Dewen Zeng

dzeng2@nd.edu +1-574-220-6058
https://zengzhezz.github.io/dewenzeng/

Education

University of Notre Dame *Notre Dame, USA*
Ph.D. in Computer Science and Engineering *August 2019 - May 2025*

Huazhong University of Science and Technology *Wuhan, China*
M.S. in Electronic Engineering *September 2016 - June 2019*

Huazhong University of Science and Technology *Wuhan, China*
B.S. in Optoelectronic Information Engineering *September 2012 - June 2016*

Research Interest

Large Language Model (LLM), Self-supervised Learning, Image Recognition, Federated Learning, On-device Learning

Experience

Google Brain *June 2022 - March 2023*
Student Researcher *Mountain View, CA*

- Developed an LLM framework for allocating varying levels of computation to individual tokens within a large language model during pre-training, resulting in a 25% performance gain in 1-shot learning with minimal additional computational overhead compared to standard GPT.
- Designed a differentiable routing mechanism that enables the selective skipping of certain self-attention and FFN layers, offering precise control over performance and computation.
- Implemented this method in JAX to dynamically gather and scatter non-skipped tokens, resulting in an 18% training speed enhancement on TPUs.

Allen Institute for Cell Science *May 2021 - August 2021*
Machine Learning Research Intern *Seattle, WA*

- Explored various uncertainty estimation methods for nucleus segmentation in 3D cell images.
- Developed a segmentation failure detection system with a 0.9 F1 score based on the patch-based entropy uncertainty to automatically detect and locate potential segmentation failures.

Boston Children's Hospital *October 2020 - May 2021*
Visiting Scientist *Boston, MA*

- Developed a temporal contrastive learning framework to learn representations from unlabeled chest X-rays to improve the performance of lung and heart segmentation models with limited labels.
- Designed a technique to automatically extract lung water information from segmented chest X-rays captured at various time points, facilitating continuous health monitoring.

Technical skills

Programming Languages/Tools Python, Java, C/C++, Matlab, Javascript, LaTeX, Pytorch, Tensorflow, Scikit-Learn, Git, AWS Cloud

Awards

2023 Second prize, Tiny and Fair ML Design Contest at ESWEEK
2023 CSE Outstanding RA, University of Notre Dame
2020 Top Winning Award, IEEE SERVICES Hackathon
2016 Outstanding undergraduate, Qiming College, HUST
2015 Second prize, National Undergraduate Electronics Design Contest (NUEDC)