



# VIKRAM VOLETI

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

 [voletiv.github.io](https://github.com/voletiv)


 [vikram.voleti@gmail.com](mailto: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; experienced in leading multiple collaborative projects with international teams. 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**  **S. 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	<b>Blue Lion Labs</b> , Canada — AI Advisor	<i>Oct 2020 - present</i>
	<ul style="list-style-type: none"> <li>• Provide technical guidance and mentorship on the design and development of AI/ML systems</li> <li>• Mentored co-op students and interns, published research papers from work led by them</li> </ul>	
	<b>NextAI</b> , Canada — AI Scientist-in-Residence	<i>Apr-Sep 2019, Mar-Sep 2020</i>
	<ul style="list-style-type: none"> <li>• Provided scientific support to start-ups selected in yearly co-horts of NextAI accelerator</li> </ul>	
	<b>Playment</b> , Bengaluru, India — Computer Vision Consultant	<i>Jan-Jun 2018</i>
	<ul style="list-style-type: none"> <li>• Provided technical guidance on semantic segmentation models for autonomous driving</li> </ul>	
	<b>TalentSprint</b> , Hyderabad, India — Mentor, Foundations of AI & ML (inaugural program)	<i>Jan-May 2018</i>
	<ul style="list-style-type: none"> <li>• Designed and delivered tutorials on machine learning, mentored industry professionals</li> </ul>	
SERVICE	<b>Organizer</b> — <b>ICCV 2021</b> - Differentiable 3D Vision and Graphics workshop	<i>Feb-Oct 2021</i>
	<b>OWCV 2021</b> (Canadian Computer Vision workshop), Canada	<i>Feb-Apr 2021</i>
	<b>GRAPHQUON 2020</b> (Canadian Computer Graphics workshop), Canada	<i>Oct-Dec 2020</i>
	<b>Reviewer</b> — NeurIPS 2024, CVPR 2024, ICML 2023, Journal on Computer Vision and Image Understanding, CVPR 2022, ACML 2021, NeurIPS 2021, ICCV 2021, CVPR 2021 ( <i>Outstanding Reviewer</i> ), ICLR 2020, NeurIPS 2020, ICML 2020, NeurIPS 2019, CCAI @ ICLR 2020, CCAI @ NeurIPS 2019, LLD @ ICLR 2019	
THESIS PROJECTS	<i>Supervisor:</i> Prof. Christopher Pal, Mila, Computer Science, University of Montreal, Canada	
	<b>Doctoral thesis</b> — “Conditional Generative Modeling for Image, 3D Animation, Video” [arXiv]	<i>2023</i>
	<ul style="list-style-type: none"> <li>• <i>Images</i>: Multi-Resolution Continuous Normalizing Flows [8], Non-Isotropic Denoising Diffusion [6]</li> <li>• <i>3D animation</i>: neural inverse kinematics with 3D human pose prior [7]</li> <li>• <i>Video</i>: Neural ODEs [13], Masked Conditional Video Diffusion models [5]</li> </ul>	
	<i>Supervisor:</i> Prof. Rajiv Sahay, Electrical Engineering, IIT Kharagpur, India	
	<b>Master’s thesis</b> — “De-fencing of Images using RGB-D Data” [15]	<i>2014</i>
	<ul style="list-style-type: none"> <li>• Elimination of fence-like occlusions, and inpainting of images using RGB-D data</li> <li>• Nominated for Best Project Award among three departments, research published at ICAPR 2015 [15]</li> </ul>	
	<b>Bachelor’s thesis</b> — “Identification of Bilabial Lip Closures in Audio and Video”	<i>2013</i>
	<ul style="list-style-type: none"> <li>• Measurement of synchronization between audio and video using bilabial cues in both modes</li> </ul>	
TALKS (SELECT)	<ul style="list-style-type: none"> <li>• Ph.D. thesis “Conditional generative modeling for images, 3D animations, video” [slides, arXiv]</li> </ul>	<i>Sep 2023</i>
	<ul style="list-style-type: none"> <li>• “Diffusion models for solving video tasks” — INRIA, France [slides]</li> </ul>	<i>Feb 2023</i>
	<ul style="list-style-type: none"> <li>• “MCVD: Masked Conditional Video Diffusion” — NeurIPS 2022, New Orleans, USA [slides]</li> </ul>	<i>Dec 2022</i>
	<ul style="list-style-type: none"> <li>• “SMPL-IK: Learned Morphology-Aware Inverse Kinematics for AI Driven Artistic Workflows” — SIGGRAPH Asia, Daegu, South Korea [slides, video]</li> </ul>	<i>Dec 2022</i>
	<ul style="list-style-type: none"> <li>• “Score-based Denoising Diffusion Models - a tutorial” — Mila, Canada [slides, video]</li> </ul>	<i>Sep 2022</i>
	<ul style="list-style-type: none"> <li>• “Denoising Diffusion GANs” — Mila, Canada [slides]</li> </ul>	<i>Feb 2022</i>
	<ul style="list-style-type: none"> <li>• “Continuous Normalizing Flows” — Mila, Canada [slides]</li> </ul>	<i>Sep 2020</i>
	<ul style="list-style-type: none"> <li>• “GANs: the story so far” — Summer Symposium on AI Research, India [slides, video]</li> </ul>	<i>Jul 2020</i>
	<ul style="list-style-type: none"> <li>• “A brief tutorial on Neural ODEs” — Mila, Canada [slides, video]</li> </ul>	<i>Jul 2020</i>
	<ul style="list-style-type: none"> <li>• “Simple Video Generation using Neural ODEs” — IIIT Hyderabad, India [slides]</li> </ul>	<i>Jan 2020</i>
	<ul style="list-style-type: none"> <li>• Tutorial on “GANs” — AI for Social Good Summer Lab, Montreal</li> </ul>	<i>May 2019</i>
	<ul style="list-style-type: none"> <li>• “Image de-fencing using RGB-D data” — MPI Informatics, Saarbrücken, Germany [slides]</li> </ul>	<i>Feb 2018</i>
	<ul style="list-style-type: none"> <li>• “Intuition behind LSTMs” — IIIT Hyderabad, India [slides]</li> </ul>	<i>Feb 2018</i>
	<ul style="list-style-type: none"> <li>• Tutorial on “Back-propagation” — IIIT-Hyderabad, India [slides]</li> </ul>	<i>Aug 2017</i>
PAST INTERNSHIPS	<b>KU Leuven</b> , Belgium — <i>Supervisor:</i> Prof. Ingrid Verbauwhede, ESAT	<i>Summer 2013</i>
	<ul style="list-style-type: none"> <li>• Designed and implemented carry-free arithmetic operations in Verilog; simulated circuits in Xilinx</li> </ul>	
	<b>IIT Kharagpur</b> , India — <i>Supervisor:</i> Prof. Aurobinda Routray, Electrical Engineering	<i>Summer 2012</i>
	<ul style="list-style-type: none"> <li>• Made a gesture recognition program in MATLAB using Hidden Markov Models</li> </ul>	
	<b>Imperial College</b> , UK — <i>Supervisor:</i> Prof. Peter Cheung, Electrical & Electronics	<i>Summer 2011</i>
	<ul style="list-style-type: none"> <li>• Circuits and Systems Research Group; measured intra-die power variation in sub-nm FPGAs</li> </ul>	

TEACHING EXPERIENCE	<b>University of Montreal</b> , Montreal, Canada — Guest Lecturer <span style="float: right;"><i>Nov 2022</i></span> • Representation Learning (IFT 6135) by Prof. Aishwarya Agrawal <b>University of Montreal</b> , Montreal, Canada — Teaching Assistant <span style="float: right;"><i>Sep-Dec 2020</i></span> • Fundamentals of Machine Learning (IFT 6390) by Prof. Ioannis Mitliagkas <b>Summer Symposium on AI Research</b> , India — Guest Speaker <span style="float: right;"><i>Jul 2020</i></span> <b>University of Montreal</b> , Montreal, Canada — Teaching Assistant <span style="float: right;"><i>Sep 2019</i></span> • Fundamentals of Machine Learning (IFT 6390) by Prof. Ioannis Mitliagkas <b>IVADO/Mila Deep Learning School</b> , Montreal, Canada — Teaching Assistant <span style="float: right;"><i>Sep 2019</i></span> <b>AI for Social Good Summer Lab</b> , Montreal, Canada — Lecturer <span style="float: right;"><i>May 2019</i></span> <b>TalentSprint</b> , Hyderabad, India — Mentor, Foundations of AI & ML (inaugural program) <span style="float: right;"><i>Jan-May 2018</i></span> • 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. Bocquet, 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](#)