WANG ZHENG

(+86)13931805565 • zheng-wa22@mails.tsinghua.edu.cn • github.com/Cheeseofneo

EDUCATION

USTEP, Electrical Engineering and Information System, Tokyo University Oct 2024 - Mar 2025
Lab:TAKEUCHI LAB, research about efficent diffusion model

M.S.C., Electrical Engineering, Tsinghua University Sep 2022 - Jun 2025
Lab:OMIS(Optical Measurement and Imaging Systems) Lab 3.82/4.0 GPA

B.S.E., Electrical Engineering, Tianjin University Sep 2018 - Jun 2022
Lab:MNMT(Micro-Nano Manufacture and Technology) Lab 3.72/4.0 GPA

HONORS

- Liu Bao Scholarship(2019)
- Andon Trust(2020)
- Merit Student(2021)
- ICM Finalist(Top 1%, 2021)
- Membership of China Instrument and Control Society(2022)
- Tsinghua College Scholarship(2024)
- NVIDIA AI-Agent Group(2024)

PUBLICATIONS

Wang, Z., Ma, R., Zeng, C., Liu, L., Li, X., Wang, X., & He, B. A fast and precise autofocus method using linear array CCD. In Optical Metrology and Inspection for Industrial Applications XI (Vol. 13241, pp. 446-456). SPIE.

Li, Z., Su, Y., Yang, R., Xie, Z., Wong, N., & Yang, H. Quantization Meets Reasoning: Exploring LLM Low-Bit Quantization Degradation for Mathematical Reasoning. arXiv preprint arXiv:2501.03035. (**Under submission**)

ACADEMIC PROJECTS

Low-bits Denoising Diffusion Models for Masked Images Nov 2024-Mar 2025
TAKEUCHI Lab, University of Tokyo

- Accelerate the diffusion process in two aspects: quantize the model parameters (W1A1) to save memory and make the model learn general representation of images to reduce the sample steps.
- Analyze the effects of prompt information in different types of noise on the BiDM(Zheng et al.) model and reasoning by decomposition(PCA & SVD) and redistribution of basic noise components.
- (Still Working) Is the representation ability of LDM related to the synthesized quality by masking learning?

Exploring LLM Low-Bit Quantization Degradation for Mathematical Reasoning Dec 2024-Feb 2025
Cooperated with Prof.Yang, HKPolyU

- Systematically evaluate the impact of quantization on mathematical reasoning tasks. We mainly test the GPTQ and AWQ on Llama-3 models.
- Introduce a multidimensional evaluation framework combining qualitative capability analysis and develop math ability recovery strategies by stepwise DPO method.

A fast and precise autofocus method using linear array CCD June 2023-Oct 2024
Master's thesis: **laboratory of measurement and computer vision**

- Design an active autofocus microscope using linear array CCD whose accuracy at the nanometer level and speed at the millisecond level. Develop a focus calibration algorithm, which conducts weighted fusion to the results.
- Propose a centroid extraction algorithm at sub-pixel level based on multiscale feature extraction, and fit linear mapping using sliding finite impulse response filter.

Deep Learning on Small Sample Pointsets for 3D Segmentation Sep. 2021-May 2022
Bachelor's thesis: **State key laboratory of precision measuring technology and instruments**

- Based on a sample-aware data augmention infrastructure, proposed a novel network for point segmentation **PASN)**. Adjusted feature extractor- PointNet for degeneracy to fit our specific segmentation task.

- Propagate segmentation loss to augmentor and regress pointwise and shapewise matrixs for input samples. Experiments proved accuracy of segmentation increase by **5%** and higher robustness than PointNet on our dataset

INTERNSHIP EXPERIENCE

Baidu, Beijing China: AI Research Intern,*Team Apollo* July 2024- Oct 2024

- For multisensor perception model, propose an infra of model fusion based on shared weights and backbone fusion.. Also conducts a cross training strategy for multi-objection and utilize distillation to compensate accuracy.
- For bev2instance module in bev perception model, conduct novel optimization the self-attention and deformable attention modules based on precision descent and infra compression and prove efficient in training accuracy.

Momenta, Shenzhen China: System Archicture Intern, *Team Middleware of System* March 2023-July 2023

- Design the timestamp synchronization architecture of ADAS system, and construct simulated signal generator in **signal simulation platform**, which involved validating signal pathways and creating abnormal signals.
- Developed an testbench-**board verification platform** based on SOMEIP about resource schedulng of hardware, the signal accuracy and communication correctness on a high-concurrence scene.

OTHER WORK EXPERIENCE

ICM 2021, Finalist Award(top 1%): Jan 2020 - Mar 2020

- Proposed **PRE(Page-Rank Entropy) model** vector network of musicians and quantify influence of musicians by weights of nodes at the time-span and genre-span perspectives.
- Construct local and global similarity evaluation model based on cosine simlarity of multi-dimensional eigen vectors extracted by *PCA*. Utlized k-means clustering method to classify significance of eigens between and within genres.
- Adopted *silding window automatic regression model* as time series forecasting on main eigens to predict the long and short-term developing trend of musical genres.