



# 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	<b>Deep learning for image, video, 3D:</b> expert at machine learning research and development; experienced in leading multiple collaborative projects with international partners in industry and academia. <b>Projects</b> include: <ul style="list-style-type: none"><li>• Video prediction/generation with denoising diffusion models [3][1], non-isotropic diffusion models [4]</li><li>• Text to 3D using dreamfusion, NeRF, DMTet [2]; 3D human pose estimation and inverse kinematics [5]</li><li>• Image generation using normalizing flows [6][9]; video generation using Neural ODEs [13], GANs [15][16]</li><li>• Contributed to projects on 4D generation, simulation [10], fairness/uncertainty [7], federated learning</li></ul>
EDUCATION	 <b>Mila, University of Montreal</b> , Canada <span style="float: right;"><i>2018 - 2023</i></span> <b>Ph.D.</b> in Computer Science — <i>Supervisor:</i> Prof. Christopher Pal <i>Thesis:</i> Conditional generative modeling for images, 3D animations, and video [3][4][5][9][13] [arXiv][slides]  <b>Indian Institute of Technology (IIT), Kharagpur</b> , India <span style="float: right;"><i>2009 - 2014</i></span> Dual Degree ( <b>B.Tech. (Honours)</b> + <b>M.Tech.</b> ) in Electrical Engineering with Master's specialization in Instrumentation and Signal Processing [17]
WORK EXPERIENCE	 <b>S. Stability AI</b> , Canada (Remote) — Research Scientist <span style="float: right;"><i>Apr 2023 - present</i></span> <ul style="list-style-type: none"><li>• Leading AI research and development on generating 3D objects, images, videos from text</li><li>• Released a state-of-the-art diffusion model for video generation [1], and a large dataset of 3D objects [2]</li></ul>  <b>Meta</b> (formerly <b>Facebook</b> ), Menlo Park, USA — Research Intern <span style="float: right;"><i>Aug 2022 - Feb 2023</i></span> <i>Team:</i> AI for Metaverse (AI4RL); <i>Supervisors:</i> Dr. Yashar Mehdad, Dr. Barlas Oguz <ul style="list-style-type: none"><li>• Led the technology development for generating 3D objects, videos from text; dreamfusion, NeRF</li><li>• Applied expertise at neural graphics for 3D rendering; implemented hands-on in PyTorch</li><li>• International AI team; technology transitioned into a Meta end product, adopted by other teams</li></ul>  <b>Unity Technologies</b> , Montreal, Canada — MITACS Research Intern <span style="float: right;"><i>Oct 2021 - Aug 2022</i></span> <i>Team:</i> Deep Pose, Unity Labs; <i>Supervisor:</i> Dr. Boris Oreshkin <ul style="list-style-type: none"><li>• Built AI-assisted user-editable 3D character animation workflow; trained novel 3D human pose prior</li><li>• Published at SIGGRAPH Asia [5], incorporated technology into a Unity product</li></ul>  <b>Google</b> , Mountain View, USA — Research Intern <span style="float: right;"><i>Sep-Dec 2019</i></span> <i>Team:</i> Google AI Perception; <i>Supervisors:</i> Dr. Bryan Seybold, Dr. Sourish Chaudhuri <ul style="list-style-type: none"><li>• Investigated the scope of deep semi-supervised learning for active speaker detection in video</li><li>• Hands-on implementation in TensorFlow; collaborated with TPU team to code Neural ODE in Jax</li></ul>  <b>IIIT Hyderabad</b> , India — Research Fellow <span style="float: right;"><i>May 2017 - Aug 2018</i></span> <i>Supervisors:</i> Prof. C. V. Jawahar, IIIT-Hyderabad, Prof. Vinay Namboodiri, IIT Kanpur <ul style="list-style-type: none"><li>• Synthesized videos in Indian languages using GANs; developed automated video dataset pipeline</li><li>• Full paper published at ICASSP 2019 [15], short paper published at CVPR 2018 Workshop [16]</li></ul>  <b>GreyOrange Robotics</b> , Gurugram, India — Image Processing Engineer <span style="float: right;"><i>Feb 2016 - May 2017</i></span> <ul style="list-style-type: none"><li>• Developed computer vision solutions for embedded robotics in real time in C++/Python</li><li>• Solely responsible for code development, testing of video processing module, camera drivers, server</li></ul>  <b>Airbus</b> , Bengaluru, India — Associate Engineer <span style="float: right;"><i>Jul 2014 - Feb 2016</i></span> <ul style="list-style-type: none"><li>• Avionics software development following standard avionics coding guidelines (DO-178B)</li></ul>
ADDITIONAL WORK EXPERIENCE	<b>Blue Lion Labs</b> , Canada — AI Advisor <span style="float: right;"><i>Oct 2020 - present</i></span> <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 <span style="float: right;"><i>Apr-Sep 2019, Mar-Sep 2020</i></span> <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 <span style="float: right;"><i>Jan-Jun 2018</i></span> <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) <span style="float: right;"><i>Jan-May 2018</i></span> <ul style="list-style-type: none"><li>• Designed and delivered tutorials on machine learning, mentored industry professionals</li></ul>

AWARDS	PhD dissertation nominated for Dean’s Award	<i>Sep 2023</i>
	Outstanding Reviewer at CVPR 2021	<i>May 2021</i>
	Microsoft Diversity Award for Doctoral Research, \$6,000	<i>Dec 2020</i>
	MITACS Accelerate Research Internship, \$30,000	<i>Oct 2020</i>
	University of Montreal entrance scholarship, \$37,000	<i>Sep 2018</i>
	IIIT Hyderabad merit scholarship for summer school, \$1,000	<i>Jul 2017</i>
SERVICE	<b>Organizer</b> — <b>ICCV 2021</b> - Differentiable 3D Vision and Graphics workshop	<b><i>Feb-Oct 2021</i></b>
	<b>OWCV 2021</b> (Canadian Computer Vision workshop), Canada	<b><i>Feb-Apr 2021</i></b>
	<b>GRAPHQUON 2020</b> (Canadian Computer Graphics workshop), Canada	<b><i>Oct-Dec 2020</i></b>
	<b>Reviewer</b> — CVPR 2024, ICML 2023, Journal on Computer Vision and Image Understanding, CVPR 2022, ACML 2021, NeurIPS 2021, ICCV 2021, CVPR 2021 ( <b><i>Outstanding Reviewer</i></b> ), ICLR 2020, NeurIPS 2020, ICML 2020, NeurIPS 2019, CCAI @ ICLR 2020, CCAI @ NeurIPS 2019, LLD @ ICLR 2019	
TEACHING EXPERIENCE	<b>University of Montreal</b> , Montreal, Canada — Guest Lecturer	<b><i>Nov 2022</i></b>
	• Representation Learning (IFT 6135) by Prof. Aishwarya Agrawal	
	<b>University of Montreal</b> , Montreal, Canada — Teaching Assistant	<b><i>Sep-Dec 2020</i></b>
	• Fundamentals of Machine Learning (IFT 6390) by Prof. Ioannis Mitliagkas	
	<b>Summer Symposium on AI Research</b> , India — Guest Speaker	<b><i>Jul 2020</i></b>
	<b>University of Montreal</b> , Montreal, Canada — Teaching Assistant	<b><i>Sep 2019</i></b>
	• Fundamentals of Machine Learning (IFT 6390) by Prof. Ioannis Mitliagkas	
	<b>IVADO/Mila Deep Learning School</b> , Montreal, Canada — Teaching Assistant	<b><i>Sep 2019</i></b>
	<b>AI for Social Good Summer Lab</b> , Montreal, Canada — Lecturer	<b><i>May 2019</i></b>
	<b>TalentSprint</b> , Hyderabad, India — Mentor, Foundations of AI & ML (inaugural program)	<b><i>Jan-May 2018</i></b>
	• Designed and presented tutorials on machine learning, and mentored industry professionals	
PAST INTERNSHIPS	<b>KU Leuven</b> , Belgium — <i>Supervisor</i> : Prof. Ingrid Verbauwhede, ESAT	<b><i>Summer 2013</i></b>
	• Designed and implemented carry-free arithmetic operations in Verilog; simulated circuits in Xilinx	
	<b>IIT Kharagpur</b> , India — <i>Supervisor</i> : Prof. Aurobinda Routray, Electrical Engineering	<b><i>Summer 2012</i></b>
	• Made a gesture recognition program in MATLAB using Hidden Markov Models	
	<b>Imperial College</b> , UK — <i>Supervisor</i> : Prof. Peter Cheung, Electrical & Electronics	<b><i>Summer 2011</i></b>
	• Circuits and Systems Research Group; measured intra-die power variation in sub-nm FPGAs	
SKILLS	C/C++, CUDA, HTML/CSS, Javascript, Jax, Keras, L <sup>A</sup> T <sub>E</sub> X, 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)	• Ph.D. thesis “Conditional generative modeling for images, 3D animations, video” [ <a href="#">slides</a> , <a href="#">arXiv</a> ]	<i>Sep 2023</i>
	• “Diffusion models for solving video tasks” — <a href="#">INRIA</a> , France [ <a href="#">slides</a> ]	<i>Feb 2023</i>
	• “MCVD: Masked Conditional Video Diffusion” — NeurIPS 2022, New Orleans, USA [ <a href="#">slides</a> ]	<i>Dec 2022</i>
	• “SMPL-IK: Learned Morphology-Aware Inverse Kinematics for AI Driven Artistic Workflows” — SIGGRAPH Asia, Daegu, South Korea [ <a href="#">slides</a> , <a href="#">video</a> ]	<i>Dec 2022</i>
	• “Normalizing flows” — Learning Representations (course), University of Montreal, Canada	<i>Nov 2022</i>
	• “Score-based Denoising Diffusion Models - a tutorial” — Mila, Canada [ <a href="#">slides</a> , <a href="#">video</a> ]	<i>Sep 2022</i>
	• “Denoising Diffusion GANs” — Mila, Canada [ <a href="#">slides</a> ]	<i>Feb 2022</i>
	• “Continuous Normalizing Flows” — Mila, Canada [ <a href="#">slides</a> ]	<i>Sep 2020</i>
	• “GANs: the story so far” — <a href="#">Summer Symposium on AI Research</a> , India [ <a href="#">slides</a> , <a href="#">video</a> ]	<i>Jul 2020</i>
	• “A brief tutorial on Neural ODEs” — Mila, Canada [ <a href="#">slides</a> , <a href="#">video</a> ]	<i>Jul 2020</i>
	• “Simple Video Generation using Neural ODEs” — IIIT Hyderabad, India [ <a href="#">slides</a> ]	<i>Jan 2020</i>
	• Tutorial on “GANs” — <a href="#">AI for Social Good Summer Lab</a> , Montreal	<i>May 2019</i>
	• “Image de-fencing using RGB-D data” — MPI Informatics, Saarbrücken, Germany [ <a href="#">slides</a> ]	<i>Feb 2018</i>
	• “Intuition behind LSTMs” — IIIT Hyderabad, India [ <a href="#">slides</a> ]	<i>Feb 2018</i>
	• Tutorial on “Back-propagation” — IIIT-Hyderabad, India [ <a href="#">slides</a> ]	<i>Aug 2017</i>

## THESIS PROJECTS

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

- Doctoral thesis** — “Conditional Generative Modeling for Image, 3D Animation, Video” [arXiv] **2023**
- Image generation using Multi-Resolution Continuous Normalizing Flows [9], Non-Isotropic Denoising Diffusion Models [4]
  - 3D animation using neural inverse kinematics with 3D human pose prior [5]
  - Video prediction using Neural ODEs [13], Masked Conditional Video Diffusion models [3]

*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] “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](#)
- [2] *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](#)
- [3] *NeurIPS 2022* - “MCVD: Masked Conditional Video Diffusion for Prediction, Generation, and Interpolation”, **V. Voleti**, A. Jolicoeur-Martineau, C. Pal [arXiv](#)
- [4] *NeurIPS 2022 Workshop* - “Score-based Denoising Diffusion with Non-Isotropic Gaussian Noise Models”, **V. Voleti**, C. Pal, A. Oberman [arXiv](#)
- [5] *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](#)
- [6] *Submitted to a journal* - “Multi-Resolution Continuous Normalizing Flows”, **V. Voleti**, C. Finlay, A. Oberman, C. Pal [arXiv](#)
- [7] *ICLR 2022* - “FairCal : Fairness Calibration for Face Verification”, T. Salvador, S. Cairns, **V. Voleti**, N. Marshall, A. Oberman [arXiv](#)
- [8] *Frontiers in Artificial Intelligence (journal)* - “Generative Models of Brain Dynamics”, M. Ramezani-Panahi, G. Abrevaya, J.C. Gagnon-Audet, **V. Voleti**, I. Rish, G. Dumas [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](#)