Vikram Voleti

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

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

Expertise

Deep learning for generative media : image, video, 3D, and beyond Expert at machine learning research and development with a proven track record in leading international collaborative projects across industry and academia. Past projects include:

- Video generation with denoising diffusion models [2][3][5], 3D video (4D) generation [1]
- Text/Image to 3D using NeRF [2][4]; 3D human pose estimation and inverse kinematics [7]
- Image generation using normalizing flows [8]; video generation using Neural ODEs [13], GANs [14]
- Contributed to projects on 4D generation, simulation [10], fairness/uncertainty [9]

Education

Mila, University of Montreal, Canada — ***2018 - 2023***

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

Thesis: Conditional generative modeling for images, 3D animations, and video [5][6][7][8][13] [arXiv][slides]

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 [15]

Work Experience

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

- Leading research and development of cutting-edge AI models for videos, images, 3D, and 4D from text
- *Released*: Stable Video Diffusion (SVD) [3], Stable Video 3D (SV3D) [2], Stable Video 4D (SV4D) [1], Stable Zero123; contributed to 3D objects dataset Objaverse-XL [4], 3D codebase threestudio

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 using 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 [7], 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 deep semi-supervised learning for active speaker detection in video

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 [14], short paper published at CVPR 2018 Workshop

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

- Developed computer vision solutions for embedded robotics in real time in C++/Python

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

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

Awards

CIPPRS John Barron Doctoral Dissertation Award	*May 2024*
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*

Skills

C/C++, CUDA, Jax, Keras, MATLAB, OpenCV, Python, PyTorch, R, Shell, SLURM, Tensorflow

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

Additional Work 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

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

---

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 — NeurIPS 2024, CVPR 2024, 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

Thesis Projects

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

Doctoral thesis — "Conditional Generative Modeling for Image, 3D Animation, Video" [arXiv] ***2023***
- *Images*: Multi-Resolution Continuous Normalizing Flows [8], Non-Isotropic Denoising Diffusion [6]
- *3D animation*: neural inverse kinematics with 3D human pose prior [7]
- *Video*: Neural ODEs [13], Masked Conditional Video Diffusion models [5]

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

Master's thesis — "De-fencing of Images using RGB-D Data"[15] ***2014***
- Elimination of fence-like occlusions, and inpainting of images using RGB-D data
- Nominated for Best Project Award among three departments, research published at ICAPR 2015 [15]

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

---

Talks (Select)

- Ph.D. thesis "Conditional generative modeling for images, 3D animations, video" [slides, arXiv] *Sep 2023*
- "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*
- "Score-based Denoising Diffusion Models - a tutorial" — Mila, Canada [slides, video] *Sep 2022*
- "Denoising Diffusion GANs" — Mila, Canada [slides] *Feb 2022*
- "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*
- "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*

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

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

---

Research Papers (Select)

[1] "SV4D - Dynamic 3D Content Generation with Multi-Frame and Multi-View Consistency", Y. Xie, CH Yao, **V. Voleti**, H. Jiang, V. Jampani arXiv

[2] "SV3D - Novel multi-view synthesis and 3D generation from a single image using latent video diffusion", **V. Voleti**, CH Yao, M. Boss, A. Letts, D. Pankratz, D. Tochilkin, C. Laforte, R. Rombach, V. Jampani arXiv

[3] "Stable Video Diffusion: Scaling latent video diffusion models to large datasets", A. Blattmann, T. Dockhorn, S. Kulal, D. Mendelevitch, M. Kilian, D. Lorenz, Y. Levi, Z. English, **V. Voleti**, A. Letts, V. Jampani, R. Rombach arXiv

[4] *NeurIPS 2023* - "Objaverse-XL - A Universe of 10M+ 3D Objects", M. Deitke, R. Liu, M. Wallingford, H. Ngo, O. Michel, A. Kusupati, A. Fan, C. Laforte, **V. Voleti**, S. Y. Gadre, E. VanderBilt, A. Kembhavi, C. Vondrick, G. Gkioxari, K. Ehsani, L. Schmidt, A. Farhadi arXiv

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

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

[7] *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

[8] *Annals of Mathematics and Artificial Intelligence* - "Multi-Resolution Continuous Normalizing Flows", **V. Voleti**, C. Finlay, A. Oberman, C. Pal arXiv

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

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

[15] *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