

# LE ZHUO

(+86) 13009522410 • zhuole1025@gmail.com • GitHub • Twitter

## EDUCATION

**Beihang University**, Beijing, China

09/2019 – 07/2024 (expected)

B.Eng. in Computer Science and Technology, GPA: 3.8/4

Core Course: Linear Algebra (93/100), Mathematical Analysis(2) (99/100), Probability and Statistics (96/100), Discrete Mathematics(2) (100/100), Mathematical Modelling (100/100), Data Structure and Programming (100/100), Computer Organization (93/100), Algorithms (98/100), Signals and Systems (95/100)

## PUBLICATIONS

- **Le Zhuo**, Ruibin Yuan, Jiahao Pan, Yinghao Ma, Yizhi Li, Ge Zhang, Si Liu, Roger Dannenberg, Jie Fu, Chenghua Lin, Emmanouil Benetos, Wenhui Chen, Wei Xue, and Yike Guo, “LyricWhiz: Robust Multilingual Lyrics Transcription by Whispering to ChatGPT,” **ISMIR’23**, [Arxiv] [GitHub]
- **Le Zhuo\***, Zhaokai Wang\*, Baisen Wang\*, Yue Liao, Stanley Peng, Chenxi Bao, Miao Lu, Xiaobo Li, and Si Liu, “Video Background Music Generation: Dataset, Method and Evaluation,” In Submission to **ICCV’23**, [Arxiv]
- **Le Zhuo\***, Qiaosong Qi\*, Aixi Zhang, Yue Liao, Yongliang Wang, Xiaobo Li, and Si Liu, “DiffDance: Cascaded Motion Diffusion Model for Music-driven Dance Generation,” In Submission to **MM’23**
- Ruibin Yuan, Yinghao Ma, Yizhi Li, Ge Zhang, Xingran Chen, Hanzhi Yin, **Le Zhuo**, et al., “MARBLE: Music Audio Representation Benchmark for Universal Evaluation,” In Submission to **NeurIPS’23**, [Arxiv] [Page]

## RESEARCH INTERESTS

**Fields** Generative Models, Multi-modal Learning (e.g., Music, CV), AI for Science

**Models** Diffusion Models, Energy-based Models, Normalizing Flows, GANs, Transformers

## RESEARCH EXPERIENCE

**Mila - Québec AI Institute**

Montreal, Canada

Research Intern, Advisor: Prof. Jian Tang

06/2023 – ongoing

- Building an LLM-based protein foundation model to provide a unified perspective for protein and language tasks by treating protein as a foreign language and aligning protein-centric tasks with language tasks
- Working on diffusion generative models to solve rigid antibody-antigen docking problem

**Beijing Academy of Artificial Intelligence**

Beijing, China

Research Intern, Advisor: Dr. Jie Fu

01/2023 – 05/2023

- Built the MARBLE benchmark, a fair, reproducible, and extendable eval suite for general music understanding, covering 9 pre-trained models and 14 downstream MIR tasks on 8 public datasets
- Designed the first automatic lyrics transcription system that can perform zero-shot, multilingual, long-form lyrics transcription by integrating two advanced models from OpenAI – Whisper and GPT4
- Introduced the first large-scale, weakly supervised, and copyright-free multilingual lyric transcription dataset, MulJam, consisting of 6,031 songs with 182,429 lines and a total duration of 381.9 hours

**Alibaba Inc.**

Beijing, China

Research Intern, Advisor: Dr. Aixi Zhang

08/2022 – 12/2022

- Proposed a cascaded diffusion model with transformer-based architecture for music-to-dance generation
- Aligned CLIP embedding space of music and motion to boost conditional generation with classifier-free guidance
- The proposed model can generate long-form (20+ seconds) and high-resolution (60 FPS) dance sequences that conform to body constraints and have global structure given the input music, achieving state-of-the-art FID

**Beihang University, CoLab**

Beijing, China

Research Intern, Advisor: Prof. Si Liu

09/2021 – 08/2022

- Introduced SymMV, the first video and symbolic music dataset with detailed annotations
- Proposed a benchmark framework with transformer-based architecture that utilizes music priors of chords, melody, and accompaniment along with video-music relations for video background music generation
- Designed an objective metric based on the video-music CLIP model to evaluate video-music alignment

## SELECTED PROJECTS

---

### Anime Character Generation and Editing

10/2022 – 02/2023

- Collaborated with Lambdanalytique Inc. to develop AI illustration tools that can perform controllable anime character generation and editing by integrating advanced techniques, such as Dreambooth, ControlNet, etc.
- Implemented an image-to-live2d pipeline to create live2d characters with dynamic expressions conditioned on the single image input, which has been introduced to game artists to reduce their workload and leave them more time to design instead of wasting time on drawing details

### Diffusion-based Out-of-Distribution Detection

05/2022 – 08/2022

- In-depth investigation and reproduction of state-of-the-art diffusion models and energy-based models
- Observed diffusion models tend to assign higher likelihood to out-of-distribution (OOD) data, and explained this phenomenon by splitting the backward diffusion process into sampling stage and denoising stage
- Proposed reconstruction-based approach for OOD detection and achieved 0.95+ AUROC in CIFAR10 vs. SVHN

## SELECTED AWARDS

---

- |  |                  |
|--|------------------|
| • <b>National Scholarship</b> (top 0.1% national-wide)   | 2020             |
| • <b>Outstanding Students</b> , Beihang University (top 5%)  | 2020             |
| • <b>First Prize</b> in the 32 <sup>th</sup> <b>Fengru Cup Competition</b> , Beihang University (top 5%) | 2022             |
| • <b>Scholarship of Academic Records</b> , Beihang University (top 5%)                                   | 2020, 2021, 2022 |

## SKILLS

---

**Programming Languages:** Python, C/C++, Java, Bash, Verilog, Assembly (ranked by proficiency)

**Libraries and Tools:** Git, L<sup>A</sup>T<sub>E</sub>X, PyTorch, TensorFlow, Matplotlib, Unix, Docker

**Languages:** College English Test - Level 6: 688/710 (**top 0.1%** in Beihang University)