**第六届Sky Hackathon**

**参赛项目书**

参赛学校：燕山大学

参赛队名：发际线的情况很对

指导老师：陈晶

团队成员：辛晨，张能，于明民，马信宏，何飞

### 项目说明：

第六届Hackthon以 AI 助力防疫主题。身着白色防护服的工作人员辛苦奋战在抗疫一线，我们可以通过AI的技术辅助防疫工作，例如口罩识别检测和口罩佩戴语音提醒，小区门禁二维码识别等场景的应用, 学习 AI项目在疫情防控中的设计理念。使用ASR进行引导提示，Object Detection对戴口罩情况进行识别，同时使用TTS进行语音播报提醒。

团队在比赛规定时间内完成了以上三个部分功能实现，在比赛过程中实现了对网络数据的获取、清理、标注，对YOLOV5-S、YOLOV5-M、YOLOX-S、YOLOX-M等模型进行选择与训练，并且使用TensorRT进行模型的部署。使用基于预训练的citrinet-512替换比赛预设的QuartzNet，并且搜集数据进行标注fine tune了ASR模型。TTS部分收集并标注数据训练了Tacotron2声学模型。

上述训练情况使用如下系统环境完成：

|  |  |
| --- | --- |
| Item | 参数 |
| GPU | P40(ASR/TTS)/3090(目标检测) |
| 显存 | 24G/24G |
| OS | Ubuntu 18.04 |
| 内存 | 50G/64G |
| Nemo | 1.4.0 |
| TensorRT | 7.1.3 |
| CUDA | 10.2 |
| Cudnn | 8.0 |

主要配置环境以及TRT测试效果均严格安装JetPacket 4.4 安装。

### 工作说明：

下面从语音识别(ASR)、语音合成(TTS)、计算机视觉目标检测(Object Detection)三个部分分别进行工作汇报。

1. ASR模型数据收集、标注以及训练部署
   1. 数据收集

ASR部分收集自身以及周边好友数据50条，其中男女声均有，年龄不等，每条均为不同人进行录制。

部分数据展示如下：

|  |  |  |
| --- | --- | --- |
| Id | Text | Duration |
| 1 | 你好大白请让我进入小区 | 4.0 |
| 2 | 你好大白请让我进入小区 | 4.89 |
| 3 | 你好大白请让我进入小区 | 2.0 |

* 1. 训练以及部署

模型部分首先采用AIShell2预训练的模型QuartzNet 15X5作为预训练的网络，后发现citrinet-512相对比QuartzNet在收集的数据集上错误率耕地，于是使用Citrinet-512替换QuartzNet模型，并使用基于预训练的迁移学习模式进行网络训练。

模型最终训练300轮。

1. TTS模型数据收集、标注以及训练

由于Nemo1.4对于中文TTS预训练模型并没有进行提供，所以采用从头训练的方式进行训练。

2.1 TTS数据集收集以及标注

TTS共收集两个数据集：

第一个数据集由单人录制，50条数据，录制人为25岁成年男性，普通话，单声道，采样率为44100，wav格式，每条语音前后保留0.5s静音片段。数据集中每句话均添加随机拼凑的人名，希望以此提高一定的TTS生成能力。数据展示如下：

|  |  |  |
| --- | --- | --- |
| Id | Text | Duration |
| 23 | 晚上好杨亚平欢迎回家二维码有效请您通过 | 9.7034 |
| 24 | 中午好宋小杰欢迎回家二维码有效请您通过 | 10.193 |
| 25 | 上午好宋小杰欢迎回家二维码有效请您通过 | 10.6115 |

第二个数据集收集自腾讯TTS开放API，使用ffmpeg对采样率和码率进行修改，最终为单声道、采样率44100，wav格式。使用“小叮当“作为人名，进行录制，数据集内容展示如下：

|  |  |  |
| --- | --- | --- |
| Id | Text | Duration |
| 1 | 你好，小叮当，欢迎回家，二维码有效，请您通过 | 4.21 |
| 2 | 你好，小叮当，欢迎回家，二维码有效，请您通过 | 4.8 |
| 3 | 你好，小叮当，欢迎回家，二维码有效，请您通过 | 4.97 |

2.2 TTS模型训练与选择

对两个数据集均进行了多次训练。

数据集一训练轮次6000次后，依然存在较大底噪、发音不清情况，对于“你好、您好、上午好”等等字眼，以及人名部分存在较大杂音，推断是由于数据集内容过于庞杂导致的，在最终比赛提交时选择放弃该模型。

数据集二由低码率使用ffmpeg提高至训练需要的44100之后运行3000轮次，依然存在一定底噪、PESQ分数相对较低情况(PESQ均分大概1.3-1.5)，恰逢端午，比赛时间延长，于是进行加训，由于无法完成继续训练，所以采用从头训练方式训练了6000轮，PESQ均分提高到1.8-2.0，底噪与失真效果略有降低。

最终选择数据集二，训练6000次的模型作为最终比赛使用模型。

2.3TTS声码器

比赛中途曾尝试加载不同声码器，对比加载了wavelow和melgan两个模型，wavelow在训练出来的模型生成频谱上，相对比melgan产生了更大噪音，所以选择melgan作为最终声码器。

1. 目标检测数据收集、标注以及训练。

3.1 目标检测数据收集

比赛初期和“无辑”队伍合作对收集了口罩二分类数据集(戴口罩、不戴口罩)，图片大概8000张，然后发现比赛官方数据集为三分类数据集(戴口罩、不戴口罩、没戴好口罩)，于是对8000多张图片进行重新标注，标注完成后再次训练。

由于训练波动较大，检查label标签后发现，第三类(没戴好口罩)标签太少，不足标签整体的二十分之一，所以重新收集数据集，从头收集第三类标签的图片，然后重新归纳整理，得到了第二个数据集，数据集大概1100张图片，标签比率相对均衡。

在比赛后期Ken老师在群内发出信息，最终比赛验证数据集会有一定的数据增强，于是使用roboflow网站对两个数据集均进行进行随机数据增强，然后进行后续训练，以便得到更好的效果。

3.2 模型训练

比赛期间放弃使用TAO而是使用Pytroch做框架的Yolov5-6.0进行训练，由于本次比赛对速度提出要求，所以放弃yolox，使用yolov5，对两个数据集分别使用yolov5-s和yolov5-m进行训练，生成了四个best模型、四个last模型。

训练均为300epoch，使用batch\_size为16，由于yolov5早停机制，第一个数据集训练结束轮次为200轮和220轮，第二个数据集均完成了300轮训练。

在生成pt模型之后，我们队根据GitHub仓库：<https://github.com/wang-xinyu/tensorrtx/tree/master/yolov5> 对模型进行TensorRT文件的生成，产生了八个推理引擎文件。推理batch\_size均为1。

在6月11日完成了推理脚本，分别对预先发布的小测试数据进行测试，其结果如下：

|  |  |
| --- | --- |
| 模型名称 | 最终mAP |
| Last-m-1 | 73.38 |
| Best-m | 63.76 |
| Last-m-2 | 65.88 |
| Best-m-2 | 65.88 |
| Last-s-1 | 72.03 |
| Best-s-1 | 65.41 |
| Last-s-2 | 67.86 |
| Best-s-2 | 67.70 |

其中last、best表示yolo训练过程中保存的最终模型以及根据mAP评分最高的模型，s、m表示模型大小，1、2表示使用的数据集。

根据结果可以进行以下推断：

1. 虽然第三类标签少，但是依然有学习到新的内容，应该关闭yolo的早停，可能出现更好的效果。
2. 数据量大还是鲁棒性更好一点。

最后根据比赛情况选择了mAP最高的last-m-1作为最终模型。

### 团队介绍

辛晨 何飞



于明民 张能



马信宏

### 总结

1. 收获

这次比赛是一个完整的项目收获，虽然核心的算法、模型并不需要重点考虑其原理和思维，但是前期的数据收集、处理、标注，训练之后的结果测试，再验证，根据情况再训练，调整和优化，这些流程上也是很大的体验，特别是如何判断训练情况等等。总体而言收获很大。

1. 遗憾

这次比赛最终判定结果在ASR和TTS部分上过多的考虑自定义结果，比如ASR部分只有一句话，TTS部分则是根据自己的数据集打分，存在一种自己当裁判的情况，导致团队比赛思维过于求胜，而非求知，甚至使用脚本获取更高分数，对这种行为并不满意。希望下次比赛的时候，可以更多的注意力转移到数据收集、处理、以求在最终结果上能以堂堂正正的能力获得更高的分数。

总体而言，比赛感受很好，期待下一次比赛！