



孙博涵

中共党员



哈尔滨工程大学 自动化类
大三年级本科生（2022年入学）



当前兴趣：大语言模型 语音处理
个人主页：[Bohan Sun](#)



山东济南
高中毕业于山东省实验中学



sunbohan856@gmail.com
18846751353（微信同号）



教育背景：国防七子211专业第一名 + 控制科学与工程学科评估为A（全国第六）

2022.09——至今 哈尔滨工程大学 智能科学与工程学院 自动化类

前五学期学习、综测成绩均排专业第一名（1/62）（2024年专业推免比例34%）

必修课加权平均分93.80分 CET4 596 / CET6 518

所学课程：工科数学分析(100) 线性代数(98) 自控原理(98) 理论力学(98) 概率论(96) 数电模电(94)



科研经历

一、基于多条件输入flow的高保真唱歌语音转换 兴趣驱动 第一作者 2024.3——2025.1

经历介绍：面对歌声转换速度慢及保真度低的问题，我们基于VITS模型（一个主流TTS系统）加入hubert-soft语音表征学习模块，且额外提取语音的音调（基频F0）和情感作为normalizing flow的输入条件，并添加基于iSTFT的解码器来加快重构波形的速度。在项目中我负责模型代码的改写与模型全部训练工作(1.使用Opensinger数据集在VCTK预训练权重基础上训练 2.消融实验研究), 并参与introduction和experiment部分的写作和demo编写整理工作。本篇论文预印本（arXiv）被字节音乐大模型Seed Music报告引用，IEEE Signal Processing Letters (SCI Q2)在投。

二、大语言模型新架构 中科院自动化所类脑智能国重 科研实习生 2024.11——至今

实习介绍：比起基于softmax attention的Transformer架构（运算复杂度与序列长度平方成正比），Linear架构(以Mamba为代表)将运算的时间与空间复杂度降至线性，但其仍然面临着recall表现差的问题。我当前熟悉了各阶段Linear attention的思路及其Chunk-wise Parallel形式的推导，学习了GLA、DeltaNet算子（推导笔记记录在CSDN:[Brock Sun](#)），熟悉Flash Linear Attention库，还学习了使用triton编写高效计算的kernal。当前正在参与构建用于SFT的AI Feedback数据集，从训练数据角度出发提升recall表现。

三、基于视觉语音的多模态情感交互数字虚拟人 国创项目 第一作者（负责人） 2024.7——至今

项目介绍：为解决当前数字人输入模态单一导致智能水平低，且情感交互能力差的问题，我们构建了视觉语音多模态输入、情感交互强的基于LLM的数字人。我本地部署了yolov8模块，并编写控制框架调用其他组件的API来构建数字人，其包含语音识别、表情识别、LLM、情感语音合成等模块，并借鉴了国内优质开源项目fay数字人的程序和人模，使得语音驱动面部，并通过Metahuman项目在UE引擎上构建3D数字虚拟人。



竞赛及荣誉

曾获全国大学生数学竞赛一等奖，美国大学生数学建模比赛H奖，及若干校级科技比赛一等奖

曾获国家奖学金（前1%）及黑龙江省三好学生、中国船级社奖学金、校优秀共青团干部、校三好学生、校优秀学生干部、校一等奖学金（多次）等等荣誉



其他

- 同时担任校吉他社社长、专业班班长、行政班团支书、预科生班副班主任、篮球班班长的职务
- 参加2024中国自动化大会（山东青岛），并在会员之夜晚宴上吉他弹唱
- 本人花两个月时间认真自学了李沐老师的动手学深度学习（pytorch版），了解深度学习与机器学习的基本概念，熟悉python，熟悉使用pytorch框架实现深度学习。