Mathieu Tuli

Research Direction

My research experience has centered around generative media. My current focus is on **generative image and video modeling**, developing techniques for interactive user experiences. I believe humans are primarily visual reasoners, and I hope to explore that thesis in developing products and higher forms of intelligence through image/video. My research to date covers the following topics:

- Large-scale Video Models: Open-Source implementation of Meta's MovieGen model and training/inference pipeline.
- Video Diffusion: Temporally consistent portrait animation using morphable multi-view diffusion models. (P11)
- Image Diffusion: Generative, reference-driven image retargeting using diffusion models. (P10)
- 3D Reconstruction: Reconstructing 3D facial animation sequences from 2D videos. (P8)
- Efficient Autoregressive Animation: Building real-time, photorealistic digital human solutions. (P7, P9)
- Text-Based Agents: Imbuing text-based agents with robust instruction following. (P5)
- CNN Optimization: Frameworks for optimizing CNNs for digital pathology and image classification. (P1, P2, P4)

Education

MSc in Computer Science · University of Toronto (Vector Institute Affiliate) With High Distinction - cGPA: 4.0/4.0 Dissertation: Learning to Follow Instructions in Text-Based Games Advisors: Sheila McIlraith and Scott Sanner

BASc in Computer Engineering · University of Toronto

With Honours - cGPA: 3.61/4.0

Experience

Senior Al Research Scientist · LG Electronics · Toronto Al Lab (Digital Media Team)01/2024 - PresentIn my current role as a Senior Al Research Scientist, I am focused on image/video models for personalized generativemedia and digital humans. I am responsible for establishing and executing high-impact foundational research projects.

- Researching 2D panoramic scene reconstruction from sparse references with spatially aware diffusion inpainting.
- Developed a state-of-the-art, large-scale multi-view morphable diffusion model for 2D portrait avatar animation and constructing controllable, 4D Gaussian avatars for real-time applications.
- Built a reference-driven inpainting diffusion model for generative image retargeting.
- Conducted large-scale synthetic data generation for training foundational 3D facial animation models using diffusion.
- Designed a SoTA feedforward 3D animation model, capable of driving photorealistic 3D Gaussian Avatars for realtime interactive experiences, and achieving 46x faster inference than baselines (P9).
- Implemented scalable ML infrastructure for training large-scale generative models, including training/testing pipelines, automated evaluation frameworks, and shared codebases for streamlined product integration.

Al Research Scientist · LG Electronics · Toronto Al Lab (Digital Media Team) 11/2022 - 12/2023

- Developed a novel face tracker based on UV-space optical flow to reconstruct 3D meshes from 2D videos. Published at CVPR 2024 (P8) and beats SoTA by as much as 54% on existing benchmarks.
- Scaled 3D facial animation training to multi-identity datasets through automatic speaker identification, achieving 15% improvement over SoTA with zero-shot generalization (P7).
- Built a digital human product pilot, featuring real-time streaming & orchestrated models with ONNX Runtime, in C++.

Graduate Machine Learning Researcher · University of Toronto

- Developed a neuro-symbolic framework (paper published at NeurIPS, P5) enabling transformer-based language models to follow temporally extended instructions and improve task completion in RL environments.
- Improved instruction following accuracy by 25%, showing value in teaching agents the temporal semantics of goals.
- Developed LLM prompt engineering techniques to generate these symbolic representations automatically.

Machine Learning Research Scientist · University of Toronto Multimedia Lab

- Published works at CVPR and ICCV covering topics in machine learning optimization for CNNS (P4, P2, P1).

Machine Learning Research Engineer · ANVIL AI

- Built AI systems to extract customer insights from unstructured data through video and audio processing.

Projects and Services

Open-Source MovieGen · GitHub	Present
Open-source implementation of Meta's large-scale MovieGen video model, including training and inference	. Includes
all technical innovations, such as the Temporal Autoencoder, Llama-inspired Transformer backbone, and	all parallel
optimizations. Actively pre-training on WebVid.	
Conference Reviewer	active
CVPR (2022, 2023, 2024, 2025), NeurIPS (2023, 2024), ICML (2024, 2025), ICLR (2025), AAAI (2025).	
Learning	Present
Self-learning CUDA, primarily through the textbook 'Massively Parallel Processor Programming.'	
jarvis.nvim · GitHub	07/2024
Lua plugin with popup-ui for LLM-prompting and chatting directly in Neovim.	
Aligning Language Models · paper	11/2021
Investigated the use of constrained beam search on language models for improved dialogue state tracking.	

Skills

Large-Scale ML	Distributed training and inference on multi-node multi-GPU clusters: DDP, FSDP, ten- sor/sequence parallelism.
Generative AI	Large-scale Diffusion models (image, video) and Transformers (flow matching, LLMs, autoregressive animation). Hands-on, daily research experience coding architectures, build- ing training/testing codebases for various tasks, and optimizing distributed systems.
Programming Languages	Python, C, C ⁺⁺ , and Go. Experience in SQL and Solidity.
ML-Related Frameworks	PyTorch, CUDA, Triton, Hugging Face, and ONNX (+Runtime).
Game Dev./Graphics	3D Geometry and Video/Image Analysis. OpenCV, Unreal Engine, Blender, and OpenGL.
Project Management	Experienced in long-term project planning and cross-functional collaboration in AI research environments.
Research Leadership	Proven track record leading product-focused foundational research groups and projects leading to successful publications and product demos.

09/2020 - 10/2022

05/2018 - 08/2019

05/2020 - 08/2020

Research [Google Scholar]

- P11 *CAP4D: Creating Animatable 4D Portrait Avatars with Morphable Multi-View Diffusion Models*. Preprint. Felix Taubner, Ruihang Zhang, Mathieu Tuli, David B. Lindell.
- P10 Generative Reference-Driven Image Retargeting. Preprint. Mathieu Tuli, Kaveh Kamali, David B. Lindell.
- P9 FaceFormer++: An Efficient and Robust Audio-to-3D Facial Animation Model. Preprint. Eu Wern Teh, Mathieu Tuli, Prashant Raina, David B. Lindell.
- P8 3D Face Tracking from 2D Video through Iterative Dense UV to Image Flow. Conference on Computer Vision and Pattern Recognition (CVPR). Felix Taubner, Prashant Raina, **Mathieu Tuli**, et al. June 2024.
- P7 A Study of 2D-Augmented 3D Speech-Driven Face Animation. arXiv. Mathieu Tuli, Felix Taubner, et al. March 2024.
- P6 *PrismAvatar: Real-time animated 3D neural head avatars on edge devices.* arXiv. Prashant Raina, Felix Taubner, **Mathieu Tuli**, et al. March 2024.
- P5 Learning to Follow Instructions in Text-based Games. Conference on Neural Information Processing Systems (NeurIPS). Mathieu Tuli, Andrew C. Li, Pashootan Vaezipoor, Toryn Q. Klassen, Scott Sanner, Sheila A. McIlraith. December 2022.
- P4 *Exploiting Explainable Metrics for Augmented SGD*. Conference on Computer Vision and Pattern Recognition (CVPR). Mahdi Hosseini*, **Mathieu Tuli***, Konstantinos Plataniotis (2022). June 2022.
- P3 Instruction Following in Text-Based Games. 3rd Wordplay: When Language Meets Games Workshop, at NAACL. Mathieu Tuli, Andrew C. Li, Pashootan Vaezipoor, Toryn Q. Klassen, Scott Sanner, Sheila A. McIlraith. July 2022.
- P2 *Towards Robust and Automatic Hyper-Parameter Tuning*. Optimization for Machine Learning Workshop, at NeurIPS. **Mathieu Tuli**, Mahdi Hosseini, Konstantinos Plataniotis. December 2021.
- P1 *CONet: Channel Optimization for Convolutional Neural Networks*. NeurArch Workshop, at ICCV. Mahdi Hosseini, Jia Shu Zhang, Zhe Liu, Andre Fu, Jingxuan Su, **Mathieu Tuli**, Konstantinos Plataniotis. October 2021.

(*) denotes equal contribution