

重点人群项目技术选型文档

一、简介

本项目是一个在大型公共场所、交通枢纽、企事业单位等地，需要对特定人群进行有效监控，以提高安全性和防范潜在威胁。重点人群识别系统是一种基于人脸识别视觉技术的智能安防解决方案，旨在通过高效的人脸识别和行为分析技术，实现对特定人群的准确识别与监测。

基于百度的FastDeploy框架，在NVIDIA Jetson Orin NX 16G边缘设备上部署，经过测试，项目也可以在瑞芯微RK3588上进行部署。

FastDeploy是一款全场景、易用灵活、极致高效的AI推理部署工具，支持云边端部署。提供超过 160+ Text, Vision, Speech和跨模态模型开箱即用的部署体验，并实现端到端的推理性能优化。包括 [物体检测](#)、[字符识别\(OCR\)](#)、[人脸](#)、[人像扣图](#)、[多目标跟踪系统](#)、[NLP](#)、[Stable Diffusion文图生成](#)、[TTS](#)等几十种任务场景，满足开发者多场景、多硬件、多平台的产业部署需求。

二、架构图



三、技术选型

1.人脸检测 SCRFD模型

项目通过网络摄像头采集实时的监控数据，对每一帧的图像进行人脸检测，检测到人脸并将其采集下来以供人脸识别算法使用。

在推理速度方面，SCRFD通常被认为是一种轻量级的人脸检测器，具有较快的推理速度。它能够在图像或视频中快速准确地检测出人脸区域，为后续的人脸识别提供准确的输入。

SCRFD是基于Single Shot MultiBox Detector (SSD) 架构的改进版本。它在传统SSD模型的基础上引入了一些关键的改进，以提高人脸检测的性能和精度。以下是SCRFD模型的一些主要特点和改进：

- 1.高分辨率：SCRFD模型采用了高分辨率的输入图像，通常为640x640或更大的尺寸。这使得模型能够更好地捕捉小尺寸的人脸特征，提高检测精度。
- 2.上下文感知：SCRFD通过引入多层级的特征金字塔结构，充分利用图像中的上下文信息。这种多尺度特征表示使模型能够更好地处理不同尺寸和尺度的人脸，并提高检测的鲁棒性。
- 3.多任务学习：SCRFD模型同时进行人脸检测和关键点定位的任务。这样的多任务学习可以进一步提高人脸检测的准确性和多样性。
- 4.基于Anchor的检测：SCRFD采用了一组称为锚框 (Anchor) 的预定义框，用于检测和定位人脸。这些锚框具有不同的尺度和长宽比，以覆盖各种人脸形状和大小。

SCRFD模型经过大量的人脸数据集训练和优化，可以在实时或近实时的速度下进行高效的人脸检测。它在许多人脸识别、人脸表情分析、人脸属性分析等应用中得到广泛应用。

2.人脸识别 insightface

基于insightface的 Partial_FC模型对人脸图片进行embedding提取，根据余弦相似度计算人脸间的相似值。

Partial_FC 是一种具有创新性的人脸识别模型，通过将人脸图像的特征映射到一个高维嵌入空间，并利用角度余弦距离进行分类，实现了更加准确和鲁棒的人脸识别。人脸识别网络的FC层所需的空间及计算开销与人脸id的数量呈线性关系，当人脸id数量庞大时，往往需要使用分布式训练，数据并行会有巨大通信开销，而模型并行往往受到单块GPU显存掣肘。Partial FC通过近似softmax loss，在不损失模型性能的情况下可以通过堆叠GPU的数量来加速训练，同时能够解决数据长尾分布与类间冲突的问题。

Partial_FC 是人脸识别领域中的一个模型或算法，其相对于传统的全连接层 (Fully Connected Layer) 在人脸识别任务中具有以下优势：

- 1.参数量减少：Partial_FC 使用了部分连接 (Partial Connection) 的方式，即只连接部分神经元，而不是全连接所有神经元。这导致了模型的参数量大幅减少，减少了训练和推理时的计算负担，提高了效率。
- 2.减轻过拟合风险：部分连接的方式可以减少模型的复杂度，从而减轻了过拟合的风险。在人脸识别等任务中，过拟合是一个常见的问题，而 Partial_FC 的使用可帮助降低这一风险，提高模型的泛化能力。
- 3.稀疏性：部分连接导致模型中的连接变得更为稀疏，即每个神经元只与一部分其他神经元相连接。这种稀疏性有助于提高模型的可解释性，并且在一定程度上降低了计算成本。
- 4.加速训练和推理速度：由于参数量减少、减轻了过拟合风险以及稀疏性等因素，Partial_FC 可以加速模型的训练和推理速度。这对于实时性要求较高的人脸识别应用来说尤为重要。
- 5.保持模型性能：尽管 Partial_FC 减少了连接和参数量，但其在人脸识别任务中仍然能够保持较高的识别性能。这意味着可以在不牺牲准确率的情况下获得更高的效率和速度。

综上所述，insightface Partial_FC 在人脸识别任务中的优势主要体现在减少参数量、减轻过拟合风险、稀疏性、加速训练和推理速度以及保持模型性能等方面，为人脸识别应用的性能和效率提供了有力支持。

3.人脸分类

在jetson上基于YOLOv5训练的分类模型

1. 速度和效率：YOLOv5相对于之前的版本，如YOLOv3和YOLOv4，在速度和效率方面有所提升。YOLOv5采用了一系列优化策略，包括网络结构设计、模型精简和推理加速等，使得在保持较高检测准确率的前提下，提高了检测速度和效率。
2. 模型轻量化：YOLOv5相对于之前的版本，如YOLOv3和YOLOv4，模型参数量更小，模型结构更简单，这使得在移动端设备上的部署更加轻便和高效，适合于一些对模型大小和计算资源有限制的场景。
3. 准确率提升：YOLOv5在保持速度和效率的同时，通过引入一些新的技术和策略，如新的数据增强方法、损失函数的优化等，提升了目标检测的准确率，特别是在小目标检测和遮挡目标检测等方面有了更好的表现。
4. 易用性：YOLOv5具有简单易用的特点，通过PyTorch框架实现，并提供了开源代码和预训练模型，使得用户能够方便地进行模型训练、部署和应用，并且具有一定的灵活性和可定制性，适用于不同的应用场景和需求。
5. 支持多种应用场景：YOLOv5具有较强的通用性，不仅可以用于普通的目标检测任务，如行人检测、车辆检测等，还可以应用于特定领域的检测任务，如人脸检测、口罩检测等，满足了各种不同应用场景下的需求。

YOLOv5在速度、效率、准确率和易用性等方面都有一定程度的优势，适用于各种不同的目标检测应用场景，并且在实际应用中取得了较好的效果。

在RK3588上基于paddlecls训练的分类模型

1. 高性能：PaddleClas基于百度PaddlePaddle深度学习框架构建，优化了模型结构和算法，具有较高的训练和推理性能。其在多个公开数据集上取得了领先水平的分类性能，包括ImageNet、CIFAR等。
2. 丰富的模型库：PaddleClas提供了丰富的预训练模型库，包括ResNet、MobileNet等经典模型，以及一些轻量级模型，满足了不同场景下的需求。
3. 灵活性和可定制性：PaddleClas提供了易于使用的API和模型训练工具，支持用户快速搭建、训练和部署自定义分类模型。用户可以根据实际需求进行模型定制和优化，适应不同的应用场景。
4. 多端部署：PaddleClas支持多端部署，包括移动端、服务器端等多种硬件平台，可以灵活应用于移动应用、云端服务等不同场景，满足了多样化的部署需求。
5. 持续更新和维护：PaddleClas作为PaddlePaddle生态系统的一部分，得到了百度团队的持续更新和维护，保持了较高的性能和稳定性，同时不断地引入新的技术和算法，保持了竞争力。

PaddleClas具有高性能、丰富的模型库、灵活性和可定制性、多端部署以及持续更新和维护等优势，适用于各种不同规模 and 要求的图像分类任务。

4.数据库 mysql