Vikram Voleti

Research Scientist at Stability AI; PhD candidate at Mila; former Research Intern at Google, Unity, Meta

voletiv.github.io | vikram.voleti@gmail.com | Google Scholar | LinkedIn

Expertise

Deep learning for image, video, 3D: expert at machine learning research and development; experienced in leading multiple projects collaborating with international partners in industry and academia. **Projects** include:

- Video prediction/generation using denoising diffusion models [1], deriving non-isotropic diffusion models [2]
- Text to 3D using dreamfusion, NeRF, DMTet; 3D human pose estimation and inverse kinematics [3]
- Image generation using normalizing flows [4][9]; video generation using Neural ODEs [13], GANs [15][16]
- Contributed to projects on 4D generation, simulation [10], fairness/uncertainty [5], federated learning [6]

Education

Mila, University of Montreal, Canada — ***Sep 2018 - present (Aug 2023)***

Ph.D. in Computer Science — *Supervisor*: Prof. Christopher Pal

Conditional generative models for image, 3D animation, video [1][2][3][9][13]

Indian Institute of Technology (IIT), Kharagpur, India — ***2009 - 2014***

Dual Degree (**B.Tech. (Honours) + M.Tech.**) in Electrical Engineering with Master's specialization in Instrumentation and Signal Processing [17]

Work Experience

Stability AI, Canada (Remote) — Research Scientist — ***Apr 2023 - present***

- Leading AI research and development on generating 3D objects, images, videos from text

Meta (formerly **Facebook**), Menlo Park, USA — Research Intern — ***Aug 2022 - Feb 2023***

Team: AI for Metaverse (AI4RL); *Supervisors*: Dr. Yashar Mehdad, Dr. Barlas Oguz

- Led the technology development for generating 3D objects, videos from text; dreamfusion, NeRF
- Applied expertise at neural graphics for 3D rendering; implemented hands-on in PyTorch
- International AI team; technology transitioned into a Meta end product, adopted by other teams

Unity Technologies, Montreal, Canada — MITACS Research Intern — ***Oct 2021 - Aug 2022***

Team: Deep Pose, Unity Labs; *Supervisor*: Dr. Boris Oreshkin

- Built AI-assisted user-editable 3D character animation workflow; trained novel 3D human pose prior
- Published at SIGGRAPH Asia [3], incorporated technology into a Unity product

Google, Mountain View, USA — Research Intern — ***Sep-Dec 2019***

Team: Google AI Perception; *Supervisors*: Dr. Bryan Seybold, Dr. Sourish Chaudhuri

- Investigated the scope of deep semi-supervised learning for active speaker detection in video
- Hands-on implementation in TensorFlow; collaborated with TPU team to code Neural ODE in Jax

IIIT Hyderabad, India — Research Fellow — ***May 2017 - Aug 2018***

- *Supervisors*: Prof. C. V. Jawahar, IIIT-Hyderabad, Prof. Vinay Namboodiri, IIT Kanpur
- Synthesized videos in Indian languages using GANs; developed automated video dataset pipeline
- Full paper published at ICASSP 2019 [15], short paper published at CVPR 2018 Workshop [16]

GreyOrange Robotics, Gurugram, India — Image Processing Engineer — ***Feb 2016 - May 2017***

- Developed computer vision solutions for embedded robotics in real time in C++/Python
- Solely responsible for code development, testing of video processing module, camera drivers, server

Airbus, Bengaluru, India — Associate Engineer — ***Jul 2014 - Feb 2016***

- Avionics software development following standard avionics coding guidelines (DO-178B)

Other Professional Experience

Blue Lion Labs, Canada — AI Advisor — ***Oct 2020 - present***

- Provide technical guidance and mentorship on the design and development of AI/ML systems
- Mentored co-op students and interns, published research papers from work led by them [6][8]

NextAI, Canada — AI Scientist-in-Residence — ***Apr-Sep 2019, Mar-Sep 2020***

- Provided scientific support to start-ups selected in yearly co-horts of NextAI accelerator

Playment, Bengaluru, India — Computer Vision Consultant — ***Jan-Jun 2018***

- Provided technical guidance on semantic segmentation models for autonomous driving

TalentSprint, Hyderabad, India — Mentor, Foundations of AI & ML (inaugural program) — ***Jan-May 2018***

- Designed and delivered tutorials on machine learning, mentored industry professionals

Awards

Outstanding Reviewer at CVPR 2021 — *May 2021*
Microsoft Diversity Award for Doctoral Research, $6,000 — *Dec 2020*
MITACS Accelerate Research Internship, $30,000 — *Oct 2020*
University of Montreal entrance scholarship, $37,000 — *Sep 2018*
IIIT Hyderabad merit scholarship for summer school, $1,000 — *Jul 2017*

Service

Organizer — **ICCV 2021** - Differentiable 3D Vision and Graphics workshop — ***Feb-Oct 2021***
OWCV 2021 (Canadian Computer Vision workshop), Canada — ***Feb-Apr 2021***
GRAPHQUON 2020 (Canadian Computer Graphics workshop), Canada — ***Oct-Dec 2020***

Reviewer — ICML 2023, Journal on Computer Vision and Image Understanding, CVPR 2022, ACML 2021, NeurIPS 2021, ICCV 2021, CVPR 2021 (***Outstanding Reviewer***), ICLR 2020, NeurIPS 2020, ICML 2020, NeurIPS 2019, CCAI @ ICLR 2020, CCAI @ NeurIPS 2019, LLD @ ICLR 2019

Teaching Experience

University of Montreal, Montreal, Canada — Guest Lecturer — ***Nov 2022***
- Representation Learning (IFT 6135) by Prof. Aishwarya Agrawal

University of Montreal, Montreal, Canada — Teaching Assistant — ***Sep-Dec 2020***
- Fundamentals of Machine Learning (IFT 6390) by Prof. Ioannis Mitliagkas

Summer Symposium on AI Research, India — Guest Speaker — ***Jul 2020***

University of Montreal, Montreal, Canada — Teaching Assistant — ***Sep 2019***
- Fundamentals of Machine Learning (IFT 6390) by Prof. Ioannis Mitliagkas

IVADO/Mila Deep Learning School, Montreal, Canada — Teaching Assistant — ***Sep 2019***

AI for Social Good Summer Lab, Montreal, Canada — Lecturer — ***May 2019***

TalentSprint, Hyderabad, India — Mentor, Foundations of AI & ML (inaugural program) — ***Jan-May 2018***
- Designed and presented tutorials on machine learning, and mentored industry professionals

Past Internships

KU Leuven, Belgium — *Supervisor*: Prof. Ingrid Verbauwhede, ESAT — ***Summer 2013***
- Designed and implemented carry-free arithmetic operations in Verilog; simulated circuits in Xilinx

IIT Kharagpur, India — *Supervisor*: Prof. Aurobinda Routray, Electrical Engineering — ***Summer 2012***
- Made a gesture recognition program in MATLAB using Hidden Markov Models

Imperial College, UK — *Supervisor*: Prof. Peter Cheung, Electrical & Electronics — ***Summer 2011***
- Circuits and Systems Research Group; measured intra-die power variation in sub-nm FPGAs

Skills

C/C++, CUDA, HTML/CSS, Javascript, Jax, Keras, LaTeX, MATLAB, OpenCV, OS X, Python, PyTorch, R, Shell, SLURM, Tensorflow, Ubuntu, Verilog, Windows

Deep learning, computer vision, machine learning, research and development, generative modeling, NeRF, score-based diffusion models, normalizing flows, Neural ODEs, GANs, Transformers, image generation, video prediction, 3D pose estimation, 3D rendering, text-to-image, text-to-3D, text-to-4D

Talks (Select)

- "Diffusion models for solving video tasks" — INRIA, France [slides] — *Feb 2023*
- "MCVD: Masked Conditional Video Diffusion" — NeurIPS 2022, New Orleans, USA [slides] — *Dec 2022*
- "SMPL-IK: Learned Morphology-Aware Inverse Kinematics for AI Driven Artistic Workflows" — SIGGRAPH Asia, Daegu, South Korea [slides, video] — *Dec 2022*
- "Normalizing flows" — Learning Representations (course), University of Montreal, Canada — *Nov 2022*
- "Score-based Denoising Diffusion Models - a tutorial" — Mila, Canada [slides, video] — *Sep 2022*
- "Denoising Diffusion GANs" — Mila, Canada [slides] — *Feb 2022*
- "Score-based Generative Models with SDEs" — Mila, Canada [slides] — *Feb 2021*
- "Continuous Normalizing Flows" — Mila, Canada [slides] — *Sep 2020*
- "GANs: the story so far" — Summer Symposium on AI Research, India [slides, video] — *Jul 2020*
- "A brief tutorial on Neural ODEs" — Mila, Canada [slides, video] — *Jul 2020*
- "Simple Video Generation using Neural ODEs" — IIIT Hyderabad, India [slides] — *Jan 2020*
- Tutorial on "GANs" — AI for Social Good Summer Lab, Montreal — *May 2019*
- "BigGAN" — Mila, University of Montreal, Canada [slides] — *Oct 2018*
- "Image de-fencing using RGB-D data" — MPI Informatics, Saarbrücken, Germany [slides] — *Feb 2018*
- "Intuition behind LSTMs" — IIIT Hyderabad, India [slides] — *Feb 2018*
- Tutorial on "Back-propagation" — IIIT-Hyderabad, India [slides] — *Aug 2017*

Thesis Projects

Supervisor: Prof. Christopher Pal, Computer Science, University of Montreal, Canada

Doctoral thesis — "Conditional Generative Modeling for Image, 3D Animation, Video" [1][2][3][9][13] ***2023***

- Image generation using Multi-Resolution Continuous Normalizing Flows, Non-Isotropic Denoising Diffusion
- 3D animation using neural inverse kinematics with 3D human pose prior
- Video prediction using Neural ODEs; video prediction, generation, interpolation using Masked Conditional Video Diffusion models

Supervisor: Prof. Rajiv Sahay, Electrical Engineering, IIT Kharagpur, India

Master's thesis — "De-fencing of Images using RGB-D Data" [17] ***2014***

- Elimination of fence-like occlusions, and inpainting of images using RGB-D data
- Nominated for Best Project Award among three departments, research work published at ICAPR 2015

Bachelor's thesis — "Identification of Bilabial Lip Closures in Audio and Video" ***2013***

- Measurement of synchronization between audio and video using bilabial cues in both modes

---

Research Papers (Select)

[1] *NeurIPS 2022* - "MCVD: Masked Conditional Video Diffusion for Prediction, Generation, and Interpolation", **V. Voleti**, A. Jolicoeur-Martineau, C. Pal [arXiv]

[2] *NeurIPS 2022 Workshop* - "Score-based Denoising Diffusion with Non-Isotropic Gaussian Noise Models", **V. Voleti**, C. Pal, A. Oberman [arXiv]

[3] *SIGGRAPH Asia 2022* - "SMPL-IK: Learned Morphology-Aware Inverse Kinematics for AI-Driven Artistic Workflows", **V. Voleti**, B. N. Oreshkin, F. Bocquelet, F. G. Harvey, L. Ménard, C. Pal [arXiv]

[4] *Submitted to a journal* - "Multi-Resolution Continuous Normalizing Flows", **V. Voleti**, C. Finlay, A. Oberman, C. Pal [arXiv]

[5] *ICLR 2022* - "FairCal : Fairness Calibration for Face Verification", T. Salvador, S. Cairns, **V. Voleti**, N. Marshall, A. Oberman [arXiv]

[6] *CVIS 2022 (Oral)* - "Plankton-FL: Exploration of Federated Learning for Privacy-Preserving Training of Deep Neural Networks for Phytoplankton Classification", D. Zhang, **V. Voleti**, A. Wong, J. Deglint

[7] *Frontiers in Artificial Intelligence (journal)* - "Generative Models of Brain Dynamics", M. Ramezanian-Panahi, G. Abrevaya, JC. Gagnon-Audet, **V. Voleti**, I. Rish, G. Dumas [arXiv]

[8] *FSS at AAAI 2022* - "Towards Generating Large Synthetic Phytoplankton Datasets for Efficient Monitoring of Harmful Algal Blooms", N. Bamra, **V. Voleti**, A. Wong, J. Deglint [arXiv]

[9] *ICML 2021 Workshop* - "Improving Continuous Normalizing Flows using a Multi-Resolution Framework", **V. Voleti**, C. Finlay, A. Oberman, C. Pal

[10] *ICLR 2021* - "gradSim: Differentiable simulation for system identification and visuomotor control" , K. M. Jatavallabhula, M. Macklin, F. Golemo, **V. Voleti**, L. Petrini, M. Weiss, B. Considine, J. Parent-Lévesque, K. Xie, K. Erleben, L. Paull, F. Shkurti, D. Nowrouzezahrai, S. Fidler [arXiv]

[11] *MLSys 2021* - "Accounting for Variance in Machine Learning Benchmarks", X. Bouthillier, P. Delaunay, M. Bronzi, A. Trofimov, B. Nichyporuk, J. Szeto, N. Sepah, E. Raff, K. Madan, **V. Voleti**, S. E. Kahou, V. Michalski, D. Serdyuk, T. Arbel, C. Pal, G. Varoquaux, P. Vincent [arXiv]

[12] *ICML 2020* - "Learning to Combine Top-Down and Bottom-Up Signals in RNNs with Attention over Modules", S. Mittal, A. Lamb, A. Goyal, **V. Voleti**, M. Shanahan, G. Lajoie, M. Mozer, Y. Bengio [arXiv]

[13] *NeurIPS 2019 Workshop* - "Simple Video Generation using Neural ODEs", **V. Voleti**, D. Kanaa, S. E. Kahou, C. Pal [arXiv]

[14] *ICML 2019 Workshop* - "Comparing Normalization in Conditional Computation Tasks", V. Michalski, **V. Voleti**, S. E. Kahou, A. Oritz, P. Vincent, C. Pal, D. Precup [arXiv]

[15] *ICASSP 2019* - "Cross-Language Speech Dependent Lip-Synchronization", **V. Voleti**, A. Jha, V. P. Namboodiri, C. V. Jawahar [pdf]

[16] *CVPR 2018 Workshop* - "Lip-Synchronization for Dubbed Instructional Videos", **V. Voleti**, A. Jha, V. P. Namboodiri, C. V. Jawahar (FIVER) [pdf]

[17] *ICAPR* 2015 - "A Multimodal Approach for Image De-fencing and Depth Inpainting", S. Jonna, **V. Voleti**, R. R. Sahay, and M. S. Kankanhalli [pdf, IEEE]