

# 范云潜

✉ pannenets.f@foxmail.com

in Yunqian Fan

🌐 PannenetsF

🌐 <https://pannenetsf.com/>



## 教育经历

- 2018 – 2022    **北京航空航天大学, 微电子科学与工程专业 学士, 北航优秀毕业生**  
专业 GPA: 3.8/4.0 (班级排名 1/19)  
毕业论文题目: 存内计算量化技术研究 导师: 康旺副教授, 北京优秀毕业论文奖
- 2022 – 2025    **上海科技大学, 计算机科学与技术专业 硕士**  
专业 GPA: 3.9/4.0  
研究方向: 模型压缩和机器学习分布式系统加速。导师: 范睿副教授

## Research Publications

### Conference Proceedings

- 1 Y. Fan, X. Wei, R. Gong, Y. Ma, X. Zhang, Q. Zhang and X. Liu, "Selective focus: Investigating semantics sensitivity in post-training quantization for lane detection," in *The 38th Annual AAAI Conference on Artificial Intelligence AAAI*, 2024.
- 2 J. Bai, Y. Fan, S. Sun, W. Kang and W. Zhao, "Tiny neural network search and implementation for embedded fpga: A software-hardware co-design approach," in *2021 IEEE Asian Solid-State Circuits Conference (A-SSCC)* IEEE, 2021, pages 1–3.
- 3 H. Zhang, J. Liu, W. Kang, Y. Fan, S. Fu, J. Bai, B. Pan, Y. Liu and W. Zhao, "A 40nm 33.6 tops/w 8t-sram computing-in-memory macro with dac-less spike-pulse-truncation input and adc-less charge-reservoir-integrate-counter output," in *2021 IEEE International Conference on Integrated Circuits, Technologies and Applications (ICTA)* IEEE, 2021, pages 123–124.

### Journal Articles

- 1 J. Bai, W. Xue, Y. Fan, S. Sun and W. Kang, "Partial sum quantization for computing-in-memory based neural network accelerator," *IEEE Transactions on Circuits and Systems II: Express Briefs*, 2023.

### Copyrighted Patents

- 1 Y. Fan, C. Liu, J. Xu, H. Zhang, W. Kang and B. Pan, *Automatic adaptation method and device for neural networks*, Patent of China. Application Number: 202110399619.4 Patent Number: 202208300209750, 2022.
- 2 Y. Fan, C. Liu, J. Xu, H. Zhang, W. Kang and B. Pan, *General parallel inference acceleration structure and inference accelerator design for ai*, Patent of China. Application Number: 202110399639.1 Patent Number: 2022071902253770, 2022.

技能

- 硬件编程 Verilog HDL, Xilinx HLS C++, FPGA
- 软件编程 C, C++, Python, RUST, MATLAB
- 研究方向 机器学习模型硬件部署和加速 (FPGA, GPU), 后训练模型压缩 (无标签模型压缩), 大模型推理系统 (量化, Kernel, 调度)

实习经历

- 2021 - 2024 见习算法研究员商汤研究院, 模型工具链团队
  - 基于计算图的 DNN 多平台统一量化部署工具, 被 OpenMMLab 等开源项目采用。MQBench。
  - 自动驾驶检测模型的量化研究, 提出基于特征的物理语意进行高效量化。AAAI 2024
  - 通过 Kernel 优化和系统调度策略进行大模型推理系统加速。lightllm

获奖经历

- 2024 最佳模型效率奖和排名 2/36 AI4S Cup LLM 挑战赛大模型提取 “基因-疾病-药物” 知识图谱, 北京智能科学研究院
- 排名 2/39 AICAS 2024 挑战赛 (初赛阶段), IEEE 电路与系统协会
- 未来之星奖商汤科技, 最佳实习生奖
- 2022 沈元奖章提名奖北京航空航天大学, 北航本科生最高荣誉
- 北航优秀毕设论文, 北京航空航天大学
- 北京市优秀毕设论文, 北京市教育委员会
- 商汤奖学金, 商汤科技. 中国大陆筛选 30 名进行 AI 方向研究的本科生
- 2021 Meritorious 美国大学生数学建模大赛, 数学及其应用联合会
- 一等奖北京市大学生集成电路设计大赛, 北京市教育委员会
- 企业特别奖和二等奖全国 EDA 精英挑战赛, 中国电子学会
- 2019-2022 一等奖学金, 北京航空航天大学, 社会工作、学业成绩、科研竞赛等奖项均有斩获
- 2019 校级优秀生北京航空航天大学
- 一等奖北京市大学生数学竞赛, 北京市数学会