

杜嘉炜

isjiawei.du@gmail.com | [微信](#) | [个人主页](#) | [谷歌学术](#) | [Github](#)

教育经历

国立台湾大学 | 资讯工程学系 | 硕士研究生 | **GPA: 4.15/4.3** 2022.09—2025.06 (预计)

- 主要研究方向为说话人验证, 神经音频编解码器, 音频反欺骗, 视听学习。
- 专业排名前 30%, 23/24 学年度系排第一, 硕士期间在MIRLab受到张智星教授的指导。

铭传大学 | 电脑与通讯工程学系 | 工学学士 | **GPA: 3.97/4.0** 2018.09—2022.06

- 专业排名 1/79, 6 次学业奖学金, 主要研究方向为六轴机器人应用, 指导教授为江淑盈教授。

上海交通大学 | 计算机科学与工程学系 | 交换生 2020.09—2021.01

- 主修数门重要专业课程: 计算机网络, 计算机视觉, 数字信号处理, 数字图像处理。

技术能力

- 编程语言: Python, C++, C。
- 常用工具: Linux, Git, Pytorch, LaTeX, Markdown。
- 语言能力: 英文 - 雅思 6.5, 中文 - 母语。

论文清单

Empower Typed Descriptions by Large Language Models for Speech Emotion

Haibin Wu, Huang-Cheng Chou, Kai-Wei Chang, Lucas Goncalves, **Jiawei Du**, Jyh-Shing Roger Jang, Chi-Chun Lee, Hung-yi Lee

APSIPA ASC 2024

[PDF]

- 使用 ChatGPT 处理自然语言情感描述, 在 SUPERB 上提升了 3.08% 的 SER 模型性能。

DFADD: The Diffusion and Flow-Matching Based Audio Deepfake Dataset

Jiawei Du*, I-Ming Lin*, I-Hsiang Chiu*, Xuanjun Chen, Haibin Wu, Wenzhe Ren, Yu Tsao, Hung-yi Lee, Jyh-Shing Roger Jang

SLT 2024

[Github, Demo, PDF]

- DFADD 是第一个专门使用 Diffusion 和 Flow-matching 的 TTS 模型生成的欺骗性语音的数据集。
- 相较于 ASVspoof 上训练, DFADD 上训练的反欺骗模型在面对更先进方法生成的欺骗性语音时可得到更低的 EER。

Open-Emotion: A Reproducible EMO-SUPERB for Speech Emotion Recognition Systems

Haibin Wu, Huang-Cheng Chou, Kai-Wei Chang, Lucas Goncalves, **Jiawei Du**, Jyh-Shing Roger Jang, Chi-Chun Lee, Hung-yi Lee

SLT 2024

[PDF]

- 一个开源的用户友好的语音情感识别的性能基准。

Codec-SUPERB @ SLT 2024: A lightweight benchmark for neural codec models

Haibin Wu, Xuanjun Chen*, Yi-Cheng Lin*, **Jiawei Du***, Kai-Wei Chang*, Ke-Han Lu*, Alexander Liu*, Ho Lam Chung*, Yuan-Kuei Wu*, Dongchao Yang*, Songxiang Liu, Yi-Chiao Wu, Xu Tan, James Glass, Shinji Watanabe, Hung-yi Lee

SLT 2024

[PDF]

- 一种比较现有编解码器模型的公平、轻量级性能基准, Codec-SUPERB 比赛的结果汇总。

Neural Codec-based Adversarial Sample Detection for Speaker Verification

Xuanjun Chen*, **Jiawei Du***, Haibin Wu, Jyh-Shing Roger Jang, Hung-yi Lee

Interspeech 2024

[Github, PDF]

- 一种在自动说话人验证系统 (ASV) 中采用神经音频编解码器识别对抗样本的新方法, 优于 7 种之前的 SOTA 方法。
- 全面调查和分析 15 种具有不同配置的开源神经音频编解码器。

EMO-SUPERB: An In-depth Look at Speech Emotion Recognition

Haibin Wu, Huang-Cheng Chou, Kai-Wei Chang, Lucas Goncalves, **Jiawei Du**, Jyh-Shing Roger Jang, Chi-Chun Lee, Hung-yi Lee

Preprint

[Github, Demo, PDF]

- EMO-SUPERB 利用 15 个最先进的语音自监督学习模型在六个开源 SER 数据集上对 SER 进行详尽评估。
- 通过使用 ChatGPT 生成的标签，我们在所有设置中实现了 3.08% 的平均相对增益。

Dcase 2023 task 6b: Text-to-audio retrieval using pretrained models

Chung-Che Wang*, Jiawei Du*, Jyh-Shing Roger Jang

DCASE 2023 [PDF]

- Dcase 2023 task 6b 的技术汇报，使用交叉注意力以及大型预训练模型进行语音 - 字幕检索。

研究经历

针对 SingFake 数据集的歌曲深伪检测 | 实验室子项目主要负责人 2024.03—2024.06

- 负责设计了基于高频信息以及 Codec 增强的 Wav2Vec2 - Conformer 检测方法。
- 在 SingFake 的 CtrSVDD 数据集上的 EER 相对于基线降低了 50% 以上。

视听深度伪造检测 | 台湾地区国科会子项目 2023.09—2023.11

- 探索当前最先进的视听 Deepfake 检测方法的局限性, 使用 LoRA 微调现有的视听 Deepfake 模型。
- 将我们的私有 AV Deepfake 数据集上的 mAP@50 从 13.1% 提高到 75.7%。

虚假音频检测 | 台湾地区国科会子项目 2023.07—2023.09

- 研究并应用了近五年来几乎所有的先进方法，包括 RawNet2、AASIST、Wav2Vec2-AASIST 等。
- 收集了高仿真度的深度伪造资料集，使用方法包括但不限于 StyleTTS, YourTTS, so-vits-svc 等。

音频文本跨模态学习 | Dcase 2023 比赛项目 2023.03—2023.06

- 重新设计并探索了不同预训练模型的性能语音文本检索，并尝试不同各种数据增强方法。
- 在 Dcase2023 6b 项目中排名第三。

个人项目经历

基于 YOLOv5 的 CSGO 作弊程序 | 深度学习课程设计与实现 2022.03—2022.05

- 标注了 800 张 CSGO 角色的头部和身体，并且使用 YOLOv5 对其进行训练。
- 设计了 bounding box 和鼠标的配对算法实现自动锁定功能。

针对基于有噪音的 OCT 医学影像增强算法的研究 | 数字图像处理设计与实现 2020.11—2021.01

- 负责比较不同算法中对 OCT 图像进行去噪的能力（侧窗过滤；非局部均值；BM3D；从单个图像学习自监督去噪）。
- 通过迭代去噪过程增强图像质量，并通过 PSNR（峰值信噪比）评估每种算法的增强效果。

学术活动

- 2024 International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI 2024) 审稿人。
- 2024 IEEE Spoken Language Technology Workshop (SLT 2024) 审稿人。
- Codec-SUPERB 比赛 技术支持。

在校奖项

- SLT 2024 官方会议奖助金（10 个名额，共 42 名申请者）。
- SingFake 挑战赛 CtrlSVDD 比赛中排名 13/49（总共 74 位参赛者, 130 个结果提交）。
- DCASE2023 task 6b 音频文本检索挑战赛排名 3/10。
- 大学时期 6 次系排第一奖学金，1 次班排名第一（系排名第二）奖学金。

校外活动

铭传大学陆生会 | 总务 2019.09—2020.01

- 负责总体预算规划和执行、采购、筹款、与大学行政部门和学生的沟通，规划和实施全年校园活动。

湖北在台陆生会 | 活动部干事 2018.01—2019.06

- 负责湖北籍在台陆生的集体活动规划。