BHAVAN JASANI

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

EDUCATION

Carnegie Mellon University, School of Computer Science — Pittsburgh, PA
M.S. in Robotics (Research-based) — August 2017 – August 2019

Birla Institute of Technology & Science (BITS), Pilani – K.K. Birla Goa Campus — Goa, India
Dual degree: M.Sc. Physics + B.E. Electrical & Electronics Engineering — August 2011 – August 2016

RESEARCH EXPERIENCE

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

- Built novel methods combining expert humans, non-expert human annotators, tools, and fine-tuned vision-language models to generate synthetic data to train reasoning based large vision-language models (work published in CVPR 2024)
- Developed and deployed transformer-based encoder-only and encoder-decoder visual question answering (VQA) models that integrate spatial, textual, and visual modalities to extract structured information from document images - including launching the first commercial document VQA model - Amazon Textract Queries
- Led the full cycle of applied research, product discussions, data annotation, and engineering handoff to build and launch a multi-modal AI assistant (for understanding documents, images, videos, and audio) as part of Bedrock Data Automation
- Developed visual grounding techniques to localize the language model's text predictions in the image regions for explainable ML models
- Published research at top computer vision conferences (CVPR, ICCV, ECCV), mentored PhD interns and filed patents

Carnegie Mellon University, Robotics Institute, School of Computer Science — Pittsburgh, PA
Research Assistant *(Advisor: Prof. Jeffrey Cohn – Psychology & Robotics)* — October 2017 – August 2019

- Developed a multi-modal human emotion recognition system using video and audio data with real-time, noisy annotations; addressing variable temporal lags between video segments and corresponding emotion labels
- Leveraged 3D facial landmarks, head pose, body pose, and facial action units combining classical time-series & deep neural networks for robust emotion recognition
- Discovered and quantified the influence of head movements, facial expressions, and body pose on the behaviour of people in interpersonal conversations in psychotherapy session videos, uncovering key behavioural insights including onset of depression [NIH funded] [details]

Nanyang Technological University, School of Computer Science & Engineering — Singapore
Research Staff (*Advisor: Prof. Lam Siew Kei)* — January 2016 – May 2017

- Implemented a parallel and hardware efficient DPM object detection algorithm for real-time pedestrian detection on an FPGA-based embedded system, achieving a 40% reduction in hardware resources
- Developed a bit-width optimization approach for hardware acceleration of Harris Corner Detector algorithm, reducing bit-width by 45% with only a 0.57% drop in accuracy, and achieving real-time performance at 335 FPS on HD videos

SELECTED PUBLICATIONS

- Synthesize Step-by-Step: Tools, Templates and LLMs as Data Generators for Reasoning-Based Chart VQA, *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024
- YORO - Lightweight End to End Visual Grounding, *European Conference in Computer Vision (ECCV) workshops*, 2022
- DocFormer: End-to-End Transformer for Document Understanding, *International Conference on Computer Vision (ICCV)*, 2021
- End-to-End Visual Question Answering on Document Images, *Amazon Machine Learning Conference (AMLC)*, 2021

- Exploiting Spatial Layout in Document Question Answering using Transformers*, Amazon Machine Learning Conference (AMLC)*, 2021
- Are we asking the right questions in MovieQA?, *International Conference on Computer Vision (ICCV) Workshops*, 2019 [spotlight oral presentation]
- Skeleton based Zero-Shot Action Recognition in Joint Pose-Language Semantic Space, arXiv:1911.11344, 2019
- Automatic detection of human affective behavior in dyadic conversations, *Tech. Report, CMU-RI-TR-19-53, Robotics Institute, Carnegie Mellon University*, 2019
- Threshold-guided design and optimization for Harris corner detector architecture, *IEEE Transactions on Circuits and Systems for Video Technology*, 2017

PATENTS

- Document information extraction using visual question answering and document type specific adapters, *pending US patent (filed 2023)*
- Document information extraction using visual question answering and a multi-modal transformer encoder-decoder model, *pending US patent (filed 2022)*

COMMUNITY SERVICE

- Reviewer for Conference on Computer Vision and Pattern Recognition (CVPR), International Conference on Computer Vision (ICCV), European Conference on Computer Vision (ECCV), Amazon Computer Vision Conference (ACVC) and Amazon Research awards
- Program committee for International Conference on Document Analysis and Recognition (ICDAR) 2025 and TASK-CV workshop ICCV 2019
- Reviewer for book chapters – "Data Augmentation with Python", Packt publishing, 2023

AWARDS & ACHIEVEMENTS

- **Research Assistantship, 2017 – 2019**, Carnegie Mellon University
- **DAAD WISE, 2014 scholarship**, awarded by German Academic Exchange Service for a summer research internship
- **Innovation in Science Pursuit for Inspired Research (INSPIRE), 2011 – 2016 fellowship,** from Department Of Science And Technology, Government of India, awarded to bright students majoring in natural sciences

SKILLS

Python, PyTorch, PySpark, TensorFlow, Scikit, OpenCV, AWS, MATLAB, ELAN, Blender, C/C++, Docker, Git, HTML, JS