，也就是动态的人脸识别，这种人脸识别一般有时效性的要求，也就是从检测到人脸到最后识别出人脸，匹配到该人脸的身份，要求时间限定在 2s 左右。

## 3.7 实验结果与分析

### 3.7.1 实验环境

本章中的神经网络模型是基于 Caffe<sup>[44-45]</sup>平台实现的，所以对 Caffe 平台做一个简单的介绍。

随着深度学习的发展，出现了很多优秀的深度学习框架，例如，Theano<sup>[46-47]</sup>、TensorFlow<sup>[48-49]</sup>、Torch<sup>[50]</sup>和 Caffe 等。其中，Theano 是深度学习框架中的元老，用 python 编写；TensorFlow 是谷歌推出的开源深度学习框架，通过 C/C++引擎加速，可以部署在一个或多个 CPU、GPU 的台式计算机或者服务器中，也可以使用单一的 API 应用在移动设备中；Torch 是 Facebook 推出的有大量机器学习算法支持的科学计算框架，Torch 构建模型简单，支持 GPU，高度模块化等特点；Caffe (Convolutional Architecture for Fast Feature Embedding) 是由贾扬清博士创建的，由 Berkeley Vision and Learning Center 和社区工作者一起维护的。它是由表达式，速度和模块三部分组成的深度学习框架，主要有五个组件：Blob, Solver, Net, Layer 和 Proto。

1. Caffe 通过使用 Blob 来存储数据的结构，是一个不定维的矩阵，在 Caffe 中一般用其表示一个四维矩阵，即  $N \times C \times H \times W$ ，其中  $N$  是图像的数量， $C$  是图像的通道数， $H$  是图像的高度， $W$  是图像的宽度。使用深度学习框架 Caffe 进行实验在进行模型训练之前，需要将图像变换为 Blob 的形式。

2. Solver 用来训练深度网络，每个 Solver 中都包含了一个训练网络的对象和一个测试网络的对象。

3. Net 则是由若干个 Layer 构成的。每个 Layer 的输入特征向量和输出特征向量表示为 Input Blob 和 Output Blob。

4. Proto 则基于 Google 的 Protobuf 开源项目，是一种类似 XML 的数据交换格式，用户只需要按照格式定义对象的数据成员，可以在多种语言中实现对象的序列化与反序列化，在 Caffe 中主要用于网络模型的结构定义、存储和读取。

通过使用 Caffe 平台，我们可以通过使用一个简单的语言（google protobuf）定义网络结构，又由于 Caffe 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，所以就可以通过简单的命令在 CPU 或者 GPU 上执行代码。一个完整的基于 caffe 平台的 CNN 网络模型训练流程如图 3.5 所示。

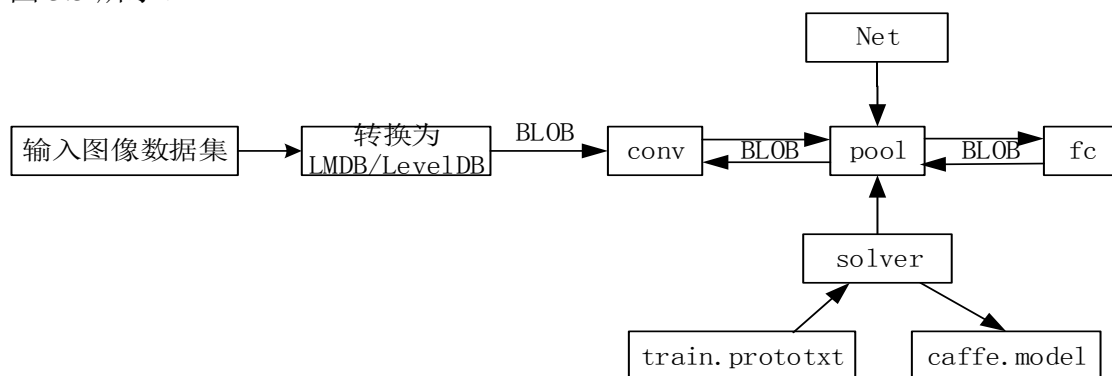


图3.5 基于 Caffe 平台的 CNN 模型流程图

由于 Caffe 具有使用方便、运行速度快、模块化程度很高，既能很方便地扩展到新的任务和设置上，也可以很方便地对自己设计的模型进行定义，并且可以通过社区和其他人进行交流等优点，所以使用此平台进行人脸检测与识别软件的开发。

### 3.7.2 实验的数据集

对改进的 RPN 网络和改进的 VGG 网络训练完成之后，需要进行测试，本文使用以下数据集进行测试，该测试数据集分为两个部分：第一部分是人脸测试数据集，包括 Fddb 数据集和 WIDER FACE 数据集；第二部分是人脸识别数据集，包括 LFW 数据集和 CelebA 数据集。

1) Fddb 数据集包括 2845 张图像, 共有 5171 个人脸作为测试集。测试集范围包括: 不同姿势、不同分辨率、旋转和遮挡等图像, 同时包括灰度图和彩色图, 标准的人脸标注区域为椭圆形。

Fddb 部分数据集如图 3.6 所示。

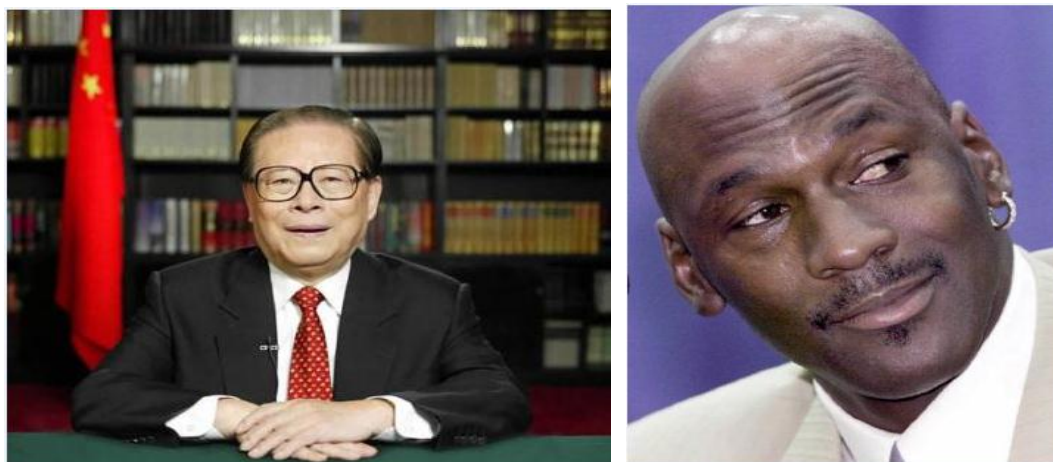


图3.6 Fddb 部分数据集示例图

2) WIDER FACE 是香港中文大学提供的一个更广泛的人脸数据, 是人脸检测基准数据集, 它包含 32203 张图像和 393703 个标注人脸图像, 其中, 标注人脸在尺度、姿势、装扮和遮挡等方面表现出了较大的变化。WIDER FACE 是基于 61 个事件类别组织的, 对于每一个事件类别, 选取其中的 40%作为训练集, 10%用于交叉验证 (cross validation), 50%作为测试集。

WIDER 部分数据集如图 3.7 所示。



图3.7 WIDER 部分数据集示例图



3) LFW (Labeled Faces in the Wild) 是一个用于研究无约束的人脸识别的数据库。该数据集包含了从网络收集的 13000 张人脸图像, 每张图像都用被拍摄人的名字进行命名。其中, 有 1680 个人有两个或两个以上不同的照片。

LFW 部分数据集如图 3.8 所示。



图3.8 LFW 部分数据集示例图

4) CelebA (Large-scale CelebFaces Attributes) 数据集包括 10K 名人, 202K 脸部图像, 每个图像 40 余标注属性。

CelebA 部分数据集如图 3.9 所示。



图3.9 CelebA 部分数据集示例图

### 3.7.3 改进的 VGG 模型的参数选择

对于网络模型的训练, 学习率的设置, 以及迭代次数的多少, 会对最终训练出来的模型产生决定性的影响。对于学习率的选择, 都是通过多次实验进行优化, 一般来说, 学习率一般是位于 0 和 1 之间的一个实数, 控制着每一轮迭代中的更新步长, 学习率如果设置的过小, 网络的收敛速度就会很慢, 而设置的过大, 会使网络发生震荡。

学习率设置的大小和模型训练的快慢有很大的关系，模型的初始训练可以将学习率设置的高一些，如 0.01，当模型的识别结果已经达到较好的准确率，可以适当减小学习率，如 0.001。图 3.10 是对训练好的模型进行测试，由图可知对样本识别的准确率达到 99.99%。

SPP 技术可以使得任意大小的特征图都可以转换成固定大小的特征向量。

```
root@admingpu-G1-Sniper-Z97: /home/admin-gpu/liu-test
ion.
I0618 15:53:34.607620 13232 net.cpp:169] pool1 does not need backward computat
n.
I0618 15:53:34.607623 13232 net.cpp:169] relu1_2 does not need backward computat
ion.
I0618 15:53:34.607627 13232 net.cpp:169] conv1_2 does not need backward computat
ion.
I0618 15:53:34.607630 13232 net.cpp:169] relu1_1 does not need backward computat
ion.
I0618 15:53:34.607635 13232 net.cpp:169] conv1_1 does not need backward computat
ion.
I0618 15:53:34.607637 13232 net.cpp:205] This network produces output prob
I0618 15:53:34.607656 13232 net.cpp:447] Collecting Learning Rate and Weight Dec
ay.
I0618 15:53:34.607664 13232 net.cpp:217] Network initialization done.
I0618 15:53:34.607667 13232 net.cpp:218] Memory required for data: 114620912
libprotobuf WARNING google/protobuf/io/coded_stream.cc:487] Reading dangerously
large protocol message. If the message turns out to be larger than 2147483647 b
ytes, parsing will be halted for security reasons. To increase the limit (or to
disable these warnings), see CodedInputStream::SetTotalBytesLimit() in google/p
rotobuf/io/coded_stream.h.
E0618 15:53:35.301839 13232 single_image.cpp:35] Using GPU #
E0618 15:53:35.461959 13232 single_image.cpp:91] max: 0.999977 i 2
root@admingpu-G1-Sniper-Z97: /home/admin-gpu/liu-test#
```

图3.10 模型的部分识别结果示意图

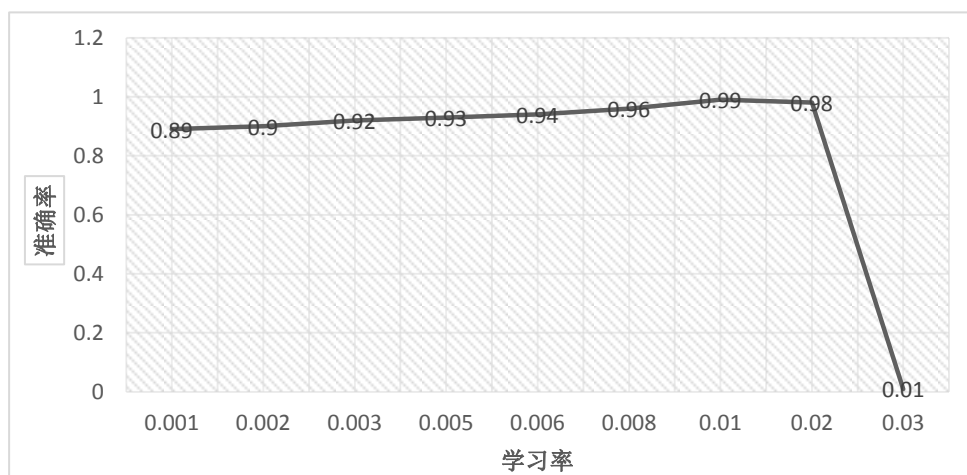


图3.11 改进的 VGG 模型对人脸识别数据集分类结果图

从图 3.11 可以看出，当学习率在 0.001 到 0.02 的区间时，VGG 网络对人脸图像的识别率成稳定性增长的趋势，但是当设置为 0.03 时，就会因为学习率过大，产生网络震荡，进而对人脸识别的准确率极低。



### 3.7.4 实验结果与分析

当完成模型的训练之后,用人脸检测数据集和人脸识别数据集分别对模型进行测试,改进的 Faster R-CNN 算法对人脸图像的检测效果更好,尤其是对人脸图像中人脸较小或较大的情况,在人脸检测的基础上,人脸识别的准确率达到 99% 左右。由表 3.2 可知,牛津大学提出的 VGG 模型在人脸识别的准确率上面达到了 98.95,而经过的改进的 VGG 模型,由于不再对输入的人脸图像大小进行限定,也就不需要对输入的人脸图像进行预处理,因为对图像的预处理,会造成部分图像信息的丢失,进而影响模型对人脸图像特征的提取,而特征对人脸识别的准确率有决定性的影响,从而,改进的 VGG 模型对人脸图像的识别有着更高的准确率。表 3.2 展示了改进的 VGG 模型和其他经典的人脸识别算法的比较。

表3.2 几种典型的人脸识别算法结果统计表

No	Method	Images	Networks	Acc.
1	DeepFace	4M	3	97.35
2	Fusion	500M	5	98.37
3	DeepID-2,3		200	99.47
4	FaceNet	200M	1	98.87
5	FaceNet+Alignment	200M	1	99.63
6	VGG	2.6M	1	98.95
7	改进的 VGG		1	99

## 3.8 本章小结

本文的第二章对 Faster R-CNN 算法和 VGG 模型进行了详细的介绍和分析,本章在第二章的基础上,接着提出用固定尺寸分割策略代替了 RPN 网络中的滑动窗口策略,提升了 Faster R-CNN 算法对人脸的检测性能,也将空间金字塔池化技术引入 VGG 网络,使得 VGG 网络输入图像的尺寸大小不再固定,进一步提升了 VGG 对人脸识别的准确率。最后对 Caffe 开源平台进行了介绍,阐述了模型交替策略、数据集的预处理、测试数据集、ReLU 函数和人脸检测与识别算法评价标准。



## 第四章 基于 Faster R-CNN 的人脸检测与识别软件设计

### 4.1 软件设计目标

本软件采用 Visual Studio 2010 MFC(Microsoft Foundation Classes)开发实现，在深度学习框之一，Caffe 平台的基础上，结合 OpenCV 计算机视觉库对人脸数据图像格式进行处理，主要完成了对人脸的检测与识别功能。主要从以下几个方面对软件进行设计：

1. 正确性和高效性：也就是该软件在满足对人脸检测和识别基本功能的同时，也应该具有很好的时效性，从输入图片到最终的结果显示，时长不应该超过 2 秒。
2. 界面的友好性：对于用户来说，可以很方便，快捷地使用。
3. 健壮性和可扩展性：软件在遇到输入的文件格式不能被正确识别时，应该及时地告知用户，而不应该让用户长久的等待。出于对可扩展性的考虑，软件应该进行模块化。
4. 功能的模块化与独立性：软件对每个功能的调用都采用菜单项及动态链接库的形式；各功能模块以独立的动态链接库形式存在，满足功能的独立性。
5. 接口的统一性与实现的封装性：在实现动态链接库时，借助面向对象的设计思想，对实现的细节进行封装，最终通过 Caffe 平台通过接口进行调用，实现了接口的统一性与实现的封装性。

综上所述，出于这些方面的考虑，以动态链接库的形式对人脸检测与识别软件进行整体设计。

### 4.2 软件系统设计

#### 4.2.1 系统结构设计

由于人脸检测与识别算法流程设计到的所有功能，都是对图像进行处理，并且最终得到的结果也是图片，出于对 Caffe 与动态链接库之间独立性的考虑，我们将只依赖 OpenCV 计算机视觉库对数据图像格式进行处理。OpenCV 是一个基于开源发行的跨平台计算机视觉库，它为物体跟踪、数据挖掘、图像处理、模式识别、机器学习、三维重建和线性代数等领域提供了多种多样的算法，并对其进行了优化和加速。基于 Faster R-CNN 的人脸检测与识别软件框架设计如图 4.1 所示。

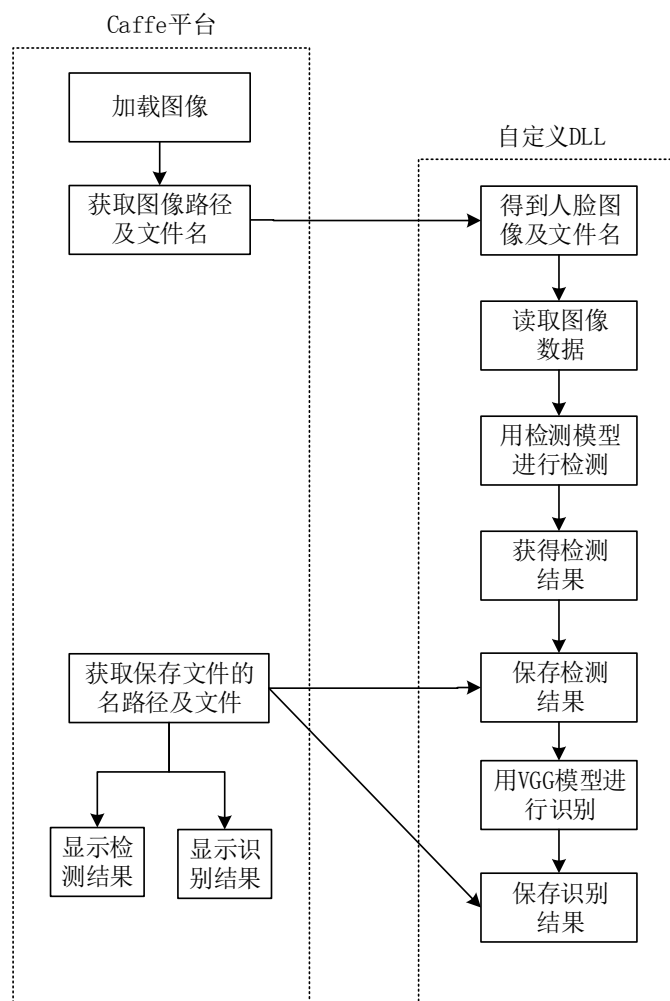


图4.1 基于 Faster R-CNN 的人脸检测与识别软件框架

#### 4.2.2 人脸检测与识别算法功能模块设计与实现

本文第三章给出的基于改进的 Faster R-CNN 算法中，主要包括两个网络模型：IRPN 网络与 Fast R-CNN 网络，两个网络共同完成对人脸图像的检测；而人脸的识别由改进的 VGG 完成。因此，本软件主要包括两个过程：首先是对加载的人脸图像通过改进的 Faster R-CNN 算法进行检测，其次将检测到的人脸送入改进的 VGG 模型进行识别，得到最终的识别结果。

因此本软件主要分为两大模块，即人脸检测模块和人脸识别模块，通过人脸检测模块，可以实现对图像的预处理，人脸图像中人脸的精准定位、特征的提取功能，而通过人脸识别模块，可以对图像中人脸进行分类识别，即将该图像中的人脸与数据集的人脸图像进行匹配和识别，假设数据集中的人脸身份都是已知的。软件的工作流程如图 4.2 所示。

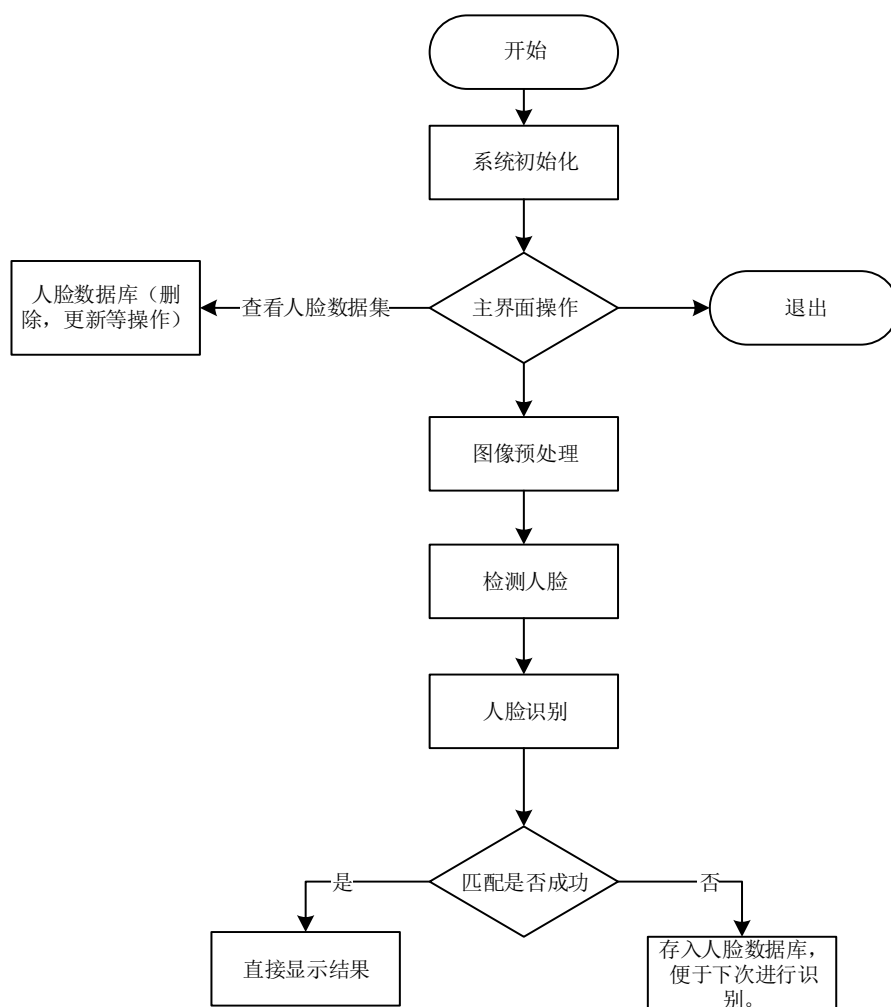


图4.2 人脸检测与识别系统的工作流程图

由图 4.2 可知, 软件主要是对人脸人脸检测与识别的实现, 人脸检测的功能类图如图 4.3 所示, 人脸识别的功能类图如图 4.4 所示。

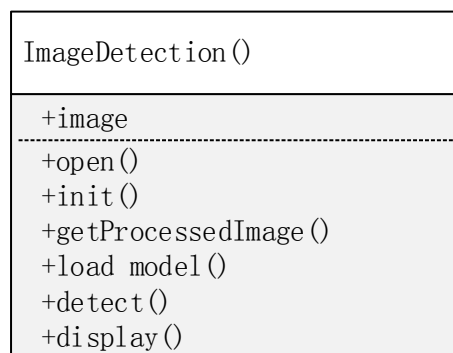


图4.3 人脸检测类图

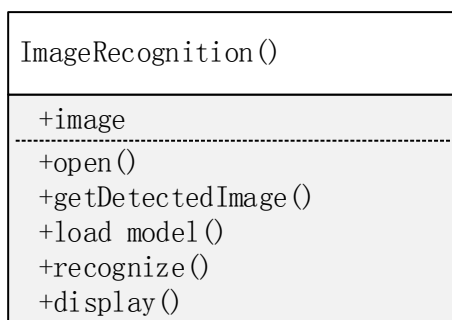


图4.4 人脸识别类图

人脸检测与识别算法功能模块设计如图 4.5 所示。

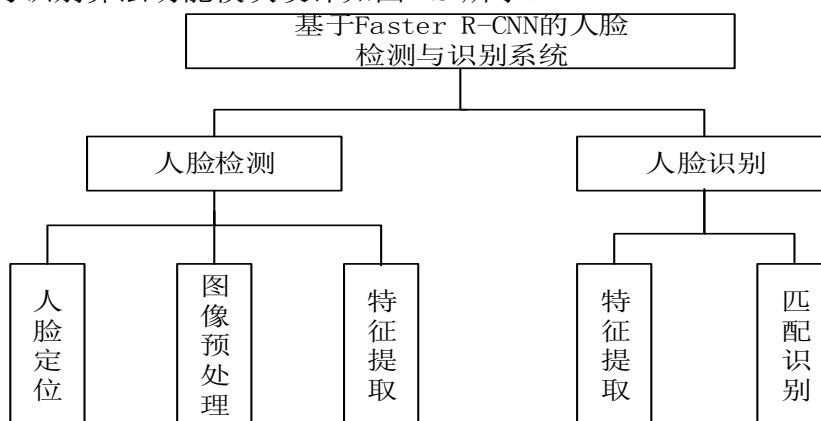


图4.5 人脸检测与识别算法功能模块设计

在对人脸图像进行匹配识别之前，需要对该人脸样本进行训练和学习，所以，当未知人脸第一次出现时，需要对该人脸的图像样本进行收集，通过提前训练好的模型提取特征，便于下次出现时的识别。为了对未知的人脸特征进行更好地提取，进而达到较好的识别率，需要收集该人脸的多张图像。收集多张人脸图像的思路是：先采集待测人的人脸正面图像，再采集通过姿态、角度、遮挡物的变化而产生该人脸的其他图像。

对图像的收集完成之后，该软件可以对曾经未知的人脸图像进行检测和识别，接下来对软件的各个功能模块进行详细的说明：

#### 1) 人脸检测

采用本文第三章阐述的基于改进的 **Faster R-CNN** 算法对收集到的人脸图像进行准确的定位，人脸检测的算法流程图如图 4.6 所示。

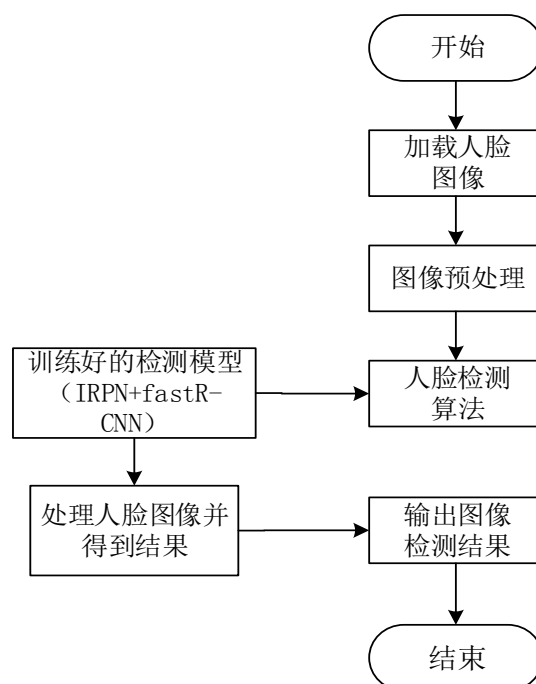


图4.6 人脸检测算法流程图

## 2) 图像预处理

将进行过定位的人脸图像进行预处理，处理的过程和第三章的第五小节相同，主要是为了提高系统的检测与识别的准确率和效率。

## 3) 特征提取

收集了人脸图像之后，接着就是对所需要的人脸特征的提取。本文采用深度卷积神经网络来提取特征，提取到的特征保证了一定程度上的位移、尺度、平移和形变等不变性。所以该软件对于不同姿态和表情的人脸，具有较强的鲁棒性。

## 4) 人脸识别

人脸的匹配与识别也就是将采集到的人脸数据和图像数据集中的数据进行对比，假设图像数据集中有该人脸的信息。本文通过已经充分训练过的改进网络模型 VGG 进行人脸的匹配和识别。

人脸识别流程图如图 4.7 所示。

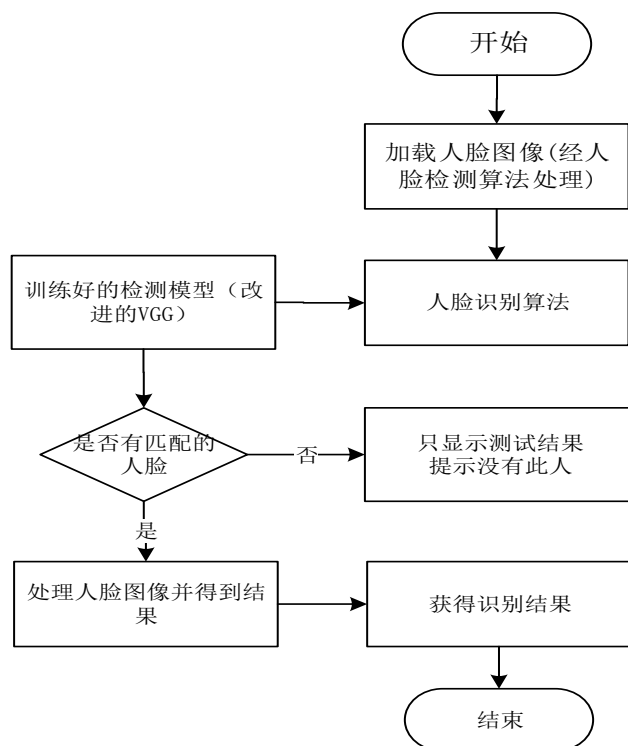


图4.7 人脸识别算法流程图

### 4.3 软件检测与识别结果

由第三章可知,训练好的模型对人脸检测与识别准确率较之前已经有了很好的提升。本小节主要是对基于改进的人脸检测与识别算法设计的人脸检测与识别软件进行测试,主要是对软件功能的测试。众所周知,一个好的软件除了尽可能地检测和识别到图像中的所有人脸,也应该有较好的鲁棒性。

#### 1) 对于软件功能的测试

通过点击该软件菜单项,是否能进行人脸人脸的检测、特征的提取和人脸的识别。软件功能测试界面如图 4.8 所示。





图4.8 软件主界面运行图

## 2) 软件性能的测试

该软件通过 RPN 网络模型和 Fast R-CNN 网络模型进行人脸的检测，和改进的 VGG 模型进行人脸识别，所以通过对该软件的测试可以从软件检测与识别的准确率和效率两方面进行。该软件的检测结果如图 4.9 所示。

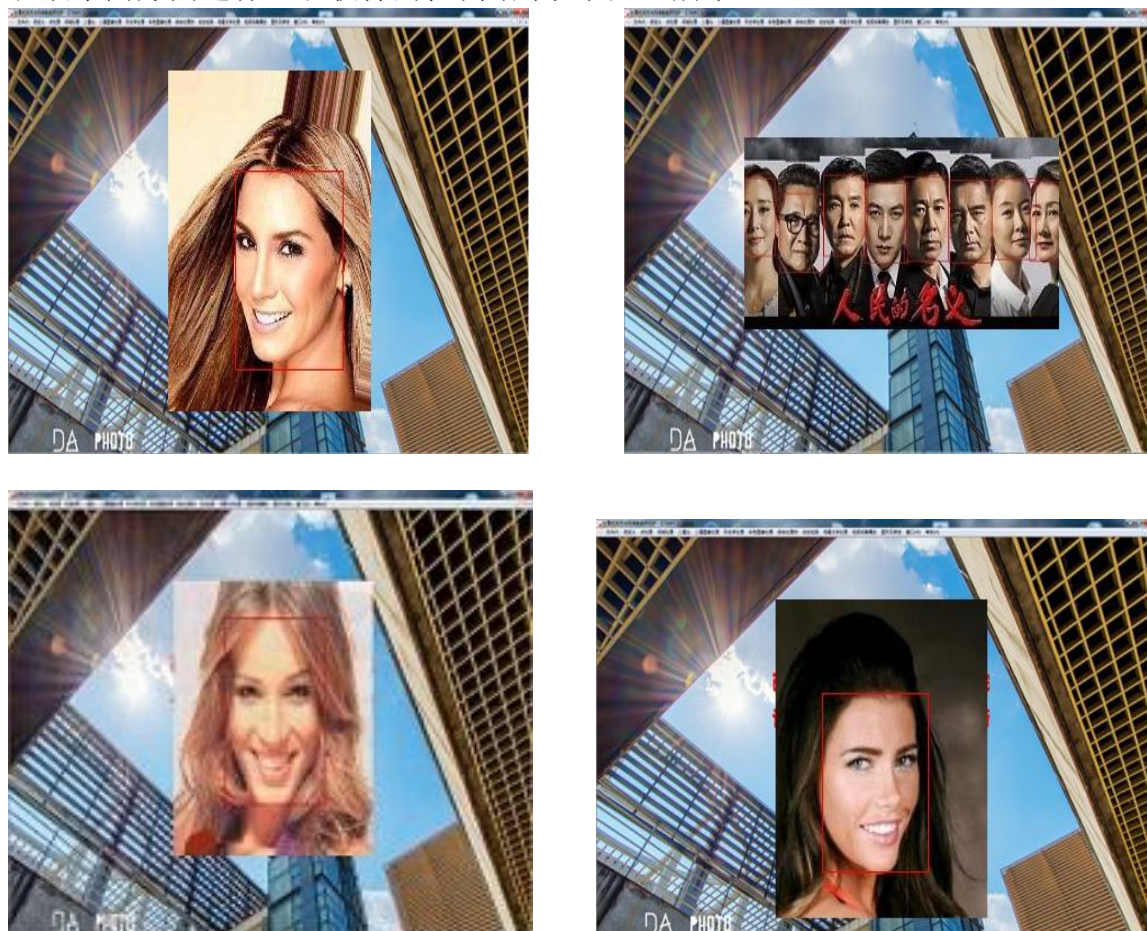


图4.9 人脸检测结果示意图



由图 4.6 可知该软件对人脸图像的检测效果比较好,将经过检测的人脸图像使用改进的 VGG 网络识别,可以加快识别的效率和准确率,因为经过人脸检测的人脸图像去除了复杂无用的背景,可以加快改进的 VGG 模型提取特征的速度,进而提高该模型对人脸的识别效率,通过对第三章的数据进行实验表明,该软件对人脸识别的准确率达到了 99% 以上。该软件的识别结果如图 4.10 所示。



图4.10 人脸识别效果示意图

## 4.4 本章小结

本章对 MFC 框架、OpenCV 进行了简单的介绍,然后在此基础上设计并实现了人脸检测与识别软件,详细介绍了该软件的框架设计和功能模块的划分,最后,根据前面介绍的人脸检测与识别算法的数据集,对该软件的检测与识别结果进行了说明与展示。

## 第五章 总结与展望

### 5.1 总结

本文在研究了人脸检测与识别的相关背景和理论知识之后,对目前人脸检测与识别用到的算法进行了深入分析。在人脸检测与识别领域,深度学习是最值得尝试的方案。自 Hinton 教授提出深度学习理论以后,就得到了学术界和工业界的广泛关注,越来越多的学者和企业使用深度学习理论去解决在人脸检测和识别中碰到的难题。深度学习理论包含了很多不同的网络模型,不同的网络模型针对不同方向的问题。其中,VGG 模型是最常用来进行人脸检测与识别的模型。与传统的人工神经网络相比,VGG 模型包含更多的隐藏层,特有的卷积和池化操作对人脸的检测和识别有着较高的准确率与效率。

通过分析模型各层提取的特征发现:高层特征是低层特征的进一步抽象,人脸的检测与识别的最终结果主要和经过抽象的高层特征关系比较大。根据这个发现,本文对 Faster R-CNN 中的 RPN 网络模型进行了改进,也将 SPP 引入 VGG 网络模型,基于改进了的 VGG 网络和 RPN 网络,开发和设计了人脸检测与识别软件。最后通过对该软件进行的测试来对本文改进的网络模型的检测与识别效率进行验证。本文的主要研究内容如下:

第一、通过对用于生成候选区域的 RPN 神经网络研究,发现 RPN 通过滑动窗口的方式在最后一层卷积特征图上进行穷举,由于滑动窗口会造成较多的窗口冗余,且窗口大小是固定的,采用固定尺寸图像分割技术代替滑动窗口,用来提升 RPN 网络的性能,将改进的图像分割模块称为 IRPN。

第二、研究学习了用于人脸识别的 VGG 网络,将空间金字塔池化技术引入 VGG 网络模型,即在 VGG 网络模型的卷积层和全连接层之间加一层 SPP,同时,通过分析网络模型中的激活函数 ReLU,在网络模型中引入了新的激活函数,用来提高模型的识别准确率。

第三、根据改进的 RPN 模型与 Fast R-CNN 模型进行人脸检测,和改进 VGG 模型进行人脸识别,利用 OpenCV, Caffe, 以及 Python 设计开发出了人脸检测和识别的软件,用来测试改进的模型的识别准确率和效率,测试表明该软件检测与识别的准确率和效率都比较高。

虽然此软件检测与识别的准确率和效率都比较高,但是在某些方面,本文还存在一些缺陷和不足:

1. 基于 Faster R-CNN 的人脸检测与识别软件的编码实现主要依赖第三方工具

完成,软件整体的代码优化不是很彻底,可以在工程实现上进一步优化,使该软件的运行时间更短,效率更高。

2. 模型的训练比较耗时,故本文的有些对比实验就没有进行。例如,两个不同模型之间共享卷积层参数的训练方式一共有三种,本文采用了轮流训练的方式,也可以采用其他的训练方式与该种方式进行对比,找到最好的训练方法;对于新的 ReLU 函数的对比等。

3. 对于 RPN 的改进,将滑动窗口的策略更改为固定尺寸分割策略,对于固定尺寸的分割可以进行更好的优化,也就是可以进一步地快速确定人脸所在的位置,也就是进行人脸检测。

## 5.2 展望

人脸检测与识别技术经过多年的研究,已经积累了较多的研究成果。传统的人脸检测与识别算法在解决人脸检测与识别问题的同时,往往比较依赖人的先验性知识,所以针对此问题,本文采用深度学习的方法进行人脸检测与识别。一般的,基于深度学习的人脸检测与识别算法会比传统的方法好很多,但是也存在很多问题,尤其是复杂环境下的人脸检测与识别。主要有以下几种情况:

- 1) 同一个人的人脸图像会因为照明条件的变化而发生剧烈的改变;
- 2) 对于人脸正面的图像和其他姿态的人脸图像的检测与识别效果完全不同;
- 3) 人脸的特征会随着时间的推移,年龄的增长而发生变化;
- 4) 图像与摄像头的距离也会对图像的质量有很大的影响,从较远距离拍摄的人脸图像质量会比较低劣,也会有噪音的影响;
- 5) 人脸的脸部可能会存在遮挡,被其他人或物体(如眼镜、帽子等)遮挡。

越来越多的研究表明人脸识别问题有着巨大的应用价值,尤其是以上这些复杂条件下的人脸检测与识别,需要研究人员进行更深一步的研究。目前的人脸检测与识别技术主要有以下的研究发展趋势:

### 1. 三维人脸检测与识别

与二维人脸图像相比,三维的人脸图像包含了人脸本身固有的信息,即人脸空间信息,因而对外界条件的变化具有很好的鲁棒性,并且收集到的三维人脸图像不随光照、姿态等条件的变化而变化。从本质上来说,二维图像是三维物体在二维空间上的投影,因为在投影的过程中损失了很多有利于识别的信息。因此,越来越多的图像处理、模式识别领域的研究人员开始对三维人脸的检测与识别进行研究。

三维的人脸包含更多的有利于人脸检测识别的信息,利用人脸的三维信息进行检测与识别将有助于克服传统的基于二维人脸图像的识别方法所遇到的困难。虽然三维

人脸数据信息量丰富,但如何提取对分类有效的特征进行人脸识别是三维人脸识别的关键研究内容,也是首要解决的问题。

## 2. 人脸的动态追踪识别

随着计算机软件和硬件的发展和应用需求的增长,基于人脸图像的检测与识别不再满足需求,需要对火车站、飞机场和海关等人口密集的地方进行实时监控的同时,也需要对监控视频中的人脸进行动态地检测、追踪和识别。对动态人脸的检测与识别研究具有十分重要的价值和意义,也有着很好的应用前景。



## 参考文献

- [1] 张翠平,苏光大.人脸识别技术综述[J].中国图象图形学报: A 辑, 2000 (11): 885-894.
- [2] 孙志军,薛磊,许阳明,等.深度学习研究综述[J]. 计算机应用研究,2012, 29(8): 2806-2810.
- [3] 梁路宏,艾海舟,徐光祐,等.人脸检测研究综述[J]. 计算机学报, 2002, 25(5): 449-458.
- [4] SUYKENS J A K, VANDEWALLE J. Least squares support vector machine classifiers[J]. Neural processing letters, 1999, 9(3): 293-300.
- [5] JOACHIMS T. Text categorization with support vector machines: Learning with many relevant features[J]. Machine learning: ECML-98, 1998: 137-142.
- [6] HSU C W, Chang C C, Lin C J. A practical guide to support vector classification[J]. 2003.
- [7] BROWN M P S, Grundy W N, Lin D, et al. Knowledge-based analysis of microarray gene expression data by using support vector machines[J]. Proceedings of the National Academy of Sciences, 2000, 97(1): 262-267.
- [8] CHANG C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.
- [9] SMOLA A J, SCHÖLKOPF B. A tutorial on support vector regression[J]. Statistics and computing, 2004, 14(3): 199-222.
- [10] REN S, He K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
- [11] ZHANG L, CHU R, XIANG S, et al. Face detection based on multi-block lbp representation[J]. Advances in biometrics, 2007: 11-18.
- [12] AHONEN T, HADID A, PIETIKAINEN M. Face description with local binary patterns: Application to face recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2006, 28(12): 2037-2041.
- [13] JIN H, LIU Q, LU H, et al. Face detection using improved LBP under Bayesian framework[C]//Image and Graphics (ICIG'04), Third International Conference on. IEEE, 2004: 306-309.
- [14] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors[J]. Cognitive modeling, 1988, 5(3): 1.
- [15] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.
- [16] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning internal representations by error

- p propagation[R]. California Univ San Diego La Jolla Inst for Cognitive Science, 1985.
- [17] YANG J, ZHANG D, FRANGI A F, et al. Two-dimensional PCA: a new approach to appearance-based face representation and recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2004, 26(1): 131-137.
  - [18] HE K, ZHANG X, REN S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.
  - [19] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3): 211-252.
  - [20] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.
  - [21] RASTEGARI M, ORDONEZ V, REDMON J, et al. Xnor-net: Imagenet classification using binary convolutional neural networks[C]//European Conference on Computer Vision. Springer International Publishing, 2016: 525-542.
  - [22] OUYANG W, WANG X, ZENG X, et al. Deepid-net: Deformable deep convolutional neural networks for object detection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 2403-2412.
  - [23] OUYANG W, LUO P, ZENG X, et al. Deepid-net: multi-stage and deformable deep convolutional neural networks for object detection[J]. arXiv preprint arXiv:1409.3505, 2014.
  - [24] SUN Y, LIANG D, WANG X, et al. Deepid3: Face recognition with very deep neural networks[J]. arXiv preprint arXiv:1502.00873, 2015.
  - [25] YANG G, HUANG T S. Human face detection in a complex background[J]. Pattern recognition, 1994, 27(1): 53-63.
  - [26] 梁路宏,艾海舟,肖习攀,等.基于模板匹配与支持矢量机的人脸检测[J]. 计算机学报, 2002, 25(1): 22-29.
  - [27] 梁路宏,艾海舟,徐光佑,等.基于模板匹配与人工神经网络确认的人脸检测[J]. 电子学报, 2001, 29(6): 744-747.
  - [28] VAN O C, HUTCHINSON W F, WILLS D P M, et al. MICRO - CHECKER: software for identifying and correcting genotyping errors in microsatellite data[J]. Molecular Ecology Notes, 2004, 4(3): 535-538.
  - [29] 阎平凡,张长水.人工神经网络与模拟进化计算[M].清华大学出版社有限公司, 2005.
  - [30] ALEXANDER C, ISHIKAWA S, SILVERSTEIN M, et al. A pattern language[M]. Gustavo Gili, 1977.
  - [31] HAN S, POOL J, TRAN J, et al. Learning both weights and connections for efficient neural



- network[C]//Advances in Neural Information Processing Systems. 2015: 1135-1143.
- [32] REN S, HE K, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.
- [33] GIRSHICK R. Fast r-cnn[C]//Proceedings of the IEEE International Conference on Computer Vision. 2015: 1440-1448.
- [34] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
- [35] 刘曙光,郑崇勋,刘明远.前馈神经网络中的反向传播算法及其改进:进展与展望[J]. 计算机科学, 1996, 23(1): 76-79.
- [36] CHUA L O, ROSKA T. The CNN paradigm[J]. IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, 1993, 40(3): 147-156.
- [37] ROSKA T, CHUA L O. The CNN universal machine: an analogic array computer[J]. IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing, 1993, 40(3): 163-173.
- [38] SHARIF R A, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: an astounding baseline for recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2014: 806-813.
- [39] CHUA L O. CNN: A paradigm for complexity[M]. World Scientific, 1998.
- [40] WEST D. Neural network credit scoring models[J]. Computers & Operations Research, 2000, 27(11): 1131-1152.
- [41] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.
- [42] BRIDLE J S. Probabilistic interpretation of feedforward classification network outputs, with relationships to statistical pattern recognition[M]//Neurocomputing. Springer Berlin Heidelberg, 1990: 227-236.
- [43] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[C]//European Conference on Computer Vision. Springer International Publishing, 2014: 346-361.
- [44] JIA Y, SHELHAMER E, DONAHUE J, et al. Caffe: Convolutional architecture for Fast feature embedding[C]//Proceedings of the 22nd ACM international conference on Multimedia. ACM, 2014: 675-678.
- [45] HINTON G E, OSINDERO S, TEH Y W. A Fast learning algorithm for deep belief nets[J]. Neural computation, 2006, 18(7): 1527-1554.

- [46] BERGSTRA J, BREULEUX O, BASTIEN F, et al. Theano: A CPU and GPU math compiler in Python[C]//Proc. 9th Python in Science Conf. 2010: 1-7.
- [47] BASTIEN F, LAMBLIN P, PASCANU R, et al. Theano: new features and speed improvements[J]. arXiv preprint arXiv:1211.5590, 2012.
- [48] ABADI M, AGARWAL A, BARHAM P, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems[J]. arXiv preprint arXiv:1603.04467, 2016.
- [49] ABADI M, BARHAM P, CHEN J, et al. TensorFlow: A system for large-scale machine learning[C]//Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI). Savannah, Georgia, USA. 2016.
- [50] BERGSTRA J, BREULEUX O, BASTIEN F, et al. Theano: A CPU and GPU math compiler in Python[C]//Proc. 9th Python in Science Conf. 2010: 1-7.

## 致谢

三年时间转瞬即逝，在西安电子科技大学学习的三年时间，赋予我人生浓墨重彩的一笔，知识的储备以及思维的丰盈，都让我感觉这三年学习生活的充实。

感谢尊敬的导师-----苗启广老师，毫无保留的传道、授业、解惑，让我在学习的过程中充分享受到学习的愉悦和知识的力量。老师的辛勤耕耘和谆谆教导让我受益匪浅，从论文的选题、结构、成文、到反复的修改完善，都倾注了老师大量的心血和精力，并提出很多宝贵的意见和建议。苗老师渊博的知识，严谨的治学态度，以及负责任的态度和真诚坦荡的处世之道，让我敬仰佩服。在此特别向我的导师苗启广老师表示我最真挚的谢意和最美好的祝福。借此论文答辩之际，要表示我最忠诚和崇高的感谢和敬意。

也要感谢与我三年同窗的同学，以及在求学期间，他们曾为我提供的帮助，才能让我不断地进步前行，更重要的收获是我们三年之间深厚的同学情谊。

另外，对本文写作中所参考和引用的国内外研究成果、著作、文献的作者表示最真挚的感谢。

因为才疏学浅，以及掌握查询资料的不足，在论文的写作中难免有疏漏和不足，还希望各位专家学者老师提出批评指正，让我能够不断完善进步。

感谢西安电子科技大学，感谢各位老师同学，感谢我的家人，因为有你们，我的人生才完整丰富，我将秉承西电的校训和各位老师的教诲，走向更美好的未来。



## 作者简介

### 1. 基本情况

尉冰，女，山西运城人，1991 年 7 月出生，西安电子科技大学计算机学院计算机技术专业 2014 级硕士研究生。

### 2. 教育背景

2010.09~2014.06 重庆三峡学院，本科，专业：软件工程

2014.09~2017.06 西安电子科技大学，硕士研究生，专业：计算机技术



西安电子科技大学  
XIDIAN UNIVERSITY

地址：西安市太白南路2号

邮编：710071

网址：[www.xidian.edu.cn](http://www.xidian.edu.cn)