

基于 TINY-YOLO 的嵌入式人脸检测系统设计*

Design of Embedded Face Detection System Based on TINY-YOLO

游 忍¹ 周春燕¹ 刘明华¹ 邵延华² 展华益^{1,3} (1 长虹 AI 实验室,四川 绵阳 621051;

2 西南科技大学信息工程学院,四川 绵阳 621010;

3 西北大学电子工程和计算机系,美国 伊利诺斯州埃文斯顿 60208)

摘要: 基于深度学习的人脸检测算法对于人脸检测性能有了很大的提升,但是大多数算法无法部署在移动端上。如 MTCNN 等算法虽然可以部署在嵌入式系统,但是在实际的应用场景中,有时不能满足需求。将 TINY-YOLO 用于人脸检测,结合 NCNN 框架,实现了在移动端的部署及商用。精度上相比于 MTCNN 有较大的提升,在实际的应用场景中表现优异,并且算法运行时间不会随着人脸数量的增加而增加,更适用于多人脸的嵌入式环境下人脸检测。

关键词: 人脸检测, TINY-YOLO, 嵌入式, 深度学习

Abstract: Depth learning based face detection algorithm has greatly improved the performance of face detection, but most of the algorithms can not be deployed on the mobile terminals. Although algorithms such as MTCNN can be deployed in embedded systems, they sometimes can not meet the requirements in actual application scenarios. In this paper, TINY-YOLO is used for face detection, and the NCNN framework is used to implement the deployment and business in mobile terminals. Compared with MTCNN, the precision of the algorithm is improved greatly, and it performs well in practical application scenarios. Moreover, the running time of the algorithm does not increase with the increase of the number of faces, so it is more suitable for face detection in multi-face embedded environment.

Keywords: face detection, TINY-YOLO, embedded, deep learning

深度学习对于人脸检测算法有很大的促进。目前基于深度学习的主流目标检测算法主要可以分为 one-stage 和 two-stage 两种, two-stage 的代表主要有 R-CNN^[1-3] 系列, one-stage 代表主要有 SSD^[4]、YOLO (You Only Look Once)^[5-7] 系列。一般情况下, two-stage 目标检测算法的精度比 one-stage 高,但是 one-stage 算法的速度更快。One-stage 算法经过改进后精度已经有了很大的提升,已经接近甚至超越了 two-stage 算法。利用 two-stage 中的 Faster R-CNN 框架, Wang, Z. Li 等人提出 Face R-CNN^[8] 人脸检测算法, 加入了 center loss、OHEM 等 tricks, 在 Fddb 数据集上 recall 达到 98% 以上。

上述基于深度学习的方法虽然在精度上有了很大的提升,但是模型很大,在嵌入式处理器上运行时间很慢。本文将 YOLO 算法的轻量化版本 TINY-YOLO 算法用于人脸检测, 得到一个可以在手机甚至电视机空调上使用的人脸检测系统。

1 系统结构设计

本方法使用深度学习算法检测人脸,意在解决嵌入式处理器环境下,深度检测算法检测模型大、运行效率低甚至无法运行的问题。为解决这一问题,从两方面着手,训练数据和深度网络模型:首先,采集大量人脸样本库,尽可能多的包括各种场景、各种姿态、各种角度、各种表情以及各类人脸,通过 K-means 聚类算法将训练数据已标注好的人脸框进行聚类,从中挑选出几类最适合的人脸框类型作为锚框,提高检测精度;此外,对卷积层参数进行调整,减少计算量,以达到压缩模型大小,提高检测速度的目的。本系统处理器采用 RK3288 芯片,结合了腾讯 ncnn 框架,它是一个专门为嵌入式端优化的高性能神经网络前向计算框架。

2 Anchor 聚类

在检测算法中,最直接的方式是预测包含目标的方框的宽、

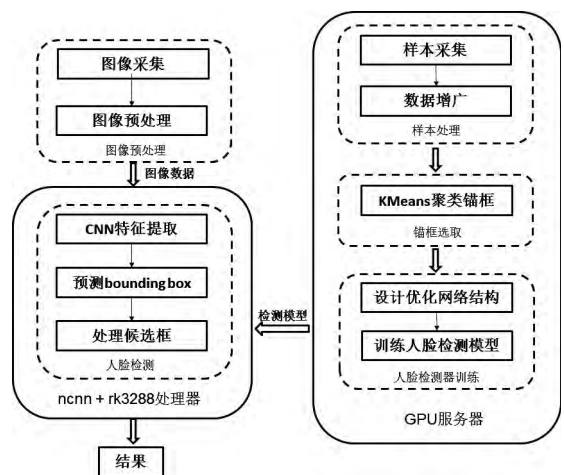


图 1 系统结构设计框图

高。但是在实际进行训练的时候,梯度将出现不稳定的现象。所以,目前主流目标检测算法都是预测默认给定方框的偏移量,这些默认给定的方框被称作 Anchor。Anchor 的尺寸应该与要预测的目标的尺寸接近,这样在进行预测的时候预测函数才能看作一个线性函数,保证结果的准确性。为了得到本系统中人脸检测算法的 Anchor,我们采用了 K-means 算法,它所采用的步骤如下:

- 1) 将所有真实标签的方框的中心坐标令为 (0,0);
- 2) 在真实标签的方框中随机选取 3 个方框作为聚类中心;
- 3) 按照以下两个步骤迭代:①根据每个方框与每个聚类中心的 IOU,将每个方框分配给与聚类中心中 IOU 最大的那个聚类中心。将真实方框分成 3 组。②在每组真实方框中,计算每个方框与该组内其它方框的 IOU 的平均值。将平均值最小的那个方框作为该组新的聚类中心,直到 IOU 平均值变化小于阈值结束。

* 长虹公司一号工程,2018-01

3 模型压缩与训练

TINY-YOLO 算法的耗时主要是在卷积部分,骨干网络如图 2 所示,含有大量的卷积层。假设卷积层的输入特征图维度是 $C_i \times W_i \times H_i$, C_i, W_i, H_i 是特征图的通道数和宽高。卷积核维度是 $C_k \times K \times K$, 其中 C 和 K 分别代表卷积核的通道数和卷积核的宽高。同理,输出的特征维度是 $C_o \times W_o \times H_o$, 则卷积核的参数量为 $C_o \times K \times K \times C_i$, 计算量为 $C_o \times K \times K \times W_o \times H_o \times C_i$ 。从上面公式中可以看出,卷积核的通道数对计算量与模型参数的大小有很大影响,所以这里我们将骨干网络中卷积核的通道数量全部减半,以减少模型大小和计算量。

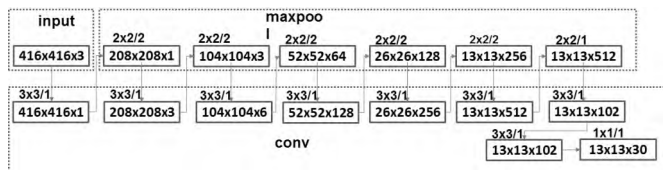


图 2 TINY-YOLO 骨干网络

本系统的人脸检测算法主要在 widerface 数据集上进行训练。测试数据集为自己创建的明星脸数据库,有 110 万张图片。训练时,对人脸数据采用了平移、旋转、部分置黑、剪切等几何变换手段,增加数据量。对数据进行了归一化以及 PCA 白化处理,使得所有特征具有相同的方差。消除图像中无关的信息,增强有关信息的可检测性。

训练的优化算法采用 SGD, batch size 取值为 64, 一共训练 30 万个 batch。开始训练的 15000 batch 内,学习速率为 10^{-4} , 200000 batch 的时候减少为 10^{-5} , 250000 batch 的时候减少为 10^{-6} 。虽然在算法进行实际检测的时候我们将图片缩放为 416x416, 但是为了检测到不同尺度的人脸, 增加模型的鲁棒性, 在训练的时候进行了多尺度训练。即将每一张图片随机地缩放到 320, 352, ..., 608 尺度范围内其中一个尺寸。

4 实验结果分析

最终的人脸检测模型结合腾讯 NCNN 框架,在主频 1.8G, 内存 2G 的 RK3288 处理器平台上实际运行场景如图 3 所示。在自己创建的明星脸数据库上测试结果为 precision 99%、recall 98%。而 MTCNN 算法的 precision 99.5%、recall 92%。虽然系统比 MTCNN 模型的 precision 少 0.5%, 但是 recall 比 MTCNN 高了 6%。当图片中只有一个人脸时,系统的平均运行时间为 183ms, MTCNN 平均时间为 137ms。但是当人脸数量增加时, MTCNN 的运行时间随着人脸数量的增加会显著增加, 但是 TINY-YOLO 算法不会随着人脸数量的增加而增加。

TINY-YOLO 算法的这个特点决定了它比 MTCNN 算法在实际应用场景中更有优势。



图 3 系统运行场景

5 结束语

与其它的人脸检测算法如 MTCNN 等比较, 不但本文方法在性能上更好, 算法运行时间不会随着人脸数量的增加而增加, 这对于工业应用有很大的优势, 基于该算法的产品已经商用, 取得较好的检测效果。

参考文献

- [1] Girshick R, Donahue J, Darrell T, et al. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation[C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2014
- [2] Girshick R. Fast R-CNN[J]. Computer Science, 2015
- [3] Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C] // International Conference on Neural Information Processing Systems, 2015
- [4] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot Multi-Box Detector[C] // European Conference on Computer Vision. Springer International Publishing, 2016:21-37
- [5] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection [C] // Computer Vision & Pattern Recognition, 2016
- [6] Redmon J, Farhadi A. YOLO9000: Better, Faster, Stronger[C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016:6517-6525
- [7] Redmon J, Farhadi A. YOLOv3: An Incremental Improvement [J]. 2018
- [8] Wang H, Li Z, Ji X, et al. Face R-CNN[J]. 2017

[收稿日期:2018.11.6]

(上接第 46 页)

在水中遨游, 模板图像和后面的图像除了角度变化外并没有其他特别大的差别。因为我们在特征点筛选环节考虑到了 2D 旋转和缩放的问题, 所以能够实现在这类情形下正确匹配的效果。由于图像质量不高, 计算量较小, 所以匹配速度非常快。综合实验结果来看, 算法在实际应用中有出色的表现。

3 结束语

本文探讨的是如何提升模板匹配算法的速度和鲁棒性。传统的算法通过计算图像块之间的相似性度量方法往往花费很多时间并且不能有效避免各种干扰带来的影响。实验结果表明, 该方法具有快速鲁棒的特点, 且具有稳定性。所以, 对以往算法进行归纳总结并提出又快又好的算法很有研究价值, 可以运用于很多具体实际场景。

参考文献

- [1] 马颂德, 张正友. 计算机视觉[M]. 北京: 科学出版社, 1998
- [2] Bradski G R, Konolige K, Rublee E, et al. ORB: An efficient alternative to SIFT or SURF [C] // 2011 International Conference on Computer Vision, 2564-2571
- [3] Debel T, Qron S, Rubinstein M. Best-Buddies Similarity for robust template matching [C] // Computer Vision & Pattern Recognition, 2015
- [4] Bian, Lin W Y, Matsushita Y, et al. GMS: Grid-Based Motion Statistics for Fast, Ultra-Robust Feature Correspondence[C] // IEEE Conference on Computer Vision & Pattern Recognition, 2017

[收稿日期:2018.12.19]