BHAVAN JASANI

bjasani@alumni.cmu.edu | https://bhavanj.github.io/ | (412) 618 – 9200
http://www.linkedin.com/in/bhavan-jasani | Google scholar

SUMMARY

Machine learning and computer vision scientist with 6+ years of applied research experience building and deploying large-scale transformer-based large language models for (1) multimodal learning -- across images, text, video, audio, and structured data (2) synthetic data generation & annotation -- with humans and AI in the loop to overcome scarce or hard-to-label data. Published in top computer vision and machine learning conferences (CVPR, ECCV, ICCV), mentored PhD interns, filed patents, and collaborated with cross-functional teams to translate research ideas into production-grade models solving business problems.

SKILLS

- **Programming:** Python, PyTorch, TensorFlow, JAX, DeepSpeed, FSDP, Hugging Face, PySpark, Gradio, C/C++
- **ML & Deep Learning:** Foundation Model Post-Training (SFT, RLHF, DPO), Pre-Training, Vision-Language Models (QWEN-VL), Vision Transformers, Multimodal Learning, LoRA/Adapters, Diffusion Models, Human-in-the-Loop Data & Annotation Pipelines, Synthetic Data Generation
- **Systems:** Multi-node Distributed Training, AWS (EC2/S3/Lambda), Docker, Kubernetes, Git

EXPERIENCE

Amazon Web Services (AWS) AI San Francisco, CA
Applied Scientist, AWS AI labs (Computer Vision) September 2019 – present

- Large-scale multi-node fine-tuning, and preference learning pipelines integrating spatial, textual, and visual modalities for visually rich document understanding (infographics, charts, tables, figures, OCR text, long context multi-page)
- Designed and deployed robust multimodal transformer architectures for document question-answering at commercial scale, as a part of Amazon Textract (related work published at ICCV)
- Developed multi-agent synthetic data generation pipelines (probe, generator, retriever, and verifier agents plus tools, templates and humans) to generate SFT and preference learning data for vision-language model improvement; related work published in CVPR using LLM & tools to generate step-by-step reasoning data, pushing chart VQA SOTA by 15%
- Led end-to-end development of a multimodal AI agent that takes documents, images, video, and audio as input, and suggests relevant questions/prompt to ask in a zero-shot way (part of Amazon Bedrock Data Automation)
- Created visual grounding methods for model explainability, linking language-model text predictions to corresponding image regions
- Mentored PhD interns, collaborated cross-functionally, published in CVPR/ICCV/ECCV, and filed patents on multimodal transformer architectures

Carnegie Mellon University, Robotics Institute, School of Computer Science Pittsburgh, PA
Research Assistant October 2017 – August 2019

- Built multimodal emotion recognition models integrating audio, video, facial action units, body pose, and 3D facial landmarks for clinical psychology datasets
- Contributed to an NIH-funded study identifying behavioral biomarkers of depression using HIPAA-compliant psychotherapy session videos, collaborating with an interdisciplinary team of psychologists, clinicians, and computer scientists

Nanyang Technological University, School of Computer Science & Engineering Singapore
Research Staff January 2016 – May 2017

- Implemented a parallel and hardware-efficient DPM object detection algorithm for real-time pedestrian detection on FPGA-based embedded system, achieving a 40% reduction in hardware resources usage
- Developed a bit-width optimization approach for hardware acceleration of the Harris Corner Detector algorithm, achieving 335 FPS on HD video with minimal accuracy loss

SELECTED PUBLICATIONS

- **Synthesize Step-by-Step: Tools, Templates and LLMs as Data Generators for Reasoning-Based Chart VQA**, *CVPR 2024*
- **DocFormer: End-to-End Transformer for Document Understanding**, *ICCV 2021*
- **YORO - Lightweight End-to-End Visual Grounding**, *ECCV Workshops 2022*
- **Are We Asking the Right Questions in MovieQA?**, *ICCV Workshops 2019 (spotlight oral)*
- **Exploiting Spatial Layout in Document Question Answering using Transformers**, *AMLC 2021*
- **End-to-End Visual Question Answering on Document Images**, *AMLC 2021*
- **Threshold-Guided Design and Optimization for Harris Corner Detector Architecture**, *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) 2017*
- **Skeleton-Based Zero-Shot Action Recognition in Joint Pose-Language Semantic Space**, *arXiv 2019*
- **Automatic Detection of Human Affective Behavior in Dyadic Conversations**, *Master's Thesis, Robotics Institute, Carnegie Mellon University 2019*

PATENTS

- **Global Prompts with Linear Adapter Tuning for Regression-Free Model Update**, *US patent 12494077, 2023*
- **Document Visual Question Answering with Multimodal Transformer Encoder–Decoder Models**, *US patent filed 2022*

COMMUNITY SERVICE

- **Reviewer** for CVPR, ICCV, ECCV, AMLC, ACVC and Amazon Research Awards
- **Program Committee Member** for ICDAR 2025 and TASK-CV Workshop at ICCV 2019
- **Book Chapter Reviewer** for Data Augmentation with Python, Packt publishing 2023

AWARDS & ACHIEVEMENTS

- **EB1-B (Outstanding Researcher) Green Card**, granted for demonstrated scientific impact by U.S. Government
- **DAAD WISE scholarship**, German Academic Exchange Service (2014)
- **INSPIRE Fellowship,** Department of Science & Technology, Government of India (2011 – 2016)

EDUCATION

- **Carnegie Mellon University, School of Computer Science** 2017 – 2019
  M.S. in Robotics (Research based – Computer Vision & Machine Learning)
- **Birla Institute of Technology & Science (BITS), Pilani – K.K. Birla Goa** 2011 – 2016
  M.Sc. Physics
- **Birla Institute of Technology & Science (BITS), Pilani – K.K. Birla Goa** 2011 – 2016
  B.E. Electrical & Electronics Engineering