项目技术指标与测试报告

1. 概述
2. 项目模型概述

本次项目构建的是，一个可针对口腔根管治疗后CT图像或X光图像进行根管牙自动定位与评分的软件系统，整个系统包括检测和评分两个主要模型：检测模型与评分模型。jpg或jpng格式的口腔CT图像或x光图像进入系统后，首先经过检测模型，得到以矩形框框定的根管牙，被框定的部分将进一步被切取并进入评分模型，评分模型将根据根管填充情况对其进行评分，系统最终输出框定出根管牙并给出评分后的jpg或jpng格式图像。

1. 主要需测试模块
2. 检测模型单独功能与性能测试；
3. 评分模型单独功能与性能测试；
4. 集成系统的功能与性能测试；
5. 项目标准与技术指标
6. 项目预期应用

集成后系统通过检测与验证后，将应用与医院或医疗机构口腔科，输入根管治疗后口腔CT图像或x光图像，以系统输出的评分结果来评价主治医生本次治疗的效果及医生水平。

1. 应用标准要求（按照口腔根管图像评分标准编写）
2. 输出检测框应准确包含根管牙全牙及单独根管部分，且检测框不允许过大而包含过多除根管及根管牙之外的其余信息；
3. 评分结果应符合以下标准：
4. 本项目技术指标（根据医院需求删改）
5. 精确度
6. 速度
7. 占用内存（计算机本身内存和显存）
8. 项目测试方案
9. 检测模型测试
10. 测试目标

使用验证集val\_txt中的图像对模型进行测试，本次使用的验证集共318图像。检测模型测试包括两个方面，第一个方面是确定一个评估标准，对预测结果的精确度进行测试，另一方面是评估超参数对网络的影响，从而筛选出最优的超参数。本模型相关超参数主要有四个:

nms衡量了预测框的密集程度，nms=1表示不允许标注框之间有重叠，nms=0表示认可网络预测出的所有标注框，在[0,1]之间进行测试，找到最适合的nms;

score\_threshold是网络确信度阈值，每个框被预测出来时，网络对于这个框的预测结果都会有一个确信程度，通过这个阈值，将确信程度低于该阈值的框全部删除，可提高预测结果的准确率；

detection\_scale是控制预测框大小的变量，由于训练时对图像进行的变换导致标注框略微比真实框大，网络将学习到这一特性并表现为最终预测框偏大，控制这一变量可使其略微缩小；

iou\_threshold是预测框和真实框的重叠程度阈值，将预测框与人工标注框对比，框重叠率越高表示预测效果越好。

调整这四个参数，选出最佳值。

最佳参数确定后，使用这一组参数来对模型的精确度进行测试。检测结果的精确度由ap(average precision)来表征，这一参数的含义及计算方法将在后面给出。测试过程中记录处理每一张图片的时长，即为模型的运行速度，同时查看并记录模型占用的内存和显存。

1. 测试步骤

①首先进行最佳超参数的筛选。四个参数中，控制其中三个不变，另外一个参数即为待优化参数，使待优化参数在某一个范围内变化，每取一个值都能获得一组数据的检测结果，这些标注框中，实际应该被预测出，并且预网络测结果也将其预测出的则为true\_positive，网络预测将其标出但实际不应该被预测出的则为false\_positive，定义一个recall变量，recall=true\_positive/应该被预测出的总量，用于表示在全部应该被预测出的标注框中，成功被预测出比例，recall越大，网络更不会错过有嫌疑的对象；定义另一个变量precision，precision=true\_positive/(true\_positive+false\_positive)，表示所有被预测出的框中，预测正确的比例，precision越大，网络预测准确度越高。前面提到的ap即是recall和precision通过一个函数综合计算出的结果。这三个参数将随着待优化参数的变化而变化，我们将这一变化表现为折线图，选择它们处于最优值时对应的参数为本模型的最优参数。

②选定最优参数后，测试得到的ap值反映了模型的精确度，ap越大，精确度越高。

1. 测试记录

①测试人员

②测试时间

③测试环境

④测试工具

⑤测试结果

1. 参数nms最优值选择：将nms选为待优化参数，在[0,1]之间每隔0.05取一个值进行测试，得到recall、precision、ap值随nms值变化折线图如下图3.1，可看出当nms取值在[0.65,0.7]之间时，这三个值可综合达到最优，模型最终选取nms=0.7;

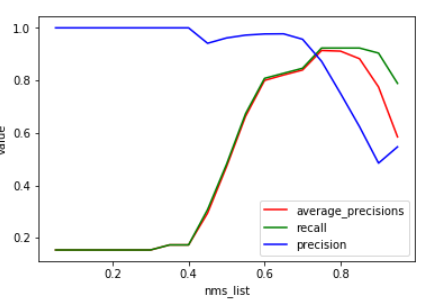
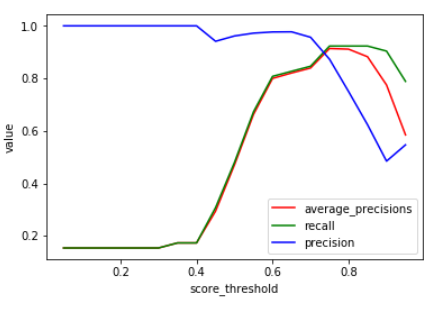


图 3.1

1. 参数score\_threshold最优值选择：将score\_threshold选为待优化参数，在[0,1]之间每隔0.05取一个值进行测试，得到recall、precision、ap值随score\_threshold值变化折线图如下图3.2，可看出当score\_threshold取值在[0.5,0.6]之间时，这三个值可综合达到最优，模型最终选取score\_threshold=0.5;



1. 参数iou\_threshold最优值选择：将iou\_threshold选为待优化参数，在[0,1]之间每隔0.05取一个值进行测试，得到recall、precision、ap值随iou\_threshold值变化折线图如下图3.3，可看出当iou\_threshold取值在xx之间时，这三个值可综合达到最优，模型最终选取iou\_threshold=0.3;
2. 参数detection\_scale最优值选择：将detection\_scale选为待优化参数，在[0,1]之间每隔0.05取一个值进行测试，得到recall、precision、ap值随detection\_scale值变化折线图如下图3.4，可看出当detection\_scale取值在xx之间时，这三个值可综合达到最优，模型最终选取iou\_threshold=xx;

e)最优参数选定后，利用该组参数处理测试集数据，本次测试选用318张图像中的49张。

6.测试结果评价

1. 评分模型测试
2. 测试目标
3. 测试环境
4. 测试工具
5. 测试步骤
6. 测试记录

①测试人员

②测试时间

③测试平台

④测试结果

6.测试结果评价

*精确度精细评估有一个单独模型，比较模型标注和手工标注，iou\_threshold（后续的最大overlaps）表示预测框与真实框的交并比，average precision（ap）是由recall和precision（这两个值都不是一个确定值，而是一个列表）算出来的评判标准，计算方法不是简单算平均，是由cumpute\_ap函数求得，recall是主要需要考虑的对象，原图没有标注时ap=0，处理方法不一定正确，是指标当中的一种，是检测模型约定的指标，数学定义不用深究。drawBox画出预测和原始标注的框（可视化）*

*nms在0.6左右 框可一定程度上重合*

1. 总结（功能及性能总结，缺点、不足及预期改进方法*（训练集可根据上下牙、切牙、乳牙等分类更加详细）*等）