Yong Wei
Linkedin
Email : weiyong1024@gmail.com
Mobile : (+86) 186-1263-2313

Education

- **Beihang University** — Beijing, China
  Master of Engineering in Electronic and Communications — *Sep. 2016 – Jan. 2019*
- **Beihang University (Honors College)** — Beijing, China
  Bachelor of Engineering in Electrical and Electronics; GPA: 3.7 — *Sept. 2012 – Jun. 2016*

Experience

- **Pony.ai** — Beijing, China
  Software Engineer - Infrastructure — *Feb 2019 - Mar 2020*
  - **Voice Logging Pipeline**: Built the pipeline of voice logging feature, allowing operators without computer-science background to record necessary info by voice. For onboard part, first chose relatively optimal devices with the hardware team and wrote relating drivers. Then added a demon process to manage the trigger and termination of recording events. For offline part, using Speech-To-Text service by Google Cloud to extract the content of recorded sounds, serializing original audio files into Google Protocol buffers and integrated meta info into the current issue reporting pipeline.
  - **Car Sound Workflow**: Updated Car Sound system to support I18N, improving the HMI for operators and passengers outside English speaking countries. First, built an internal tool to manage car sound files and suites with three main functions: Generate voice files by calling AWS Text-To-Speech service, download from and upload to Pony storage servers, and version control. Then integrate the car sound playing feature into current onboard pub-sub system.
- **Airbnb** — Beijing, China
  Software Engineer Intern - Web Full Stack — *Jun 2018 - Sep 2018*
  - **Host Retrospect Page**: Retrospect page for Airbnb host users. Added related endpoints and mobile web pages on Ruby On Rails framework. Discussed with the designers on the contents and formats of web pages.
- **Megvii (Face++)** — Beijing, China
  Research Intern — *Jan 2018 - Jun 2018*
  - **Model Search**: During the first half, after the dimension of the input graph increased by 14%, we aimed to reduce the FLOPs (float operations per second) of the CNN (convolutional neural network) of face recognition module on mobile devices to the same level as before, without damaging the performance of the whole model which is measured by $\frac{1}{10,000}$ passing rate (ROC value when $x = \frac{1}{10,000}$). The problem was solved by adding a bottleneck layer before the Inception-ResNet module to compress channels of the input feature graph by $\frac{1}{2}$. As for the second half, we aimed to improve the performance of our face recognition module referring to state-of-the-art CNN architectures. After experiments on Xception, DenseNet and several other architectures, I proposed a modified version of Google Inception V4 for our production: Instead of using the same atom modules everywhere, different Inception-ResNet modules were used for the shallow, medium and deep part of the CNN. On the other hand, the $N \times N$ kernels of the medium and deep part were replaced by a sequence of $1 \times N$ and $N \times 1$ kernels to improve model capacity. As a result, the $\frac{1}{10,000}$ passing rate was improved by 1% on most benchmarks.
  - **Data Cleaning**: Made a bunch of original multiracial face data (including $120,000$ pictures and $6,000$ videos) into benchmarks on Linux server by writing Python scripts.
- **Beihang University** — Beijing China
  Research — *Jul 2015 - Mar 2016*
  - **The 5^{th} Generation Mobile Network**: Proposed an channel estimation method based on uplink wireless data and channel sparcity with supervisor, improving upper bound of the wireless system's throughput by 28% according to simulation results.
    Paper published in Journal of Signal Processing (First author). DOI: 10.16798/j.issn.1003-0530.2017.06.002

Competitive programming

- **Google Code Jam Kickstart 2017 Round F**
  Top 5%*, rank* 108^{th} *globally.* ***scoreboard, id:WeiYong1024***

Programming Skills

- **Languages**: C++, Python | Shell, JavaScript, **Technologies**: AWS, Google Cloud | Kubernetes, React

魏雍
领英主页

邮箱: weiyong1024@gmail.com
Mobile : (+86) 186-1263-2313

教育背景

- **北京航空航天大学** 中国，北京
  电子与通信工程硕士 *2016 年 1 月 – 2019 年 1 月*
- **北京航空航天大学 (沈元荣誉学院)** 中国，北京
  电子信息工程硕士；*GPA: 3.7* *2012 年 9 月 – 2016 年 6 月*

工作经历

- **小马智行** 中国，北京
  软件工程师 - 基础架构 *2019 年 2 月 - 2020 年 5 月*
  - **行车录音工具链**: 为实现单人单车运营，简化问题记录流程，开发了行车录音工具链。对于出车阶段的需求，首先与硬件团队合作选择适用体验最佳的硬件方案并为新硬件编写驱动，同时在车载系统上增加一个守护进程用于监听和触发录音事件。对于线下处理部分，使用谷歌云提供的语音转文字服务将行车录音的内容识别并提取出来，再通过 Google Protocol buffers 将行车录音的元信息序列化。最后使用已有的 Issue 汇报系统流程将相关问题发送给 QA 同事。
  - **车载语音系统**: 为提高运营人员和使用自动驾驶出行服务的乘客的用户体验，同时减小语音播放所占的系统计算资源，升级了原有使用 Pico TTS 工具、只支持英文的车载语音系统使之支持 I18N。在离线方面，编写内部工具用于生成、管理中英、文的车载音频语料，该工具有以下三个主要功能：利用 AWS 提供的文字转语音服务生成中、英文语料、实现与内部存储服务器之间的上传下载接口，以及语音包的版本控制。对于车载系统，修改语音播放请求在消息分发系统中的格式与声音模块的逻辑使新语音预料可以在行车过程中播放。
- **爱彼迎** 中国，北京
  软件工程师实习生 - *Web* 全栈 *2018 年 6 月 - 2018 年 9 月*
  - **房东回顾页面**: 开发房东回顾页面。在 Ruby On Rails 框架下实现相关的 Endpoints 和移动段前段页面，与设计师讨论确定前端页面的样式和内容。
- **旷视科技** 中国，北京
  炼丹实习生 *2018 年 1 月 - 2018 年 6 月*
  - **模型搜索**: 在实习期前半段，由于用输入网络的图片空间维度增加了 14%，我们希望能够在不牺牲模型性能的前提下，通过模型压缩将的 CNN(卷积神经网络) 的 FLOPs(每秒钟浮点运算量) 降低到先前的水平。这里模型性能使用万一通过率（ROC 曲线上 $\frac{1}{10,000}$ 时纵坐标的值）来衡量。最终通过对网络中 Inception-ResNet 模块加入 bottleneck 层将输入特征图层数压缩一半的方式使该问题得以解决。在实习期后半段，我们希望借鉴最新的 CNN 结构研究成果进一步提升用于移动端人脸识别模块的性能。在对一些经典网络结构如 Xception、DenseNet 进行测试以后，我提出了一种基于 Google Inception V4 的网络结构用于产品：首先, 先前网络中全部使用相同的 Inception-ResNet 模块，而我针对网络中浅层、中层、深层使用不同结构的模块。另一方面，我通过将 $N \times N$ 卷积核替换为 $1 \times N$ 与 $N \times 1$ 卷积核串联的方式增加了网络容量。最终实验结果显示，上述网络结构使得模型性能在大多数数据集上提升了一个百分点。
  - **数据清理**: 编写 Python 脚本将一批多人种人脸数据 (包括 120,000 张照片和 6,000 个视频) 整理成可用于模型训练的数据集。
- **北京航空航天大学** 中国，北京
  硕士研究生 *2015 年 1 月 - 2016 年 5 月*
  - **第五代移动通信网络**: 在导师指导下提出一种利用上行数据及信道稀疏特性提升信道估计质量的方法，使得网络吞吐量提升了 28%。
    以第一作者合作发表于中文核心期刊《信号处理》。DOI: 10.16798/j.issn.1003-0530.2017.06.002

Coding 能力

- **Google Code Jam Kickstart 2017 Round F**
  前 5%, 全球排名 108^{th}。**计分板链接（*id:WeiYong1024*）**

技术栈

- **编程语言**: C++, Python 最常用 | Shell, JavaScript 有经验 **工具**: AWS, Google Cloud | Kubernetes, React