YUANMING (ALEX) TAO

ymtao5219.github.io • ytao5219@mail.com • (413) 428-8035 • Amherst, MA

Education

University of Massachusetts, Amherst **Sep 2022 – May 2024 (Expected)**

Master of Science in Computer Science, GPA: 3.79/4.0, GRE: V158 + Q167 + AW3.5 Amherst, MA

- Courseworks: Responsible AI, Software Systems, 3D Computer Vision, NLP, Information Retrieval, etc.

University of Sydney **Jul 2015 – Apr 2020**

Bachelor of Science (Honours) in Applied Mathematics, First Class Honours Sydney, Austrlia

- Thesis: "Topic modeling as a community-detection problem", supervised by Prof. Eduardo G. Altmann
- Courseworks: Bayesian Inference, Non-Linear Optimization, Stochastic Processes, Generalized Linear Models, PDEs, Computational Mathematics, Measure Theory, Abstract Algebra, etc.
- Awards: Early Researcher Scholarship to encourage undergraduate student research in complex systems
- Presented at Conference on Complex Systems, Oct 2019, NTU, Singapore

Publications

On Surgical Fine-tuning for Language Encoders **EMNLP Findings, 2023**

Abhilasha Lodha^, Gayatri Belapurkar^, Saloni Chalkapurkar^, **Yuanming Tao^**, and 4 others, ^: **Equal Contribution**

Multilayer Networks for Text Analysis With Multiple Data Types **EPJ Data Science, 2021**

Charles C. Hyland, **Yuanming Tao**, …3 others, and Eduardo G. Altmann

Experience

UMass Graphics Lab **Jun 2023 – Present**

Summer Research Intern, Prof. Evangelos Kalogerakis Amherst, MA

- Devised methods that gather user input through drag-based warping, enabling intuitive geometry and human pose editing in 2D/3D diffusion models [*diffusers*, *PyTorch*, *Flask*, *kornia*, *OpenCV*, *WebGL*, *Slurm*]
- Preparing manuscripts for submission to conferences in vision and graphics in 2024

Microsoft **Feb 2023 – Jun 2023**

Graduate Student Researcher Remote

- Developed techniques for surgical fine-tuning of LLMs based on a data-driven criteria to automatically identify and tune a smaller subset of layers using only 100 target data samples [*Huggingface*, *PyTorch*, *Slurm*, *DeepSpeed*]
- Outperformed or achieved similar performance as full model fine-tuning on nearly all GLUE and SuperGLUE downstream tasks, while potentially decreasing the time taken by 25%

Huawei **May 2020 – Dec 2021**

Software Engineer Hangzhou, China

- Implemented efficient algorithms and data structures to enable ray tracing on mobiles (282K triangles, rendered at 46 ms per frame on Kirin 9000 CPU and Mali G78 GPU) [*C++*, *Vulkan API*, *Gradle*, *Jenkins*]
- Crafted an algorithm to synthesize time-varying material and integrated it into SceneKit SDK [*C++*, *OpenCL*, *GLSL*]

Complex Systems and Data Science Group at USYD **Feb 2019 – Apr 2020**

Research Assistant, Prof. Eduardo G. Altmann Sydney, Australia

- Developed a Bayesian mutli-modal model to find topics from collections of documents (e.g. 10k of Wikipedia articles) in the presence of metadata and hyperlinks [*sklearn*, *SpaCy*, *graph-tool*, *mediawiki API*]
- Assessed and compared various models' performance using normalized mutual information, resulting in notable enhancements (~21%) in model effectiveness

Skills

- Programming: C/C++, JavaScript, Python, JAVA, C#, Bash, R, SQL, MatLab
- AI: Diffusers, PyTorch, Scikit-learn, OpenCV, TensorFlow, LangChain, DeepSpeed, Optuna
- Graphics: OpenGL, GLSL, kornia, Unity, GLSL, Vulkan
- Web: React, NodeJS, Flask, MongoDB
- Devops/Tools: Docker, Git, Jenkins