Zefang Wang | Master Student

Zhejiang University
📱 (+86) 18234041863 • ✉ zefangwang@zju.edu.cn
🌐 aden9460.github.io/Zefang-Wang

Education

Zhejiang University **Hangzhou, China**
Master of Engineering *2023.09–2026.06*

- Major: Control Engineering
- Advisor: Dr. Guanzhong Tian
- Ranking: Top 10%

North University of China **Taiyuan, China**
Bachelor of Engineering *2019.09–2023.06*

- Major: Rail Transit Signal and Control
- Ranking: Top 1%
- Honor: Outstanding Graduate

Research Experience

ENCODE Lab, Westlake University **Hangzhou, China**
Visiting Student *2025.03–now*

- Advisor: Prof. Huan Wang
- Conducted research on iOS mobile model deployment
- Developed compression methods for visual autoregressive models

Publications

Under Review
EVAR: Edge Visual Autoregressive Models via Principled Pruning, First Author

- Proposed a principled OBS-based structured pruning method for visual autoregressive models
- Introduced progressive scale-aware distillation to address gradient imbalance during next-scale autoregressive fine-tuning
- Achieved 1.8× speedup with only 10% quality loss on edge devices for single-image generation
- Project Page: `https://aden9460.github.io/EVAR/`

Under Review
OBS-Diff: Accurate Pruning For Diffusion Models in One-Shot, Fourth Author

- Proposed an OBS-based training-free multi-granularity pruning method for diffusion models
- Introduced timestep-aware Hessian construction with logarithmically decreasing weights
- Minimized visual quality degradation while accelerating inference
- Project Page: `https://alrightlone.github.io/OBS-Diff-Webpage/`

IEEE TNNLS 2024 (Accepted)
An Effective Information Theoretic Framework for Channel Pruning, Second Author

- Proposed an information-theoretic framework using entropy and rank fusion for layer-wise pruning rates
- Introduced Shapley value-based contribution evaluation as the intra-layer pruning criterion
- Demonstrated effectiveness in image classification and object detection tasks
- arXiv: `https://arxiv.org/abs/2408.16772`

Research Interests

Model Compression, Network Pruning, Knowledge Distillation, Diffusion Models, Visual Autoregressive Models, Edge AI Deployment, Binary Neural Networks