

论文类型与性质	名称	中文	基于深度学习的人脸识别系统设计与实现
		英文	Design and implementation of face recognition system based on deep learning
	类别	1. 技术攻关研究                      2. 工程项目策划 3. 工程设计或技术改造   ✓      4. 新工艺、新材料、新设备、新产品的研制与开发	
	形式	1. 工程设计   ✓ 2. 研究论文	
论文内容和意义	性质	1. 应用技术研究   ✓ 2. 技术开发	
	摘要	<p>当前，人脸识别技术飞速发展，日趋成熟，越来越广泛的应用于人们的日常生活中。但是在复杂环境中，人脸识别系统由于难以提取具有鉴别性的特征，以及过于关注全局特征而忽视局部语义特征的重要性，仍然存在人脸误检，导致准确性无法很好满足实际应用需求。本文引入人脸语义分割技术，提出基于语义特征的人脸误检筛选方法和基于全局与局部语义特征相融合的人脸识别算法，降低人脸误检率，增强人脸识别准确鲁棒性。设计并实现人脸识别系统，包含登录模块、管理模块、图像上传模块以及识别算法模块，并应用到具体的人脸识别业务场景中。</p>	
	主题词	主题词数量不多于三个，主题词之间空一格（英文用“/”分隔）	
		中文	人脸识别 深度学习 图像处理
		英文	Face recognition/ Deep learning/ Image processing

## 二、选题依据

1. 阐述选题背景，研究意义，或工程设计的价值和意义，该选题国内外技术应用现状和发展趋势，选题的先进性和实用性。

### 1.1 选题背景

随着现代信息技术的不断发展，人们对安全保障需求逐步增大，从以前平安城市的建设到现今智慧城市的建设，安防技术越来越体现出其不可或缺的重要性，相应的，基于生物特征的识别技术也受到人们越来越多的关注，而人脸识别技术便是其中最具代表性的，一直以来都是计算机视觉领域的研究热点。人脸识别技术是一种基于生物生理特征的人工智能识别技术，通过提取分析一个人的面部特征来进行智能识别<sup>[1]</sup>。

相比其他生物识别技术，如指纹识别、掌纹识别、静脉识别、虹膜识别和声纹识别，人脸识别技术因其独有的非侵入性而表现出更大的吸引力<sup>[2]</sup>。众所周知，指纹识别要求用户将手指放入传感器中，虹膜识别要求用户眼球近距离接近采集相机，声纹识别要求用户对着采集设备大声说话。相反，人脸识别仅要求用户出现在相机的视野内，这使得人脸识别成为最人性化的生物特征识别方式。不仅如此，人脸识别的潜在应用范围更广，因为它可以部署在不希望用户与系统合作的环境中，如监控系统。经过几十年的研究发展，现代人脸识别技术更具有精度高、鲁棒性好、易使用、难以假冒、性价比高的优点，因此人脸识别在人们的日常生活中得到了广泛的应用，特别是在系统安全、信息安全、人机交互方面。人脸识别的常见应用包括访问控制、欺诈检测、身份验证和社交媒体等。

人脸识别对于人类来说就是一项可以日常进行的简单任务，但是构建一个类似的计算机系统并不断提高其准确性却任重而道远。关于人脸识别的研究最早可以追溯到上世纪 70 年代，迄今为止，人脸识别技术也日趋成熟，各种新的算法不断涌现，研究者们也取得了丰硕的成果，尤其是随着深度学习的出现，人脸识别技术得到了飞速发展。但是，人脸识别技术仍然面临许多挑战，并且受到实际复杂应用环境的限制。这些限制不仅来自人脸姿态不一、面部表情多样、人脸遮挡、化妆、年龄变化、识别规模庞大等各种因素的影响，还来自图片像素质量低、光照变化、背景复杂等因素的影响，除此之外，人脸识别技术还受到数据大小和数据分布的影响。因此，现有面向复杂环境的人脸识别算法准确性、鲁棒性仍然有待提升，无法很好满足实际工

业应用场景的落地要求，而如何进一步提高复杂环境下人脸识别算法的准确率、运行效率和鲁棒性便显得更加必要、紧迫和意义重大。

## 1.2 工程设计价值意义

华为技术有限公司目前自研投入使用的人脸识别算法中应用的相关技术长时间未更新，使用的 yolov3 等技术在算法的更新迭代过程中已经渐显颓势，无法满足性能需求以及在复杂场景下的人脸识别的准确度要求，比如疫情当下人脸戴口罩的场景下，现投入使用的人脸识别算法无法满足在有遮挡条件下的人脸准确识别以及在监控条件下多个人脸的检测以及识别准确度需求，所以亟需研发性能更好且可以在复杂场景下完成识别需求的人脸识别算法，投入到例如门禁场景、监控场景以及智慧会议等多种场景的人脸识别应用中。未来也可在人脸活体检测(人脸防伪)等方向上作为前置技术进行拓展与研究。

与业界中应用的人脸识别技术不同的是，本文提出的基于语义特征的人脸误检筛选方法，利用基于人脸语义分割结果构造出的表征能力更强的人脸语义特征，结合 SVM 分类器实现人脸误检分类筛选。以及基于全局与局部语义特征结合的人脸识别算法，通过人脸分割技术显式的引入了人脸局部语义特征，然后结合局部遮挡信号采用特征融合的方式得到预期表征能力更强、鲁棒性更好的融合特征，以用于最后的人脸对比。

## 1.3 国内外研究现状

人脸识别整体流程主要可分为四个部分，如图 1 所示:人脸检测、人脸预处理、特征提取和人脸比对<sup>[3]</sup>。诚然，这其中每个部分都与人脸识别系统的运行结果好坏息息相关，业界通常指的狭义的人脸识别主要包括后三个部分，所以本小节主要从人脸检测和人脸识别两个方面来介绍与之相关的发展历史和国内外研究现状。

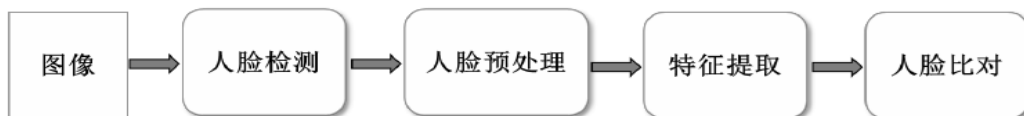


图 1 人脸识别整体流程图

### 1.3.1 人脸检测的国内外研究现状

人脸目标检测是整个人脸识别整体流程中的第一步，也是计算机视觉领域被深入研究的经典问题之一，在安防监控、人证核验、人机交互、社交等领域都有重要的应用价值。人脸检测的目的是寻找到给定图像中所有的人脸，并标出其位置，可视化结果一般多采用矩形框的形式在原图框出人脸。一个优秀的人脸检测算法应该具有高检测率和低误检率，除此之外，实际工程中更提出了实时性要求。

人脸检测方法大致可以分为三类：基于 Boosting<sup>[4]</sup>的方法、基于可变形组件模型 (Deformable Part Model , DPM)<sup>[5]</sup>的检测方法和基于深度学习的方法。

### （1）基于 Boosting 的方法

Viola 和 Jones (VJ)<sup>[6]</sup>首先提出使用积分图像的表示方法快速提取特征，并使用 AdaBoost<sup>[7]</sup>算法训练弱分类器，最后得到级联的人脸分类方法，该方法能快速剔除容易被忽略的负面信息，同时在更多类似人脸的样本上花费更多的计算时间。正是由于该方法的高效率，使得人脸检测开始应用于很多实时应用中。然而，后续的研究发现，在真实的无约束场景下，VJ 人脸检测器不能有效检测具有较大姿态变化和遮挡的人脸<sup>[8]</sup>。类似于 VJ 的框架，Mathias 等人<sup>[9]</sup>在 Boosting 的基础上引入积分通道特征，该特征本质上是基于 22 个刚性模板的多尺度检测器模型，对于每个尺度，都有 11 个模板，其中 5 个用于正面，6 个用于旋转面，进一步提升了检测效果和速度。与前面的方法不同，Liao 等人<sup>[10]</sup>提出了一种叫做标准化像素差 (Normalised Pixel Difference , NPD) 的新特征，利用深度二叉树对该特征进行学习和组合，然后采用单一的软级联增强分类器对特征树进行学习，而不依赖于特定的级联结构或位姿标记，这使得该检测器的速度和鲁棒性表现更优异。

### （2）基于可变形组件模型的检测方法

传统方法中，除了上面基于 Boosting 的方法外，还有基于 DPM 的人脸检测方法<sup>[11]</sup>。DPM 由 Felzenszwalb 于 2008 年提出，因为其本身基于组件的特性，可模拟人脸组件之间的潜在形变，所以对扭曲、性别、多姿态、多角度等人脸都具有较好的检测效果。基于 DPM 的方法在很多开放数据集上都取得了比 VJ 方法更好的效果，但是该模型计算过于复杂，很难满足实时性要求，因此在当时工程中更多选用基于 Boosting 的方法。

### （3）基于深度学习的方法

传统方法中基于经典的人工设计特征本身并不稳定，非常容易受外界环境如光照、角度、遮挡等的影响，所以在复杂场景下的人脸检测性能难以得到保证，无法很好应用于复杂场景中。深度学习出现以后，深度卷积神经网络（Deep Convolutional Neural Network, DCNN）能够很好的学习到图像中目标各个层级的特征，对外界的抗干扰能力更强，因此后续的人脸检测方法基本都是基于深度学习的方法。近年来，随着 R-CNN 系列<sup>[12-16]</sup>和 YOLO 系列<sup>[17-19]</sup>的出现与发展，基于深度学习的方法在目标检测领域中取得了突破性成果，表现出了卓越的性能。当然类似的方法思想也在第一时间应用到人脸检测方面，并取得优异的效果。一般说来，基于深度学习的人脸检测算法主要包含以下几类：多尺度的人脸检测算法、两阶段的人脸检测算法和单阶段的人脸检测算法。

多尺度的人脸检测算法可以检测多尺度的人脸，主要通过图像金字塔或者特征金字塔来实现。其中的典型代表便是 MTCNN<sup>[20]</sup>。

MTCNN 全称多任务级联卷积神经网络 (Multi-task Cascaded Convolutional Networks)，由中国科学院深圳研究院于 2016 年提出，用于人脸区域检测与人脸关键点检测，其检测效果达到了当时最优级别，其运算速度也满足了实时要求，因此被大量应用于实时人脸识别相关工程项目中。该模型框架中主要采用级联三个卷积网络的结构，利用候选框推荐加分类的思想，实现快速高效的人脸检测。这三个级联的网络分别是快速生成候选窗口的 P-Net (Proposal Network)、进行高精度候选窗口过滤选择的 R-Net (Refine Network) 和生成最终边界框与人脸关键点的 O-Net (Output Network)。

图 2 为 MTCNN 算法示意图，首先将给定的输入图片缩放到不同尺度，形成图像金字塔，以达到检测多尺度范围人脸的目的，然后便将图像金字塔通过 MTCNN 的三大主体网络，其三大网络结构如图 3 所示。第一步通过 P-Net，P-Net 是一个全卷积神经网络，对人脸区域进行初步建议，该网络会获取到大量人脸候选框和人脸关键点，同时，候选窗体还会根据边框回归进行校准；第二步是通过 R-Net，将经过 P-Net 所确定的包含候选框的图片在 R-Net 网络中训练，利用边框向量微调候选框，该网络会过滤掉大量的结果较差的候选框，以达到候选框细化选择和人脸区域优化的效果；第三步是通过 O-Net，O-Net 的网络结构比 R-Net 多一层卷积，其功能与 R-Net 作用类似，去除低质量和重叠候选窗口，同时优化五个人脸关键点坐标信息，最终输

出人脸检测结果和人脸关键点。不仅如此，在这三步中，每一步后面都采用了非极大值抑制来去除重叠候选窗，以获得更加精细准确的检测结果。



图 2 MTCNN 算法示意图

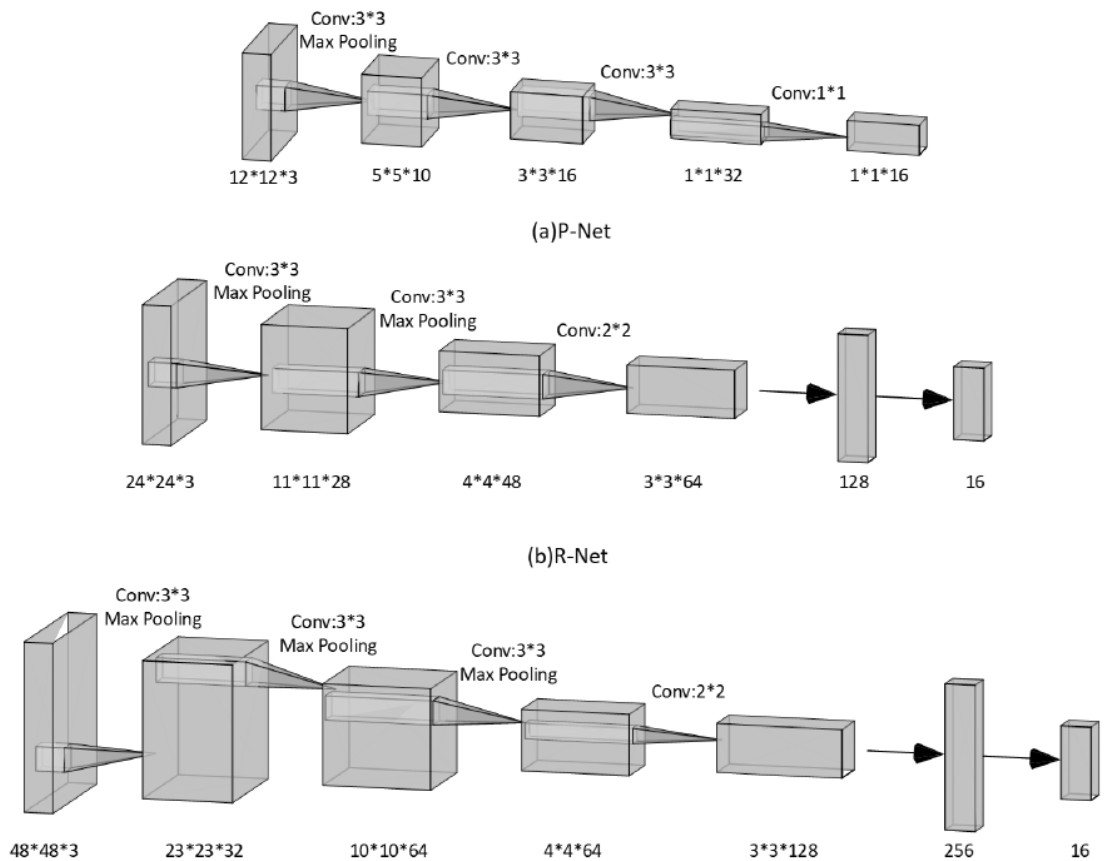


图 3 P-Net、R-Net、O-Net 网络结构图 (a)P-Net (b)R-Net (c)O-Net

相比传统算法，MTCNN 兼顾了运行性能和检测准确率，避免了传统滑动窗口加分类器思路所造成的巨大的性能消耗，使用三层级联网络，先通过较小的网络生成一批人脸目标区域候选框，然后再通过更深的网络进行更高精度的边框回归和关键点预测，最终实现了快速高效的人脸检测。除此之外，随人脸检测一同提取到的五点坐标也给后续的人脸对齐任务提供了简单的

方案。但实际运行过程中,为了满足实时要求,牺牲了检测准确度,因而存在一定的误检和漏检。

两阶段的人脸检测算法在流程上主要分为两步,第一步进行人脸候选框推荐,第二步根据候选框进行人脸目标分类与边框回归。两阶段的人脸检测算法一般具有精度高的优点,但也同时具有速度慢的缺点。其中的典型代表便是 FaceR-CNN<sup>[22]</sup>。

Face R-CNN 是一种基于通用目标检测算法 Faster R-CNN<sup>5'</sup> 的深度人脸检测算法。Faster R-CNN 一经出现便在全世界目标检测领域引起轰动,是最具代表性和最成功的目标检测方法之一,在各种具体目标检测应用中也越来越受欢迎,这其中当然也包括人脸检测应用。Face R-CNN 在 Faster R-CNN 的基础上结合人脸识别实际情况采取了三点改进创新:(1)在 Faster R-CNN 的架构上增加一个新的多任务损失函数 Center Loss 来对人脸分类进行辅助训练。(2)使用一种新的在线难例挖掘(Online Hard Example Mining, OHEM)策略获取困难样本以供后续增强训练。(3)采用多尺度的训练方式提升算法检测多尺度目标的性能。最终得到的人脸检测算法也体现出了 Faster R-CNN 的优越性,在多个最具挑战的人脸检测基准上都获得了优异的成绩。

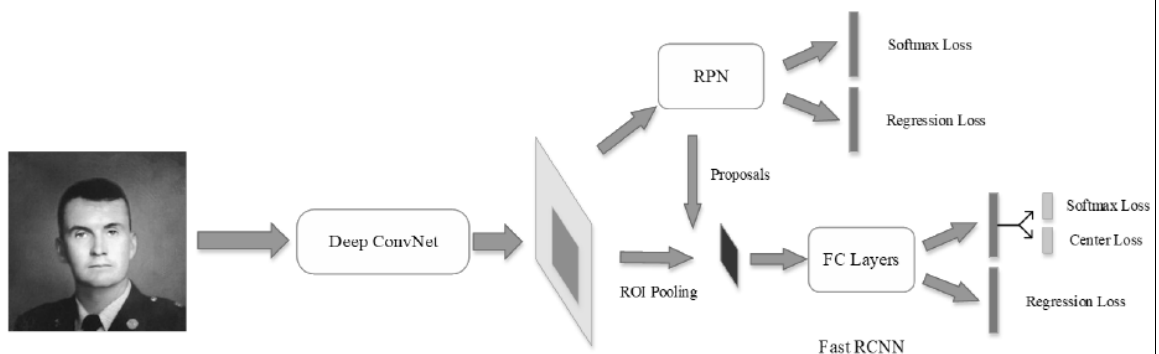


图 4 Face R-CNN 网络结构

Face R-CNN 在网络架构上沿用 Faster R-CNN 的设计,如图 4 所示,该结构中主要包含一个 Deep ConvNet,一个区域生成网络 RPN,和一个 Fast RCNN 模块。Deep ConvNet 是一个深度特征提取网络,用于提取图像深度特征,RPN 模块是建立在特征图上的一个全卷积网络,生

成一系列的目标矩形区域候选框, FastRCNN 模块对生成的区域候选框进行 ROI Pooling 操作, 得到固定大小的特征图, 然后通过两个分支的全连接层进行人脸分类和边框回归训练。

Face R-CNN 提到此前的模型大多使用 Softmax Loss 进行分类训练, 但 Softmax Loss 只关注类间差异, 忽略了类内差异, 而 Center Loss 通过最小化样本与它所属类别中心之间的距离, 可以使同类样本所学到的特征更加集中在一起, 达到进一步增大类间差异的目的。而且, Center Loss 支持与 Softmax Loss 联合优化, 利用 Softmax Loss 在最大化类间差异上的优势, 配合 Center Loss 减小类内差异的能力, 最终联合训练共同提升特征的辨识度。

OHEM 机制是一个简单而十分有效的引导技术。其关键的思路是在训练过程中收集硬样本, 即那些预测不正确的难样本, 然后在下次训练迭代时将这些样本输送给网络以增强网络分辨能力。又因为损失可以表示当前网络对样本的拟合程度, 所以可以通过样本的损失将生成的候选框进行排序, 然后只提取前  $N$  个损失最大的样本作为困难样本。

但标准的 OHEM 会遇到数据不平衡的问题, 因为在选择的困难样本集合中, 负样本的数量很可能压倒性的超过正样本的数量, 反正亦然, 而且保持正负训练样本的平衡对训练阶段极其重要。所以, Face R-CNN 在正样本和负样本上分别独立使用 OHEM 机制, 并且在每个 mini-batch 中保持正负样本的比例为 1 比 1。

除此之外, Face R-CNN 还采取了一个多尺度训练方法, 即在训练过程中, 将原始图片缩放到不同尺寸, 这样可以促使学到的模型更适应多尺度人脸, 在预测阶段, 也会进行对应的多尺度预测, 然后将不同尺度下的预测结果结合到最终的输出中。

Face R-CNN 延续了 Faster R-CNN 精度高的优点, 还能适应多尺度人脸检测, 但同样也保留了两阶段算法速度慢的缺点, 外加图像金字塔的运行消耗, 因此无法很好的适用于实时人脸检测场景。

单阶段的人脸检测算法与两阶段人脸检测算法不同, 其直接从分类网络的早期卷积层检测人脸目标。单阶段的人脸检测算法一般速度更快, 但在较大尺度和微小尺度下的检测精度偏低。其中的典型代表便是 SSH<sup>[21]</sup>。

SSH (Single stage Headless Face Detector) 是一个单阶段的人脸检测算法, 于 2017 年 8 月提出, 在当时也取得了人脸检测领域顶尖效果。SSH 将许多通用目标检测的方法思路应



用于人脸检测领域，其目的是设计出一个检测速度快、内存占用小、检测精度高的多尺度人脸检测算法。SSH 采用全卷积网络，删去了包含大量参数的全连接层，除此之外，SSH 在设计上是尺度不变的，不依靠图像金字塔来检测各种尺度的人脸，而在网络模型中利用特征金字塔通过不同检测分支来实现多尺度人脸检测，这些设计最终使得 SSH 在检测效果比较精确的同时得以保证轻量而快速。

SSH 的主干网络使用的是 VGG-16，在 VGG-16 的基础上删除了最后的全连接层，并在三个分支上使用不同的检测模块 M1、M2、M3 分别从不同特征层上检测小、中、大尺度的人脸，其网络模型结构如图 5 所示。M1 模块的步长为 8，负责检测小尺度人脸，先分别对 VGG-16 中的 conv4-3 网络层和 conv5-3 网络层的输出使用  $1 \times 1$  的卷积进行降维，减少运算量，再将 conv5-3 的特征图上采样后与 conv4-3 特征图叠加，得到兼具高纬度和高分辨率的融合特征用于检测小尺度人脸。M2 模块的步长为 16，直接使用 conv5-3 网络层的特征图作为输入，其视野更广，用于检测中尺度人脸。M3 模块通过 max pooling 操作，达到步长从 16 到 32 的变化，用于检测大尺度人脸。

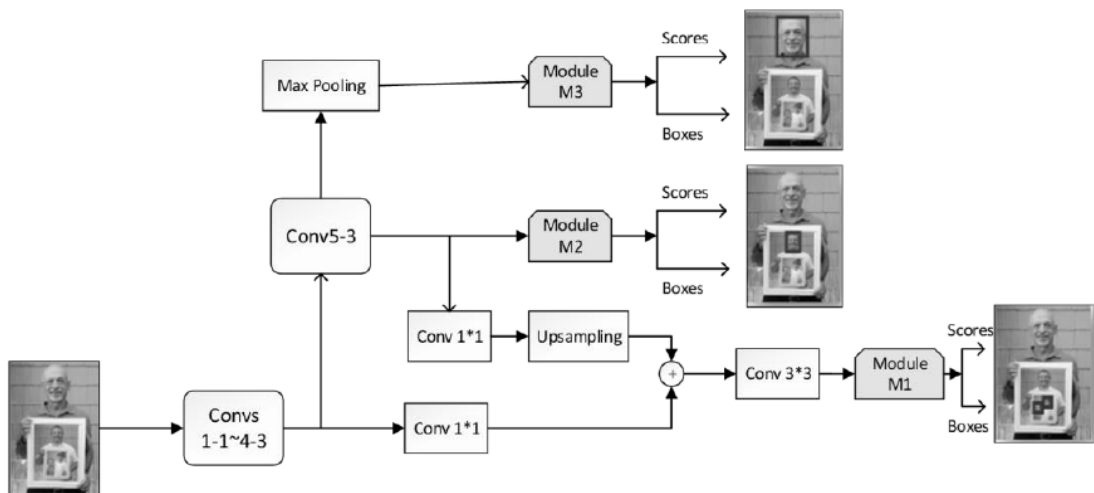


图 5 SSH 网络结构模型

SSH 凭借其单阶段检测的设计，拥有检测速度快、内存占用小的优点，其在正常尺度下也具有较高的检测准确率，比较符合实时检测需求。但相比成熟的两阶段算法，其在较大尺度和

微小尺度下的检测性能偏低，不仅如此，SSH 没有提取人脸五点坐标，因此后续需要使用专门的人脸对齐算法，增加了整体流程的时间开销。

除此之外，还有一些其他的基于深度学习的人脸检测算法<sup>[23-27]</sup>，这些算法的出现大大提高了人脸检测的效果，极大的推进了人脸检测技术在具体工业应用中落地的进程。

### 1.3.2 人脸识别的国内外研究现状

最早的人脸识别研究主要集中于使用图像处理技术来匹配描述人脸几何形状的简单特征，统称为基于几何特征的方法，它们开启了使用计算机进行自动人脸识别的先河。此后，诸如主成分分析 (Principal Component Analysis, PCA)<sup>[28]</sup> 和线性判别分析 (Linear Discriminant Analysis, LDA)<sup>[29]</sup> 之类的统计子空间方法开始得到普及，这些方法被称为基于整体的方法。同时，局部特征提取器的发展促进了基于局部特征的人脸识别方法的出现，而后，基于整体和基于局部特征的方法通过优势互补，被组合为混合方法。直到最近，深度学习技术高速发展并普及，成为包括人脸识别在内的大多数计算机视觉应用的主流方法，也将人脸识别技术推向了一个新的阶段。

七十年代初的 Kelly<sup>[30]</sup> 和 Kanade<sup>[31]</sup> 博士学位论文被认为是有关人脸识别的第一项研究工作，他们提出使用专用的边缘和轮廓检测器来描述每个人脸的特征矢量，并进行人脸匹配。而后，Shi 等人<sup>[32]</sup> 提出了一种基于测量两组面部标志之间的 Procrustes 距离<sup>[33]</sup> 的方法，以及一种测量面部标志之间的距离比率的方法。这些基于几何特征的方法有着速度快、存储需求小、实现简单的优点，但该类方法只对人脸主要器官的基本形状与结构关系进行了简单描述，忽略了大量的图像细节，因而识别率低下。

基于整体的方法使用整个人脸区域图像进行输入。Sirovich 等人<sup>[34]</sup> 和 Kirby 等人<sup>[35]</sup> 最初提出将 PCA 应用于人脸图像的训练，以便寻找在数据分布中差异最大的特征向量。Belhumeur 等人<sup>[36]</sup> 和 Etemad 等人<sup>[37]</sup> 提出了使用 LDA 的整体方法，也称为 Fisher 判别分析法<sup>[38]</sup>。在这种情况下，特征向量由于与真实面孔的相似性而通常被称为特征脸，Moghaddam 等人<sup>[39]</sup> 还提出了一种基于贝叶斯图像差异分析的方法，该方法中使用了两组特征脸进行人脸建模。基于支持向量机 (Support Vector Machines, SVM) 的方法<sup>[40]</sup> 也属于整体方法之一，SVM 让目标问题在高维

空间中变得线性可分。另一个流行的整体方法是基于人脸的稀疏表示方法，Wright 等人<sup>[41]</sup>提出使用训练图像的线性组合来表示人脸。基于整体的方法对于人脸识别的研究发展至关重要，但整体方法只考虑人脸全局特征，因而难以处理人脸局部变化的情况，而且受限于传统方法提取特征的表示局限性，所以识别率并不高。

基于局部特征的方法是指利用在人脸图像中不同位置提取的局部特征进行人脸识别的方法。第一种基于局部特征的方法是模块化特征脸方法<sup>[42]</sup>，该方法将 PCA 独立应用于人脸图像中的不同局部区域，以产生特征集。另一种非常流行的基于局部特征的方法是弹性约束图匹配 (Elastic Bunch Graph Matching, EBGM) 方法<sup>[43]</sup>，这种方法使用节点图来表示人脸，每个节点包含了在一组预定义人脸标志周围提取的 Gabor 小波系数<sup>[44]</sup>。随着计算机视觉应用中局部特征描述符的发展<sup>[45]</sup>，Ahonen 等人<sup>[46]</sup>从人脸局部区域中独立提取出 LBP (Local Binary Patterns) 描述符的直方图特征，Bicego 等人<sup>[47]</sup>提出在整个人脸图像上匹配 SIFT (Scale Invariant Feature Transform) 描述符<sup>[48]</sup>的方法。与基于整体的方法相比，基于局部特征的方法显示出对不同类型变化人脸匹配更强的适应性，但是，也丢失了基于整体方法的某些优点，因此便出现了结合这两种方法的混合方法。

混合方法结合了基于整体和基于局部特征的方法。Liu 等人<sup>[49, 50]</sup>提出了几种结合 Gabor 小波特征和不同统计子空间方法的混合方法，表现出更高的准确性。Tan 等人<sup>[51]</sup>提出使用不同局部特征进行组合以捕获互补信息。混合方法结合了最好的基于整体和基于局部特征的方法，其中一些方法组合了不同类型的特征，而另一些方法则引入了学习阶段以提高特征的判别能力。但传统算法其固有的在特征层面上的局限性导致了这些方法都难以在人脸识别这种复杂问题上取得优秀的结果，尤其是在开放的大规模人脸识别场景中，其识别效果更是远远达不到要求。

随着计算机硬件设备计算能力的大幅增强，深度学习技术飞速发展，并在计算机各大领域取得突破性成果，人脸识别领域也不例外。采用深度卷积神经网络，使用卷积操作大幅缩减参数，不需要人为进行特征设计，而更加关注网络自身的特征学习能力，通常需要在更加庞大的数据集上进行训练学习，以抽象出更具表征能力的高维特征，该方法利用大数据数据驱动的机器自主表示学习能力，大大提升了人脸识别的准确率。Huang 等人<sup>[52]</sup>于 2012 年首次在人脸识

别测试基准 LFW<sup>[53]</sup>上使用了深度学习方法并取得不错的成绩,紧接着,香港中文大学的团队提出 DeepID 算法<sup>[54]</sup>在 LFW 上测得 0.9745 的准确率,Facebook 的团队提出 DeepFace 算法在 LFW 上取得 0.9735 的准确率,它们远远超过传统人脸识别算法的效果。此后,传统人脸识别经验式驱动的模式在深度学习技术的迅速普及下被彻底打破了,人脸识别领域开始进入以数据驱动的深度学习时代,其主要技术路线也由原本的人工特征设计转变成基于深度学习的端到端自主特征学习。

深度学习人脸识别中的特征提取依赖于深度卷积神经网络,其目的是通过大量的数据自动学习到泛化能力强和区分能力高的人脸特征,目前这主要取决于三个方面:训练数据集、卷积网络结构和损失函数。训练数据集是深度学习的基础,近年来越来越多的大型人脸数据集被公开,如 CASIA-WebFace 和 MS-Celeb-1M 等,给深度学习人脸识别提供了数据保障。优秀的卷积神经网络结构有利于学习到表征能力更强的人脸特征,在网络结构设计上,人脸识别领域也沿用各种经典的卷积神经网络进行特征提取,如 AlexNet、VGGNet、GoogLeNet 和 ResNet 等,除此之外,还有结合人脸识别具体情况在经典网络上进行改进的网络架构,如 DeepFace<sup>[55]</sup>和 DeepID 系列。损失函数是规范特征学习的关键,在人脸识别网络的训练中,损失函数的设计主要分为两大类:一类是基于欧氏距离的度量学习损失,包括 Contrastive Loss、Triplet Loss 和 Center Loss 等;另一类是基于 Margin (间隔)的分类损失,具体来说,包括交叉熵损失 Softmax Loss 以及在其之上的改进损失,如 L-Soft max Loss、A-Softmax Loss、Cos-SoftmaxLoss 和 Arc-Softmax Loss 等。

在网络结构层面,随着 AlexNet 大获成功,深度学习开始应用于计算机视觉的各个领域,人脸识别也不例外。2014 年,DeepFace 在对人脸进行对齐后,使用 9 层的卷积神经网络进行训练,并在 LFW 测试集上取得了 97.35%的准确率。同年,GoogLeNet 和 VGGNet 通过增加网络深度有效提升了模型的性能,在 ILSVRC-2014 比赛中分获第一名和第二名。2015 年,FaceNet 使用私有数据集训练 GoogLeNet,并引入了三元组损失,最终在 LFW 测试集上获得了 99.63%的准确率。同年,M. Parkhi 等人发布了大型数据集 VGGFace,并用该数据集训练 VGGNet 网络,最终在 LFW 上达到了 98.95%的准确率。在 2015 年,何恺明等人提出了 ResNet,在卷积神经网络中引入了残差模块,避免了网络过深导致的梯度消失和梯度爆炸问题,使得神经网络可以大

幅度加深，从而获得更好的表达能力。2017 年，Liu 等人使用了修改过的 64 层 ResNet 网络，最终在 LFW 上得到了 99.42% 的准确率。同年，SENet 在 ImageNet 2017 比赛中大放异彩，通过在神经网络中加入 SE（Squeeze-and-Excitation）模块，可以为不同通道设置不同的权重，模型会更加关注信息量大的通道的特征，而抑制不重要的通道的特征，从而提升模型的学习能力。随后的 ArcFace 在改进 ResNet 网络的同时也加入了 SE 模块，并且获得了明确的性能提升。此外，DenseNet 作为新的计算机视觉通用模型，也被逐渐应用于人脸识别领域。在轻量级模型领域，Mobilefacenets 在 MobileNetv2 的基础上进一步针对人脸识别场景进行优化，在 LFW 测试集上得到 99.28% 的准确率。综上所述，目前人脸识别研究中主流的网络结构基本来源于通用的目标识别网络，但最近的工作已经开始针对人脸识别问题的特点自行修改部分网络结构。此外，随着神经网络搜索、注意力机制等新技术的发展，越来越多的工作开始考虑将新技术引入到人脸识别中。

除了网络结构上的优化，近年来人脸识别领域的研究主要集中于损失函数的改进，具体可分为基于深度度量学习的损失函数和基于分类的损失函数两大类。

基于深度度量学习的损失函数的主要思想是，输入由样本对或样本组构成，通过深度网络提取每个样本的特征向量，并通过特征向量计算样本之间的距离，缩小同类样本之间的距离，扩大不同类样本之间的距离，比较典型的损失函数有 contrastive loss 和 triplet loss。contrastive loss 由 LeCun 等提出，它的输入为一对样本，若两样本标签相同，则使样本间的距离尽可能小，否则，使样本间的距离大于设定的阈值。contrastive loss 只对样本间的绝对距离进行了优化，而 tripletloss 选择优化样本间的相对距离。triplet loss 选择锚样本（Anchor）、正样本（Positive）、负样本（Negative）三个样本组成一个三元组，其中锚样本和正样本的标签相同，和负样本的标签不同，优化的目标是使锚样本和正样本在特征空间上尽量靠近，同时使锚样本和负样本在特征空间上尽量远离直到大于设定的阈值。基于深度度量学习的损失函数目标明确，学习效果好，但构建有意义的样本组需要花费大量时间，且存在很多技巧，训练期间稳定性较差。

基于分类的损失函数的主要思想是将人脸识别当作分类问题去处理。最早的 DeepFace 和 DeepID 都采用 Softmax 损失，但其存在的问题是它鼓励不同类别的特征分开，但并不鼓励特

征分离很多。因此,在人脸识别这种细分类问题上,Softmax 损失的性能很难使人满意。Center loss 在此基础上引入了样本到对应类中心的距离作为惩罚项,在 LFW 测试集上的识别准确率达到 99.28%。随后,研究发现人脸特征在流形上具有角度分布特性,因此余弦/角度距离逐渐取代 L2 距离,成为研究的重点。Liu 等人在 Softmax 损失函数上加入乘积性角度间隔,提出了 L-Softmax 损失。A-Softmax 在 L-Softmax 的基础上进一步将权重  $w$  归一化,这样就可以减轻数据分布不均衡对模型性能的影响。但由于 L-Softmax 和 A-Softmax 损失难以收敛,CosFace、ArcFace 分别将乘积性角度间隔改为加性余弦间隔、加性角度间隔,使模型更容易收敛。Wang 等人提出 MV-Softmax 损失,在 ArcFace 的基础上进一步加强了对困难样本的挖掘。Zhang 等人提出 AdaCos 损失,尝试去掉损失函数中的超参数,并利用自适应比例参数在训练过程中自动加强训练监督,以减轻调参的工作量。

人脸识别领域高速发展,其识别技术也日趋成熟,各种新式算法不断涌现,也取得了不小的成果。但是,复杂环境下存在图片像素质量低、光照变化、人脸姿态不一、面部表情多样、人脸遮挡、年龄变化、识别规模庞大等各种影响因素,导致面向复杂环境的人脸识别算法准确鲁棒性不足。因此,通过大量调研并结合现有实际人脸识别系统的运行状态分析,本文发现两大亟待研究优化的问题:(1)在人脸识别整体流程中,现有人脸检测算法在复杂场景下存在部分人脸误检问题,而这对后续的人脸识别带来了极大的负面影响。(2)现有通用人脸识别算法大多都采用深度卷积神经网络直接对人脸图像进行全局特征提取,忽视局部语义特征的重要性,在复杂场景下的准确鲁棒性不足。针对上述两大关键问题,本文的研究目标是减少复杂场景下的人脸误检结果,提高复杂场景下人脸识别算法的准确鲁棒性。

## 1.4 本文选题的先进性与实用性

### 1.4.1 先进性

相比其他生物识别技术,如指纹识别、掌纹识别、静脉识别、虹膜识别和声纹识别,人脸识别技术因其独有的非侵入性而表现出更大的吸引力,同时针对现有人脸识别在复杂场景下的准确性问题,提出基于全局与局部语义特征结合的人脸识别算法,提高在复杂场景下的人脸识别准确性。

### 1.4.2 实用性

#### （1）非接触

对于人脸的照片采集不同于采集指纹信息，采集指纹信息需要接触性专业指纹采集设备，而相对于人脸图像的采集，使用的是摄像头，不需要直接接触。

#### （2）非侵扰

采集人脸照片通常采用的是摄像头自动拍摄，过程自然不需要工作人员指导，也无需被识别别人的配合。

#### （3）直观性突出

人脸识别技术所使用的依据时人的面部图像，而人脸无疑时肉眼能够判别的最直观的信息源，方便人工确认、审计，“以貌取人”符合人的认知规律。

#### （4）基础资料易于获得

采集友好，一个人的面部信息是与生俱来的，像对于指纹和虹膜来说，它的隐私性并不那么强，采集信息方便。

#### （5）简便

人脸采集的设备是摄像头，应用广泛也不需要特别专业的设备，所以简便易操作。

#### （6）快速

从摄像头中采集人脸信息很快，而且由于它的非接触性，使得在采集信息的过程中大大缩短了所需时间。

#### （7）可扩展性好

可以通过以往的视频数据和现有的摄像头及监控设备，在后台应用时可以加入出入控制、来访权限、信息检索等诸多领域，应用前景广阔。

#### （8）不易仿冒

在安全性要求高的应用场合，人脸识别技术要求识别对象必须亲临识别现场，他人难以仿冒。

## 2. 主要参考文献（列出作者、论文名称、期刊名称、出版年月）。

- [1]梁文莉. 基于独立成分分析的人脸识别算法研究[D].西安科技大学,2012.
- [2]许佩冉.人脸识别技术在安保工作中的应用及发展趋势研究[J].智能建筑,2017(12):35-37.
- [3]王雄伟. 基于人脸特征的身份认证技术及在考勤系统中的应用研究[D].江苏大学,2008.
- [4] Kearns M , Valiant L G . Learning Boolean formulae or finite automata is as hard as factoring. 1988.
- [5] Felzenszwalb P, McAllester D . Ramanan D.A discriminatively trained, multiscale, deformable part mode[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition.IEEE,2008:1-8
- [6] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition.cVPR 2001.IEEE,2001,1:I-I
- [7] Freund Y , Re. S . A DECISION-THEORETIC GENERALIZATION OF ON-LINE LEARNING AND AN APPLICATION TO BOOSTING[J]. Journal of Computer and System Sciences, 1997(1):55.
- [8] Klare B F , Klein B , Taborsky E , et al. Pushing the Frontiers of Unconstrained Face Detection and Recognition: IARPA Janus Benchmark A[C]// IEEE Computer Vision and Pattern Recognition (CVPR). IEEE, 2015.1931-1939



- [9] Mathias M , Benenson R , Pedersoli M , et al. Face Detection without Bells and Whistles[C]// ECCV. 2014.720-735
- [10] Liao S , Jain A K , Li S Z . A Fast and Accurate Unconstrained Face Detector[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2016, 38(2):211-223.
- [11] Zhu X , Ramanan D . Face Detection, Pose Estimation, and Landmark Localization in the Wild[C]// Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012.
- [12] Girshick R , Donahue J , Darrell T , et al. rich feature hierarchies for accurate object detection and semantic segmentation tech report (v5)[J]. 2017.580-587
- [13] He K , Zhang X , Ren S , et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition[J]. Pattern Analysis & Machine Intelligence IEEE Transactions on, 2015, 37(9):1904-1916.
- [14] Girshick\T R . fast r-cnn[C]// 2017.1440-1448
- [15] Ren S , He K , Girshick R , et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks[C]// NIPS. 2016.91-99
- [16] He K , Gkioxari G , Dollar P , et al. Mask R-CNN[C]// International Conference on Computer Vision. IEEE Computer Society, 2017.2961-2969
- [17] Susmitha A . Smart Recognition System for Business Predictions (You Only Look Once – V3) Unified, Real-Time Object Detection[J]. 2020.779-788
- [18] Redmon J , Farhadi A . YOLO9000: Better, Faster, Stronger[C]// IEEE Conference on Computer Vision & Pattern Recognition. IEEE, 2017:6517-6525.
- [19] Redmon J , Farhadi A . YOLOv3: An Incremental Improvement[J]. arXiv e-prints, 2018.
- [20] Zhang K , Zhang Z , Li Z , et al. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks[J]. IEEE Signal Processing Letters, 2016, 23(10):1499-1503.
- [21] Najibi M , Samangouei P , Chellappa R , et al. SSH: Single Stage Headless Face Detector[C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.4875-4884

- [22] Hu P , Ramanan D . Finding Tiny Faces[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2017.
- [23] Qin H , Yan J , Xiu L , et al. Joint Training of Cascaded CNN for Face Detection[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016.3456-3465
- [24] Hao Z , Liu Y , Qin H , et al. Scale-Aware Face Detection[C]// Computer Vision & Pattern Recognition. IEEE, 2017:1913-1922.
- [25] Yang S , Xiong Y , Chen C L , et al. Face Detection through Scale-Friendly Deep Convolutional Networks; 10.48550/arXiv.1706.02863[P]. 2017.
- [26] Wang H , Li Z , Ji X , et al. Face R-CNN[J]. 2017.
- [27] Deng J , Guo J , Zhou Y , et al. RetinaFace: Single-stage Dense Face Localisation in the Wild; 10.48550/arXiv.1905.00641[P]. 2019.
- [28] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016, 229-232
- [29] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016, 60-63
- [30]MD Kelly. Visual identification of people by computer[J]. Stanford Ai Project Memo Ai, 1973.
- [31] Kanade T . Picture Processing by Computer[J]. Science, 1969, 164(3875):19-29.
- [32] Shi J , Samal A , Marx D . How effective are landmarks and their geometry for face recognition?[J]. Computer Vision & Image Understanding, 2006, 102( 2):117-133.
- [33] Dryden I L , Mardia K V . Statistical Shape Analysis: With Applications in R. Statistical data analysis explained :, 2008.
- [34] Sirovich L , Kirby M . Low-dimensional procedure for the characterization of human faces[J]. J.opt.am.a, 1987, 4(3):519-524
- [35] Sirovich L , Kirby M . Low-Dimensional Procedure for the Identification of Human Faces. 1987:103-108
- [36] Belhumeur P N , Hespanha J , Kriegman D J . Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection[J]. Springer, Berlin, Heidelberg, 1996:711-720

- [37] Etemad K , Chellappa R . Discriminant Analysis for Recognition of Human Face Images (Invited Paper)[J]. Springer, Berlin, Heidelberg, 1997:1724-1733
- [38] Fisher R A . THE STATISTICAL UTILIZATION OF MULTIPLE MEASUREMENTS[J]. Annals of Human Genetics, 2012, 8(4):376-386.
- [39] Moghaddam B , Pentland A , Jebara T . Bayesian face recognition[J]. Pattern Recognition: The Journal of the Pattern Recognition Society, 2000(11):33:1771-1782
- [40] Phillips P J . Support Vector Machines Applied to Face Recognition[C]// MIT Press. MIT Press, 2001:803-809.
- [41] Wright J , Yang A Y , Ganesh A , et al. Robust Face Recognition via Sparse Representation[J]. IEEE Computer Society, 2009(2):210-227
- [42] Pentl A , Moghaddam B , Starner T . View-Based and Modular Eigenspaces for Face Recognition[C]// CiteSeer. CiteSeer, 1994.
- [43] Wiskott L , Fellous J M , Kuiger N , et al. Face recognition by elastic bunch graph matching[J]. IEEE Computer Society, 1997(7):775-779
- [44] Tai S L . Image Representation Using 2D Gabor Wavelets[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996:959-971
- [45] Mikolajczyk K , Schmid C . A performance evaluation of local descriptors[C]// 2003:II-257-II-263.
- [46] Ahonen T , Hadid A , Pietikainen M . Face Description with Local Binary Patterns: Application to Face Recognition[J]. 2006, 28(12):2037-2041.
- [47] Bicego M , Grosso E , Lagorio A , et al. Distinctiveness of faces: a computational approach[C]// ACM. ACM, 2008:35-35
- [48] Howard J S , Lembach M L , Metzler A V , et al. Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2[C]// IEEE International Conference on Computer Vision. IEEE Computer Society, 2003:1150-1157

- [49] Liu C , Wechsler H . Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model[J]. IEEE Transactions on Image Processing, 2002,11(4):467-476
- [50] Liu C , Wechsler, Harry. Independent Component Analysis of Gabor Features for Face Recognition.[J]. IEEE Transactions on Neural Networks, 2003,14(4):919-928
- [52] Tan X , Triggs B . Fusing Gabor and LBP feature sets for kernel-based face recognition[C]// Analysis and Modeling of Faces and Gestures. Springer-Verlag, 2007:235-249
- [53] Huang G B , Mattar M , Berg T , et al. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments[C]// Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition. 2008.
- [54] Sun Y , Wang X , Tang X . [IEEE 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - Columbus, OH, USA (2014.6.23-2014.6.28)] 2014 IEEE Conference on Computer Vision and Pattern Recognition - Deep Learning Face Representation from Predicting 10,000 Classes[J]. 2014:1891-1898.

### 三、课题内容及具体方案

#### 1. 课题内容

(1) 调研人脸识别技术的研究背景，梳理国内外在该领域的研究现状与发展趋势，分析人脸识别流程中存在的人脸误检以及准确性、鲁棒性问题。

(2) 针对人脸误检问题，分析问题原因以及现有解决方案，提出基于语义特征的人脸误检筛选算法。

(4) 针对通用人脸识别算法的准确性鲁棒性问题，分析现有解决方案大多过于关注全局特征而忽略局部语义特征的缺点，提出基于全局与局部语义特征相融合的人脸识别算法。

(4) 对系统进行需求分析，分别对应功能性需求和非功能性需求如性能分析等。

(5) 完成系统的概要设计、详细设计与实现，包含登录模块、管理模块、图像上传模块以及算法分析模块。

#### 2. 系统需求分析（应用软件工程专业描述工具描述）

在系统开发的过程中，首先要做的就是对系统的需求进行分析，在系统的功能需求方面，主要是利用应用人员用例图对系统的功能进行总结。在非功能分析方面，主要是对鲁棒性等提出要求。经过本节的需求分析，可以为人脸识别系统的设计提供明确的目标，为后续的开发提供明确的指导。e

##### 2.1 系统功能性需求

根据系统的用例分析，人脸识别系统的使用角色分别为：算法应用人员、管理人员。分别从不同角度分析系统应具备的功能。

图 1 展示了人脸识别系统的面向算法应用人员所提供的功能。包括有登录、个人信息管理、上传图片、人脸识别、退出系统等功能。其中 1:1 人脸识别提供 1:1 人脸比对功能，用户用过上传两张人脸图像，然后系统开始接受任务并对这两张图像进行 1:1 人脸比对操作，返回人脸相似度。1:N 人脸识别功能则是提供 1:N 人脸搜索功能，用户通过上传一张待搜索人脸，然后系统接受任务对这张人脸图像与本地人脸库中的人脸一一比对，返回并显示人脸搜索结果。

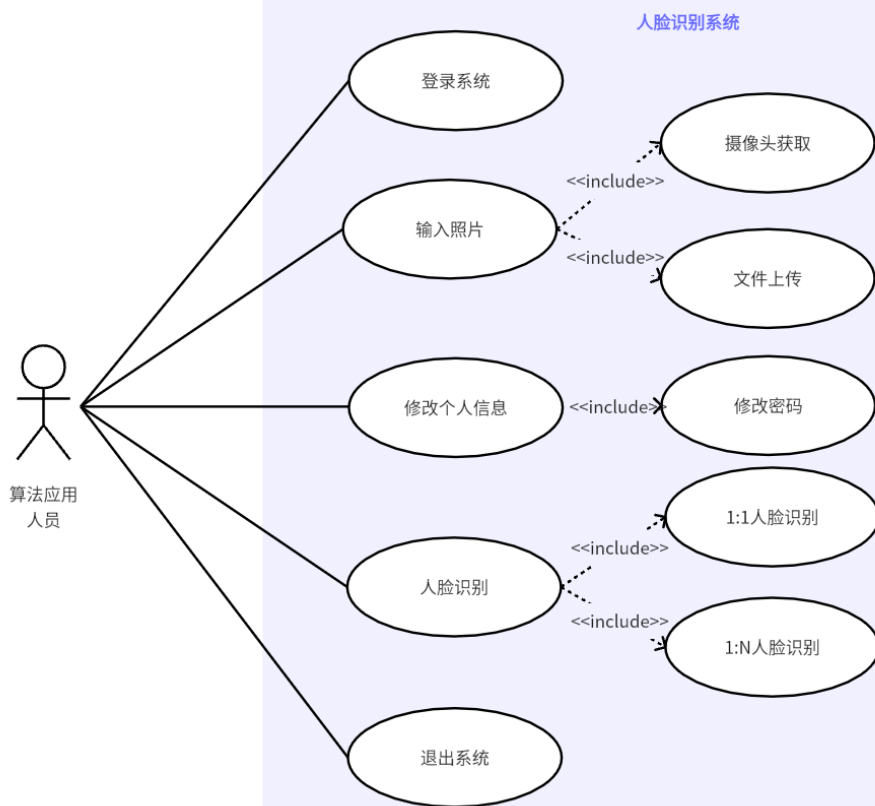


图 1 算法应用人员用例图

表 1 算法应用人员功能需求表

功能名称	功能描述	说明
登录认证	认证用户权限	无
图片上传	上传图像文件或摄像头拍摄	无
信息管理	修改密码等	无
人脸检测结果	查看人脸检测结果	返回是否有人脸，有则返回人脸框
人脸识别结果	查看人脸识别结果	返回人脸相似度或本地检索结果
退出系统	用户退出系统	无

图 2 展示了人脸识别系统面向管理人员所提供的功能。除了上传图像、登录系统、个人信

息管理、退出系统之外，管理人员还需要对普通用户进行管理，对用户上传的图片进行管理，对识别算法模块的模型进行管理。在算法模块中，管理人员可以通过更新模型，以达到应对不同算法和提升算法性能的效果。

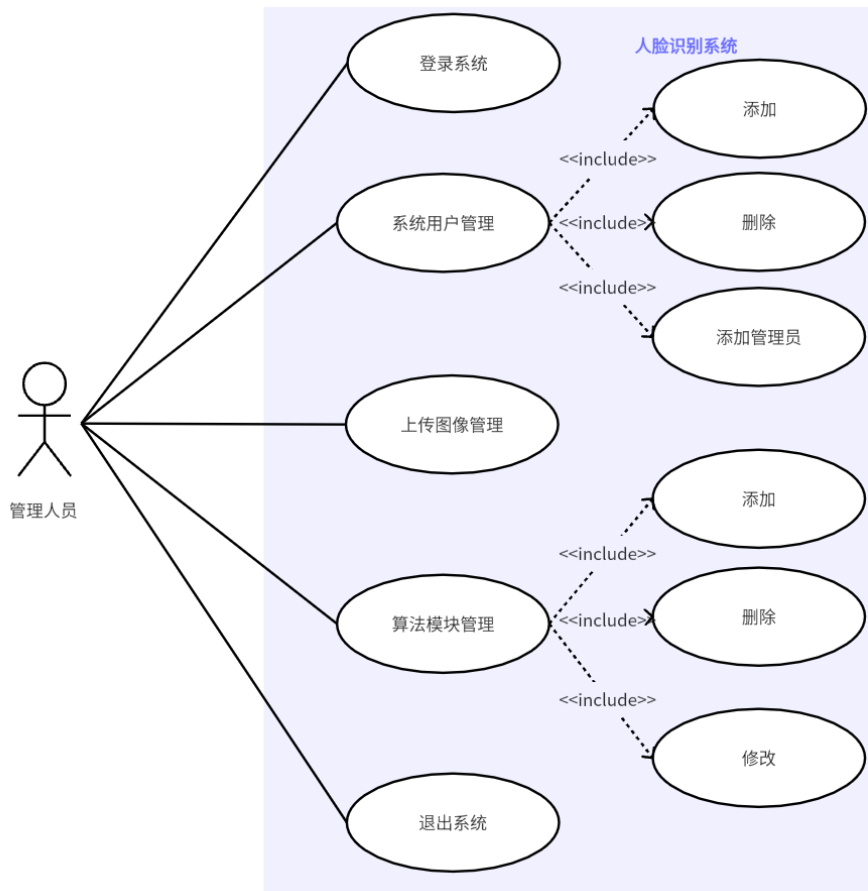


图 2 管理人员用例图

表 2 管理人员功能需求表

功能名称	功能描述	说明
登录认证	认证用户权限	无
系统角色管理	管理普通用户	无
信息管理	修改信息密码等	无
上传图片管理	管理用户上传图片是否规范	如不规范，发出警告
算法模块管理	管理算法模型	更新、替换算法模型，提升性

		能
退出系统	用户退出系统	无

## 2.2 系统性能需求

### （1）高精度

训练出来的模型要保证高精度，能够有效的提取出人脸特征，完成人脸识别的要求，具体量化指标为人脸识别准确率保持在 90%以上。

### （2）高效性

这个要求贯穿整个系统，系统需要实时的获取一个或多个图像数据。人脸检测需要快速高效的进行人脸检测，人脸识别也需要能够实时准确的呈现出识别结果，具体量化指标为人脸检测单张速度保持在 10ms 级别，人脸识别单张速度保持在 100ms 级别。

### （3）可靠性

各个模块采用的算法要成熟，理论可靠，整个系统稳定、流畅的运行。

## 3. 系统概要设计（应用软件工程专业描述工具描述）

### 3.1 系统架构

根据提出的需求分析，本文对人脸识别系统进行了功能模块划分，人脸识别系统分为了如下五个大的功能模块，分别是登录模块、管理模块、图像上传模块。数据处理模块以及识别算法模块，系统功能模块如图 3 所示。



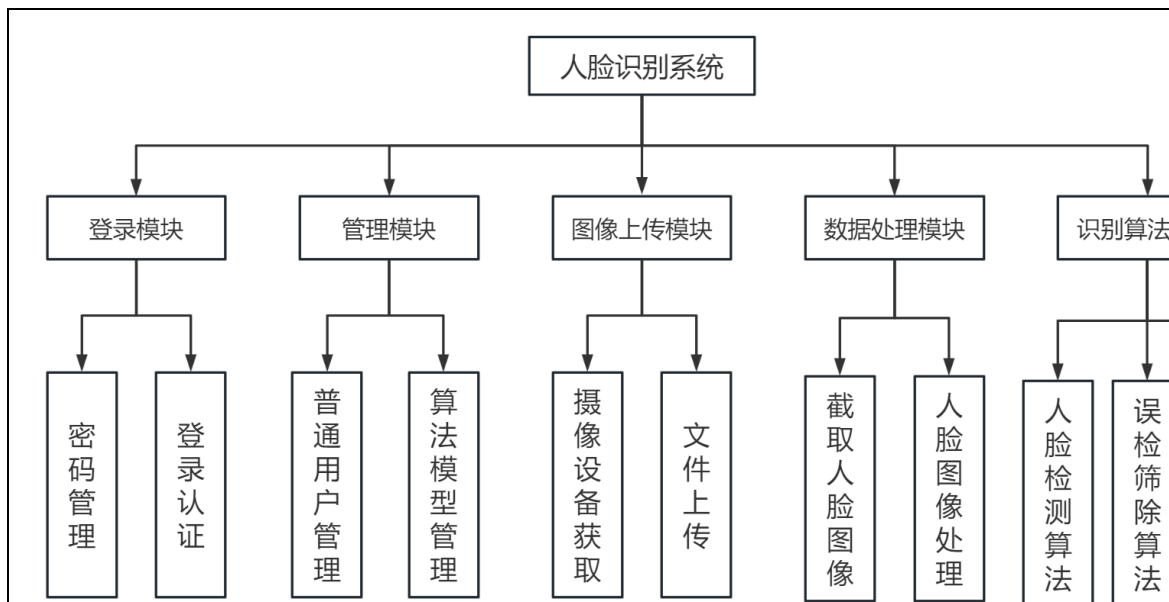


图 3 系统总体功能模块结构

### 3.2 登录模块

该模块由密码管理子模块和认证子模块组成，各模块功能如下：

#### （1）密码管理模块

为客户提供了密码修改的功能，以保证用户的信息安全。密码管理顺序图如图 4 所示，用户发出密码更新请求到密码配置服务，密码配置服务接收请求后将请求提交至数据库，提交成功后返回用密码管理界面。

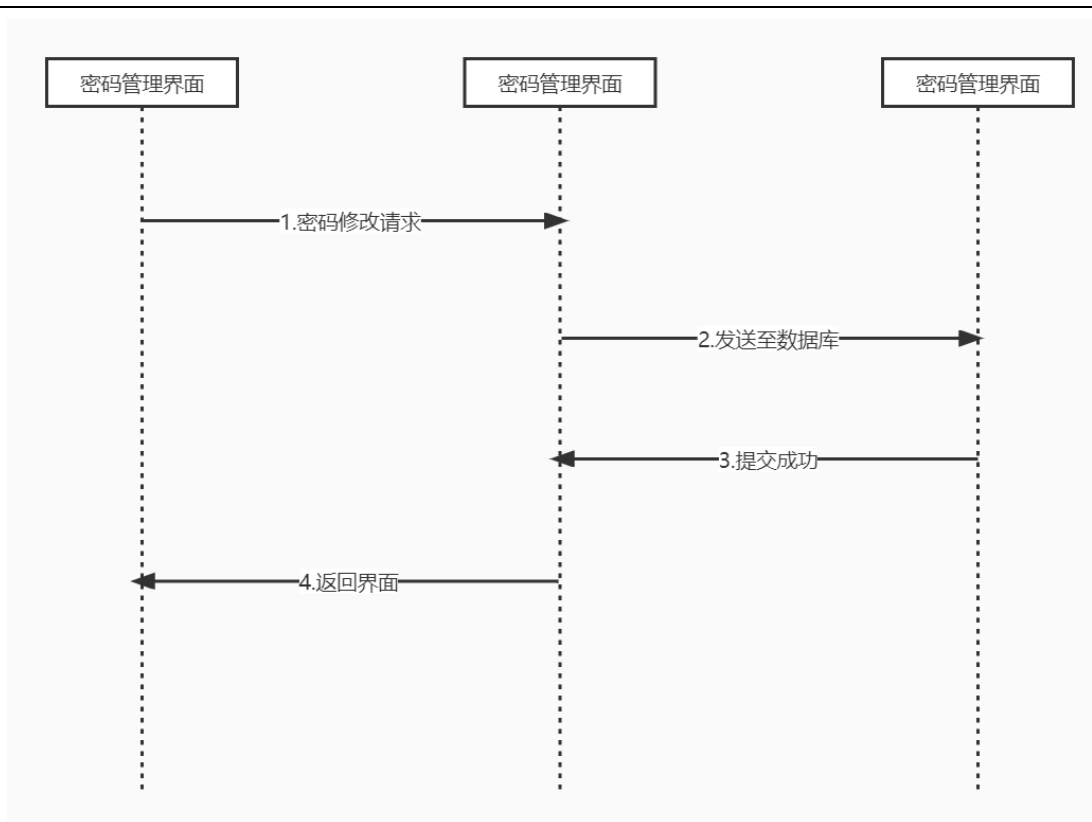


图 4 密码管理顺序图

## (2)登录认证子模块

用户登录身份验证流程图如图 5 所示。在进入系统之前，系统会判断用户是否已经登录。如果没有会话信息，系统将跳转到登录页面。用户登录后，系统将保留一段时间的登录会话信息。在此期间，进一步访问不需要登录。系统将判断用户身份属于管理员还是普通用户，并跳转到相应身份的主界面。对于具有管理员身份的用户，系统将开放所有功能供用户操作;对于普通用户身份的用户，系统将只提供部分功能供用户使用。

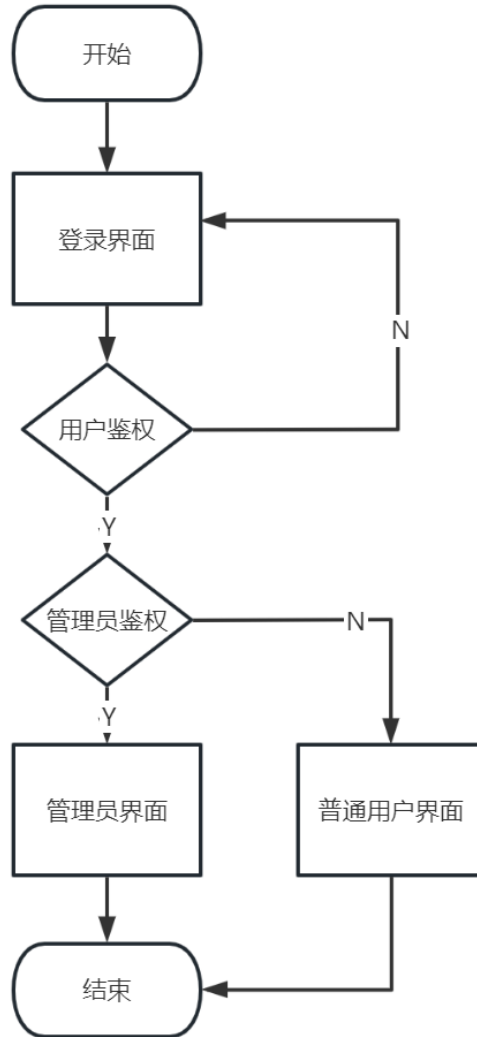


图 5 用户登录认证流程

### 3.3 管理模块

该模块由普通用户管理模块和算法模型管理子模块组成，各模块功能如下：

#### (1) 普通用户子管理子模块

可以对普通用户的权限进行管理，可以添加、修改或删除普通用户使用该系统的权限。

用户权限管理顺序图如图 6、图 7、图 8 所示，管理员通过用户管理界面对普通用户使用权限发起添加、修改或删除命令到后台相应服务。

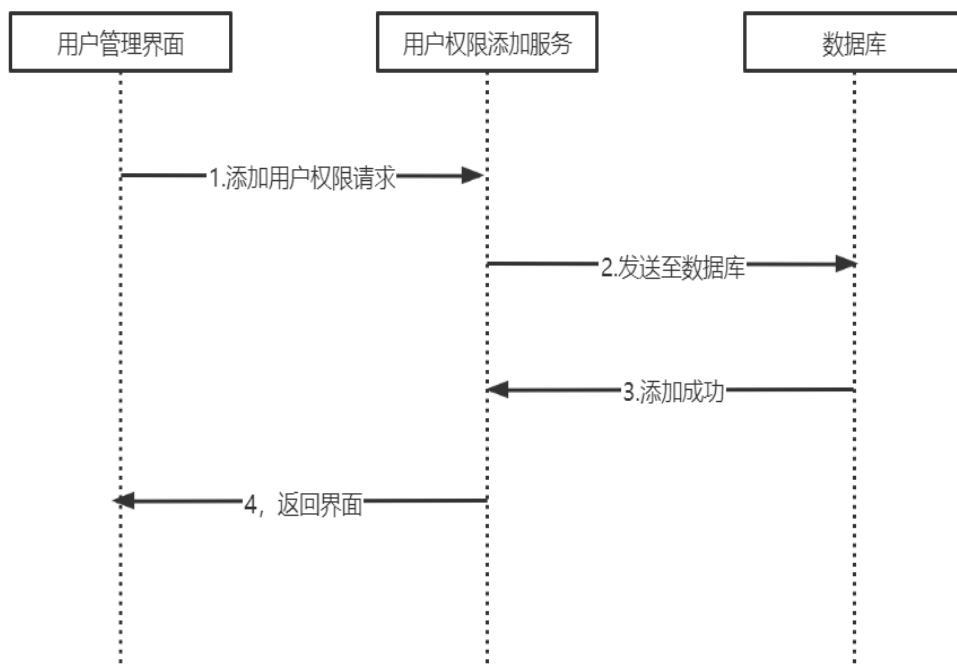


图 6 用户权限添加顺序图

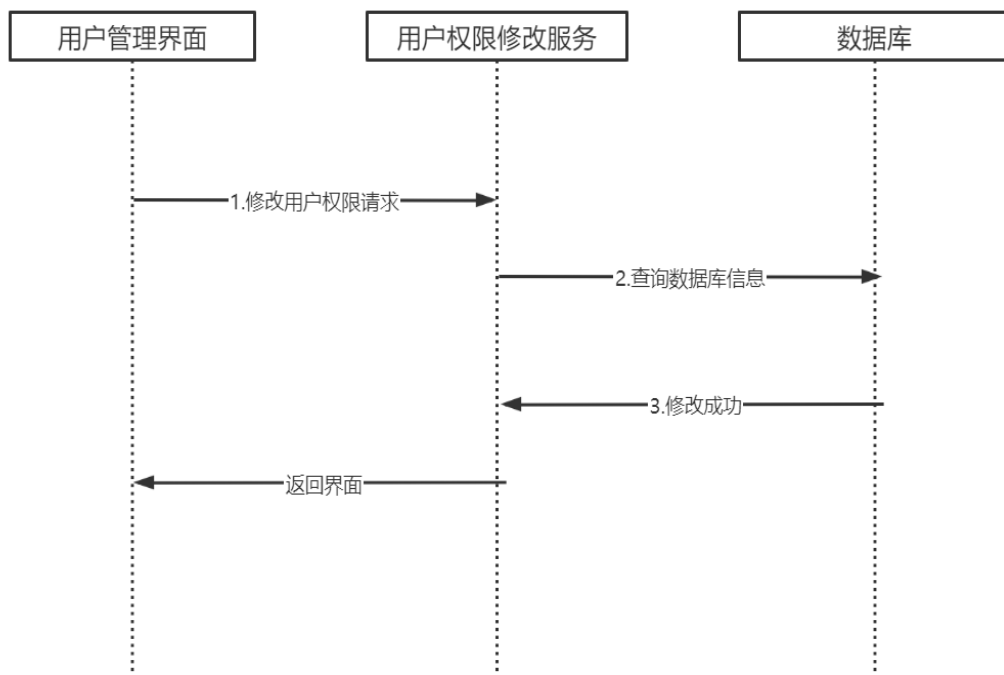


图 7 用户权限修改顺序图

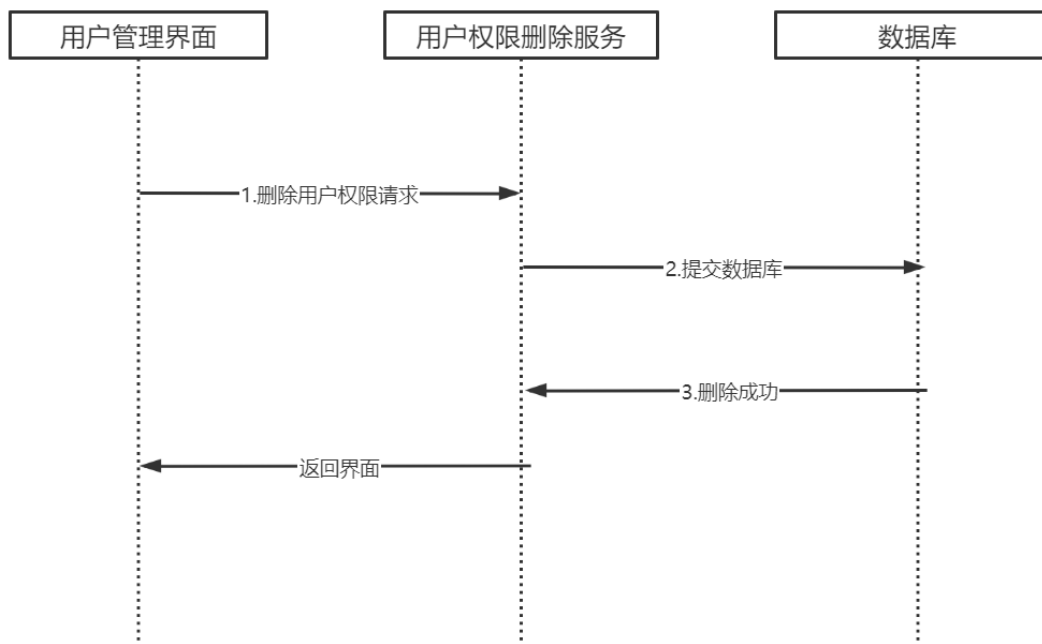


图 8 用户权限删除顺序图

## (2) 算法模型管理子模块

管理员可以通过添加、修改或删除算法模型，以达到更新或改进算法的目的。

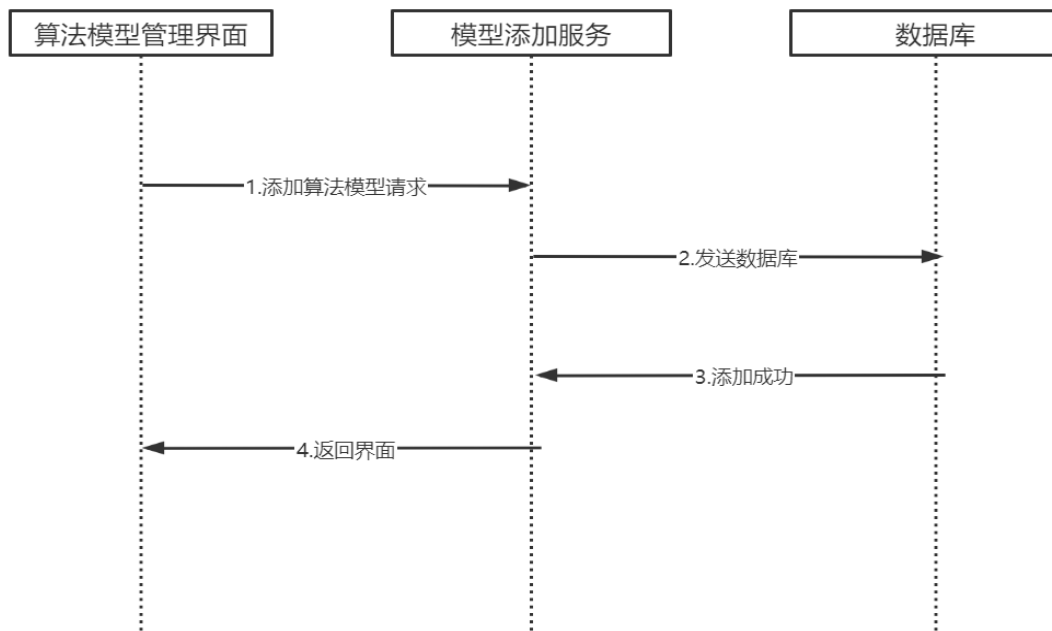


图 9 算法模型添加顺序图

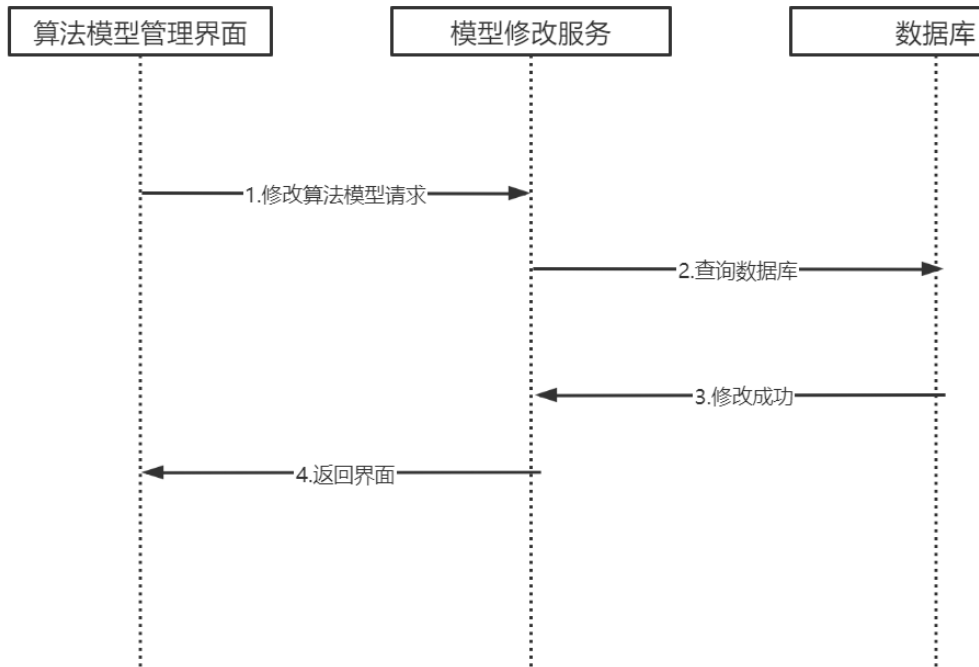


图 10 算法模型修改顺序图

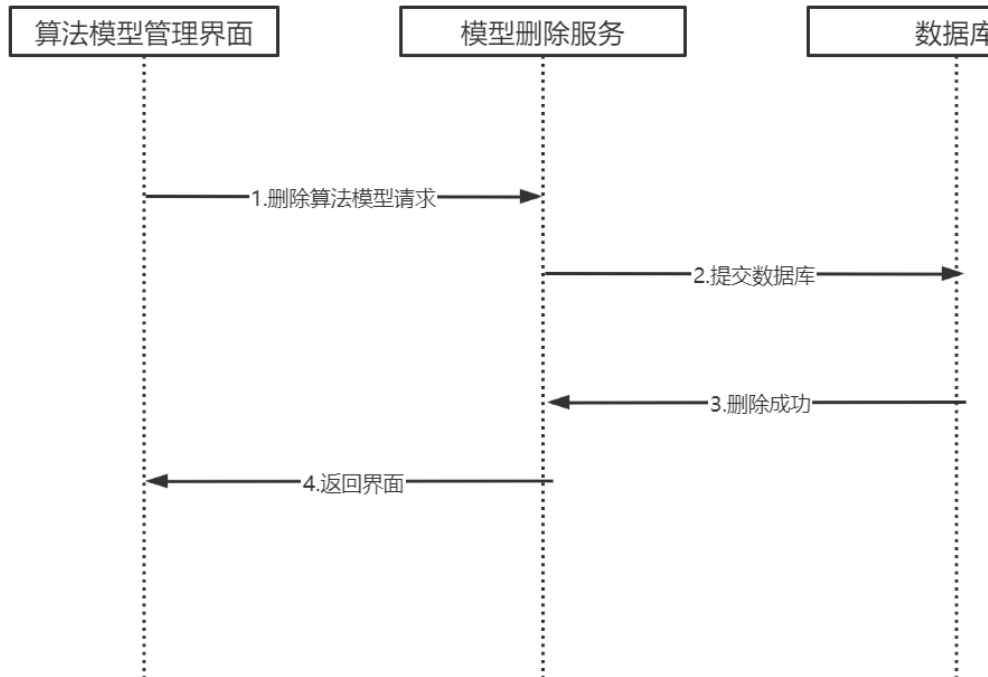


图 11 算法模型删除顺序图

算法模型管理顺序图如图 9、图 10、图 11 所示，管理员通过算法模型管理界面对系统所使用的算法模型发起添加、修改和删除命令后对应后台相应服务。

3.4 图像上传模块

图像上传模块顺序图如图 12、图 13、图 14 所示，图像上传模块允许用户使用本地上传和摄像设备获取的方式进行上传图片至数据处理模块。当用户上传图片时，系统对上传的图片进行分析，如果格式正确则发送图像处理请求至图像处理模块。

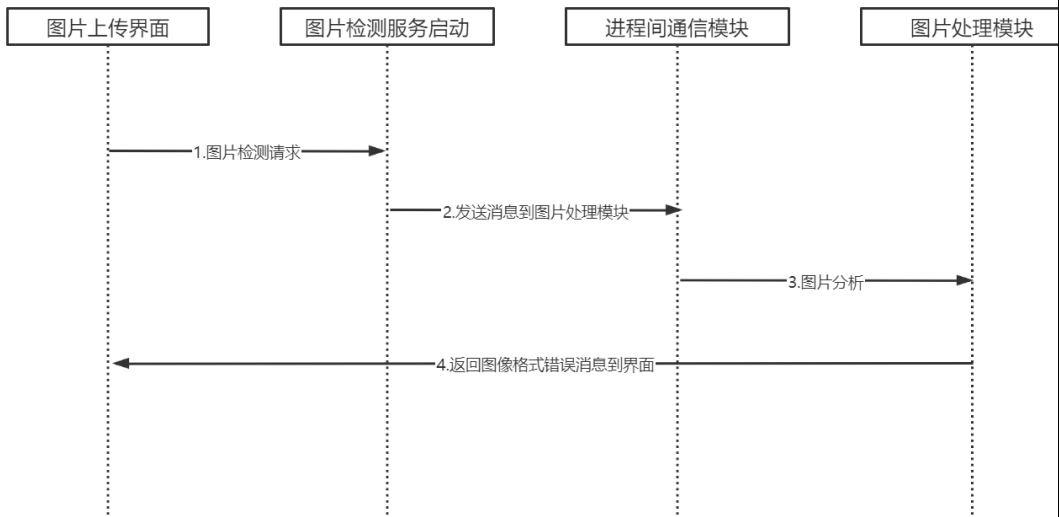


图 12 图像上传失败顺序图

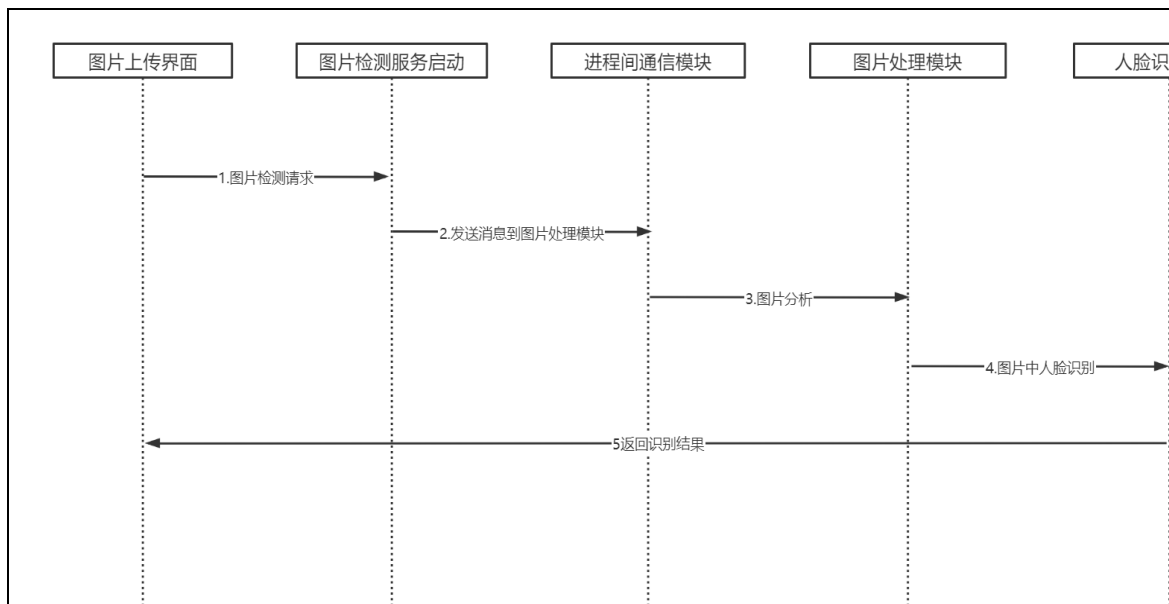


图 13 图片上传顺序图

### 3.5 数据处理模块

数据处理模块顺序图如图 14 所示，数据处理模块通过发送请求到人脸检测器，将图像输入到人脸检测器进行人脸位置检测，并将检测到的人脸图像进行截取，并进行一系列的数据处理操作。数据处理后的图像作为算法模块的输入进行后续的计算。

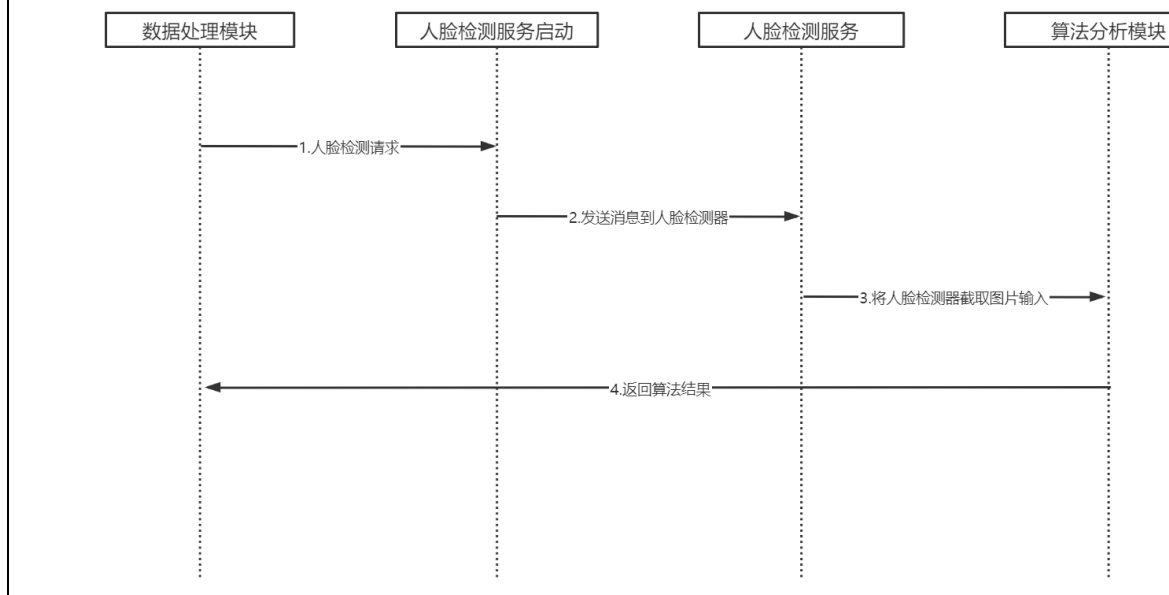




图 14 数据处理顺序图

### 3.6 识别算法模块

#### (1)人脸检测对齐子模块

人脸检测对齐子模块是整个人脸识别任务流程中的第一个功能模块，该模块主要对用户提供的图像进行人脸检测，然后对人脸检测结果进行对齐和缩放。图 15 为人脸检测对齐模块处理流程图，预期的实现方案中，人脸检测使用 MTCNN 算法模型，因为该算法方便易用，而且在进行人脸检测的同时可以提取对应的人脸五点坐标，可直接用于后面的人脸对齐。人脸对齐部分采用基于五点坐标的仿射变换方法，简单高效。经过检测对齐的人脸最后缩放尺寸为 112\*112，便得到最终的人脸检测对齐结果。

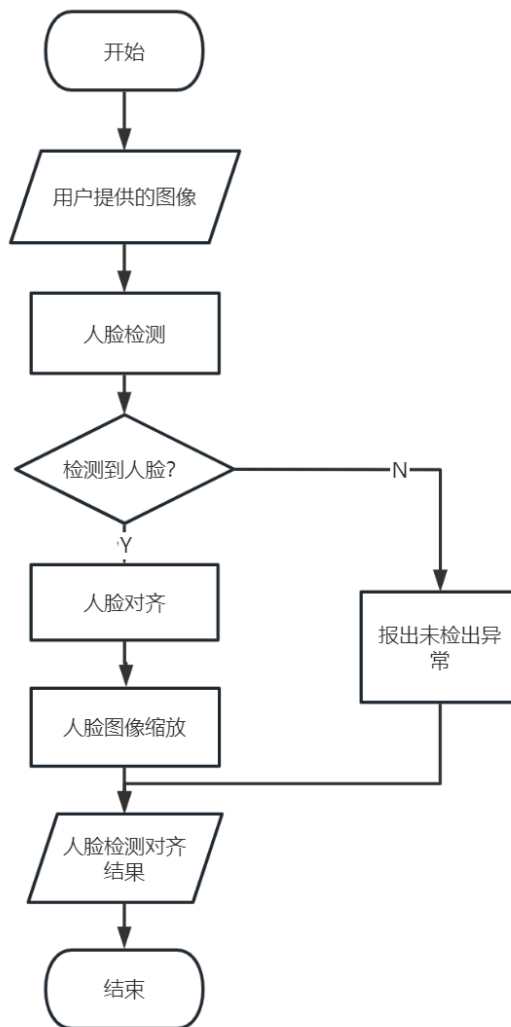


图 15 人脸检测对齐模块处理流程图

## (2) 人脸误检筛除子模块

本文提出的基于语义特征的人脸误检筛除方法 SF-FPC(Semantic Feature-based False Positive Classifier)位于人脸检测模块之后,不对前置人脸检测模块本身进行任何改动,其功能只是对检测处理结果进行分类过滤。SF-FPC 算法流程图如下图 16 所示。

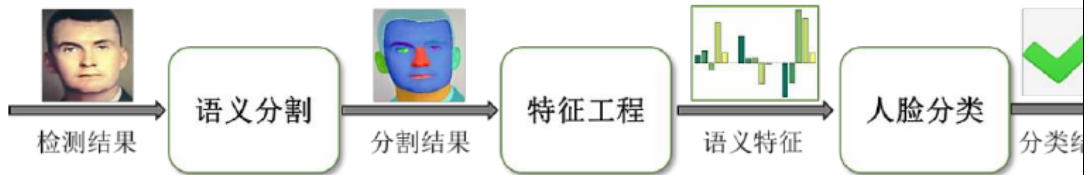


图 16 SF-FPC 算法流程图

如图 16 所示, SF-FPC 算法主要包括三个核心模块:语义分割模块、语义特征工程模块、人脸分类模块。其中,考虑到对实时性和精确度的要求,语义分割模块选用了实时语义分割网络模型 BiSeNet,该模块对输入的人脸检测结果进行精细的人脸语义分割,得到语义分割结果。语义特征工程模块采用传统特征工程处理技术,目的是基于人脸语义分割结果构造出最具表征能力的语义特征。人脸分类模块选用支持向量机(SVM)作为分类器,根据语义特征进行是否人脸分类,完成最终的人脸误检筛除。图 17 给出了人脸误检筛除模块处理流程图,使用本文提出的 SF-FPC 方法,最后保留真实人脸的人脸检测对齐结果和对应的语义分割结果作为模块输出,供后续人脸识别使用。

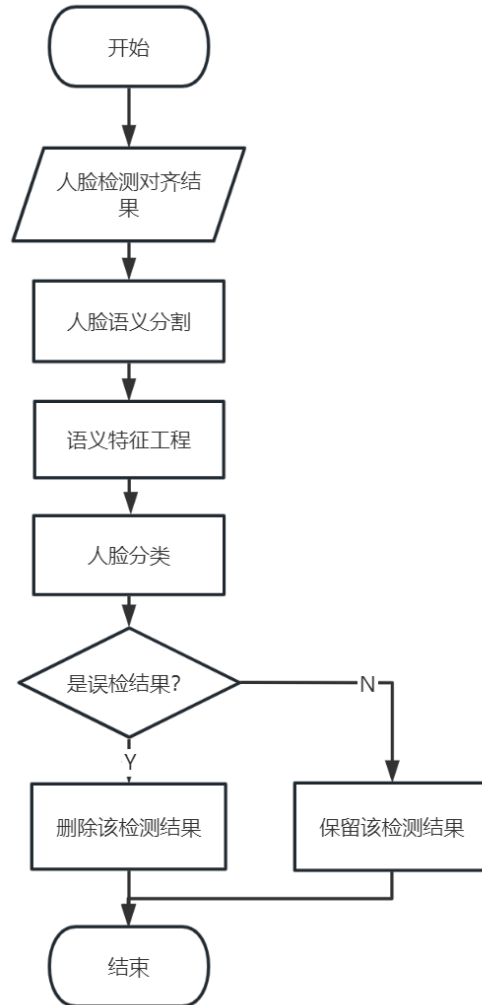


图 17 人脸误检筛除模块处理流程图

算法 1 展示了语义特征工程的具体算法步骤：

### 算法 1 基于特征工程的语义特征提取选择算法

**Input:**人脸语义分割结果 result\_seg, 分类标签 Y

**Output:** 人脸语义特征 feature\_seg

```
1: feature_origin = GetFeature(result_seg);    //提取粗略特征
2: feature_pre = Preprocess(feature_origin);    //特征预处理, 包括特征无量纲化和缺失值处理
3: feature_select_1= VarianceSelect(feature_pre); //方差选择法,返回选择后的特征
4: feature_select_2= PearsonSelect(feature_select_1, Y); //皮尔森相关系数法
5: feature_select_3= Chi2Select(feature_select_2, Y); //卡方检验
6: feature_seg= MICSelect(feature_select_3, Y); //最大信息系数选择法
7:Return feature_seg;
```

### (3) 人脸识别子模块

受到全局与局部相结合思想的启发, 本文提出了基于全局与局部语义特征相融合的人脸识别算法 FGLSF-FR(Face Recognition-based on the Fusion of Global and Local Semantic Features)。与现有大多数深度学习人脸识别算法只对全局人脸进行特征提取的方式不同, FGLSF-FR 通过人脸语义分割技术显式的引入了人脸局部语义特征, 然后结合局部遮挡信号采用特征融合的方式得到表征能力更强、鲁棒性更好的融合特征, 以用于最后的人脸比对。FGLSF-FR 的方法流程图如图 18 所示。

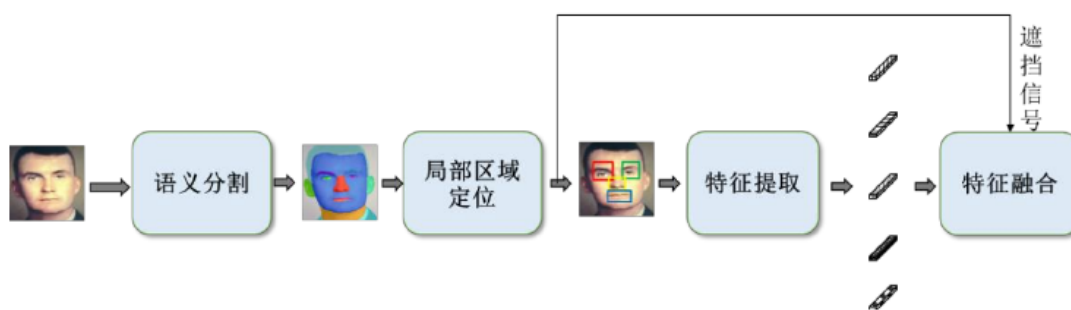


图 18 FGLSF-FR 算法流程图

如图 18 所示, FGLSF-FR 方法主要包括 4 个核心模块:语义分割模块、局部区域定位模块、特征提取模块、特征融合模块。其中, 语义分割模块是为了能够在复杂环境下也能进行精确的局部语义区域定位而服务的, 该模块对输入人脸进行精细化语义分割, 得到人脸分割结果以供

后续使用。局部区域定位模块主要根据人脸语义分割结果，得到左眼、右眼、鼻子和嘴巴 4 个重要人脸语义区域的精确位置以及相应的局部遮挡信号。特征提取模块采用深度卷积神经网络对全局人脸和 4 个局部人脸语义区域分别提取深度特征，这些特征中包含一个全局人脸特征和 4 个局部语义特征。特征融合模块结合局部遮挡信号对全局和局部特征进行融合,得到表征能力更强的融合特征。语义分割模块采用了人脸误检剔除模块中完全一致的方案，这里便不再赘述。

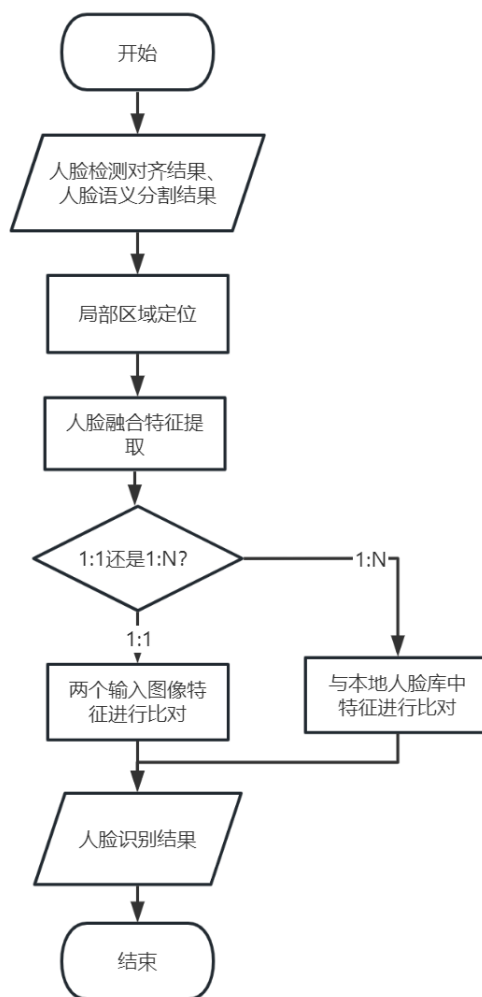


图 19 人脸识别模块处理流程图

人脸识别模块是该系统最重要的核心功能模块，它首先对输入人脸进行特征提取，然后根据具体的人脸识别任务进行对应的人脸比对操作，输出对应的识别结果。图 19 展示了人脸识

别模块处理流程图，具体实现方案中，使用了基于全局与局部语义特征相融合的人脸识别算法，首先根据人脸检测对齐结果和人脸语义分割结果进行局部区域定位，得到带局部语义区域位置的人脸图像，然后使用训练好的人脸特征提取模型提取人脸融合特征，接下来根据人脸识别任务的不同采取对应的操作，如果是 1:1 人脸比对任务，则对两个输入人脸特征进行比对，返回其人脸相似度作为人脸识别结果，如果是 1:N 人脸搜索任务，则对输入人脸特征与本地人脸库中的特征进行一一比对，返回人脸相似度最高的人脸及其相关信息。人脸相似度的计算方式仍然采用的余弦相似度。

算法 2 展示了局部区域定位算法的具体步骤：

---

**算法 2** 局部区域定位算法

---

**Input:**人脸语义分割结果 result\_seg

**Output:** 局部语义区域位置坐标 boxes, 局部遮挡信号 signal\_occ

```

1: for i in range(4) do           //4 个局部语义区域
2:   flag = i                     //指定目标操作区域
3:   image_bin = toBinarization(result_seg, flag) //对指定目标区域结果进行二值化
4:   closed = morphology_close(image_bin)       //对目标区域进行闭运算操作
5:   box = get_box(closed)                     // 计算目标区域最大轮廓旋转包围盒
6:   if box != [] then                        // 不为空,该区域未被遮挡
7:     box_ = box_opt(box, flag) // 目标框优化
8:     signal_occ[i] = 1 ,boxes[i] = box_ // 给输出结果赋值
9:   else                                    //为空, 该区域被遮挡
10:    signal_occ[i] = 0                    //对应遮挡信号置 0
11:   end if
12: end for
13: Return boxes, signal_occ              //返回结果

```

---

3.7 系统数据库设计

数据库ER图

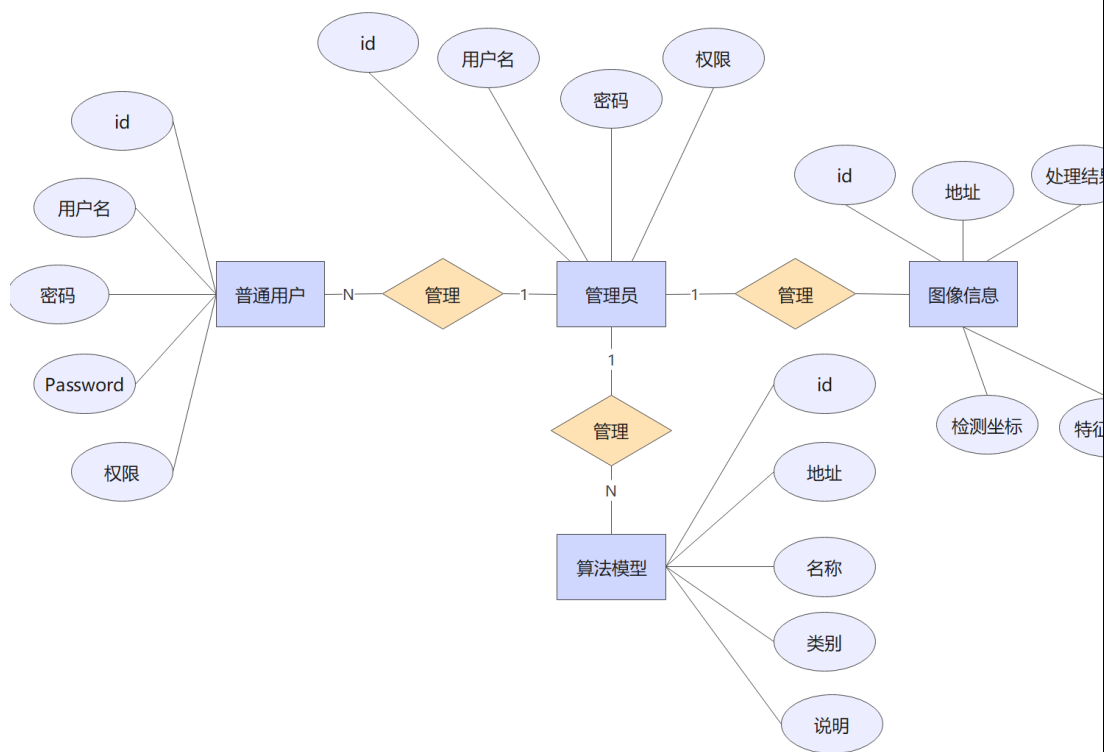


图 18 系统 ER 图

图 18 是数据库设计的 ER 图，根据图中的信息可以得知系统需要的数据库包括：用户信息表、图像信息表和算法模型信息表。数据库表汇总如表 1 所示：

表 1 系统数据表汇总

表名	功能说明	存储数据库
AdminInfo	用户信息表	MySQL

PicInfo	图片信息表	MySql
ModelInfo	算法模型信息表	MySql

表 2 为用户信息表，该表以用户 id 为主键，自增，其他信息由用户名、用户密码、用户权限，每条数据代表一个用户的信息。

表 2 用户信息表

字段名	字段类型	备注
Id	Long	表主键，自增
Username	Varchar	用户名
Password	Varchar	密码
UserType	Integer	用户权限，1 为普通用户，0 为管理员

表 3 为图像信息表，该表以图片 id 为主键，自增，其他信息有图片地址、是否存在地址、人脸框坐标、特征信息，每条数据代表一个图片的信息。

表 3 图像信息表

字段名	字段类型	备注
PicId	Long	表主键，自增
PicAdd	Varchar	图片地址，长度 200
WhetheFace	Integer	是否存在人脸，1 为存在，0 为不存在
FaceLocation	List	人脸框位置坐标
Feature	text	人脸特征信息

表 4 为算法模型信息表，该表以模型 id 为主键，自增，其他信息有模型地址、模型名称、模型类型、模型描述，每条数据代表一个模型的信息。

表 4 算法模型信息表

字段名	字段类型	备注
ModelId	Long	表主键，自增



ModelAdd	Varchar	模型地址，长度 200
ModelName	Varchar	模型名称，长度 200
ModelType	Int	模型类型，0 为大型网络模型，1 为轻量级网络模型
ModelDescribe	Varchar	模型描述，长度 200

#### 4. 拟采用的开发环境、方法、测试方案

##### 4.1 开发环境

硬件环境:

处理器: Intel(R) Core(TM) i7-10700k CPU @ 5.10GHz

显 卡: NVIDIA 2080Ti

软件环境:

操作系统: Windows 10

数 据 库: MySQL

开发语言: Python (核心算法)、Java (后台服务)

开发工具: PyCharm、IntelliJ IDEA

##### 4.2 开发方法

###### 1.2.1 算法设计

**1.人脸检测:** 人脸识别的第一个任务是人脸检测对齐，主要对用户提供的图像进行人脸检测，然后对人脸进行对齐和缩放，具体实现上，人脸检测使用训练好的 MTCNN 算法模型，可直接用于后面的人脸对齐。

**2.人脸误检剔除:** 预期使用一种基于语义特征的人脸误检剔除方法。其中考虑到对实时性和精确度的要求，语义分割选用实时语义分割网络模型 BiSeNet，该模型对输入的人脸检测结果进行精细的人脸语义分割。采用传统特征工程处理技术，目的是基于分割结果构造出最具表征能力的语义特征。人脸分类模型选用支持向量机 (SVM) 作为分类器，根据语义特征进行是否人脸分类，完成最终人脸误检剔除。

**3.人脸识别:** 基于全局与局部思想结合的思想，预期实现基于全局与局部语义特征结合的人脸识别算法，通过人脸分割技术显式的引入了人脸局部语义特征，然后结合局部遮挡信号采用特征融合的方式得到预期表征能力更强、鲁棒性更好的融合特征，以用于最后的人脸对比。

#### 4.2.2 系统开发

系统开发拟采用面向对象的分析设计方法，开发平台采用易移植的 Java 平台，便于后期的维护与移植，数据库选用 MySQL，采用数据库连接池对数据库连接进行维护。

### 4.3 测试方案

#### 4.3.1 算法测试实验设计

##### 1. 误检筛除算法

为了验证提出的人脸误检筛除方法在误检误检筛除方面的准确性，已经通过误检筛除提高整体人脸检测结果的有效性和通用性，本文设计了以下对比实验。

定性方面，首先得到人脸语义分割模型 BiSeNet-Face 在正常人脸和误检结果上的实际分割可视化效果，然后得到 MTCNN 在添加误检筛除方法前后的实际人脸检测可视化效果。

定量方面，计划在 FDRFP 数据集上对基于卷积网络的人脸误检筛除方法（Convolutional Network-based False Positive Classifier, CNN-FPC）和本文方法进行对比测试。然后选用当前流行的 3 中人脸检测算法 MTCNN、Face R-CNN 和 SSH，分别在人脸检测评测基准 FDDB 和 WIDER FACE 上添加本文方法前后进行对比测试。

##### 2. 人脸识别算法

为了对提出的基于全局与局部语义特征相融合的人脸识别算法进行最优参数模块选择以及性能测试，计划在 4 个不同的人脸识别评测基准 LFW、CFP-FP、AgeDB30 和 OFDB 上设计一下对比实验，这 4 种评测基准分别代表了普通、带侧脸、跨年龄和带遮挡 4 种不同的人脸识别场景。

(1) 采用 VGG-19、Inception-ResNet-v2 和 ResNet-101 作为基础特征提取网络，分别使用提出的特征提取网络结构和直接提取全局特征的方式进行对比实验。

(2) 使用本文提出的算法，对提出的不同特征融合方式进行对比实验。

(3) 使用本文提出的算法，对使用不同损失函数进行对比实验，损失函数选取了 Softmax、A-Softmax、Cos-Softmax 和 Arc-Softmax。

(4) 使用本文最优模块参数配置的模型与其它流行的通用人脸识别算法进行对比实验。这里选取的其它通用人脸识别算法主要有 FaceNet、SphereFace、CosineFace 和 ArcFace。

#### 4.3.2 系统测试

本系统的测试主要内容是通过模拟普通用户和管理员用户进行登录和用户管理操作。并模拟用户上传图片进行人脸人别的过程。

测试用例如表 5、表 6 所示

表 5 用户登录测试

测试项	测试用例	预期结果	测试结果
管理员登录	管理员正常登录	登录成功主页面包含管理员专用功能和普通功能	
管理员登录	用户名或密码错误	登录失败，回到登陆界面	
用户管理	管理员新增用户	新增成功，用户列表增加新建用户	
用户管理	管理员修改用户	修改成功，用户列表更新用户的修改	
用户管理	管理员删除用户	删除成功，用户列表删除所选用户	
用户登录	用户正常登录	登录成功，主页包含正常功能	
用户登录	用户名或密码错误	登录失败，回到登录界面	

表 6 图片上传分析测试

测试项	测试用例	预期结果	测试结果
上传图片	上传不符合规范图片	返回图像不符合格式规范	
上传图片	上传其他类型图片	返回部署于图像文件	
上传图片	上传不包含人脸图片	返回图像中无人脸结果	
上传图片	测试 1:1 功能	返回人脸相似度	
上传图片	测试 1:N 功能	返回本地人脸库搜索结果	

## 5. 技术难度及特色分析

(1) 实现基于语义特征的人脸误检筛除方法 SF-FPC。通过引入人脸语义分割技术，并利用特征工程构造出表征能力更强的人脸语义特征，然后结合 SVM 结构简单、泛化能力强的优点，尝试实现对人脸误检结果的有效分类筛除，进一步提高整体检测算法的性能与鲁棒性。

(2) 实现基于全局与局部语义特征相融合的人脸识别算法 FGLSF-FR。通过人脸语义分割技术进行人脸局部语义区域的定位，然后对全局和局部语义区域分别进行特征提取，显式的引

入人脸局部语义特征，最后结合局部遮挡信息对全局和局部特征进行有效的特征融合，提高最终特征的表征能力和鲁棒性，进而提升了算法在复杂场景下的性能和场景适应性。

（3）设计并实现人脸识别系统，将本文提出的人脸误检剔除方法和人脸识别算法应用到具体的人脸识别业务中，系统架构每层都作为单独的模块进行开发，只提供每层之间的通信接口，这样开发效率搭搭提高，并且每层之间相互独立存在，具有很强的拓展性以及便于后期维护。

## 6、本人主要工作描述

（1）实现基于语义特征的人脸误检剔除方法 SF-FPC，通过引入人脸语义分割技术，利用基于人脸语义分割结果构造出的表征能力更强的人脸语义特征，结合 SVM 分类器实现人脸误检分类剔除。为验证方法的性能，在 FDRFP、FDDB 和 WIDER FACE 数据集上设计多组实验。

（2）实现基于全局与局部语义特征相融合的人脸识别算法 FGLSF-FR，通过人脸语义分割技术显式的引入了人脸局部语义特征，然后结合局部遮挡信息对全局特征和局部语义特征进行特征融合，得到表征能力更强、鲁棒性更好的特征用于人脸识别。设计不同的特征提取网络结构和不同的特征融合方式，并且在 LFW、CFP-FP、AgeDB30 和 OFDB 这 4 种不同的人脸识别评测基准上进行多组实验。

（3）设计实现人脸识别系统，它包括前台客户端和后台管理。在服务端,对于客户端的请求, Server 会启动并调用 Servlet, Servlet 根据客户端的请求生成响应内容并将其传给服务器,服务器会将响应返回客户端。

## 四、工作进度的大致安排

应包括文献调研，工程设计，项目开发和调试，实验数据的分析处理，撰写论文等。

## 五、预期成果

应包括软硬件产品、文档、模型、专利、论文等

- 人脸识别系统
- 工程硕士学位论文

## 六、审核意见

### 导师意见

导师签名:

年 月 日

培养单位负责人意见

培养单位负责人签名:

年 月 日