Avnish Patel

+1-(857)693-9988 | patel.avni@northeastern.edu | **Boston, MA**
Portfolio | Github | LinkedIn

Education

Northeastern University, Boston, MA **Sept 2022-Dec 2024**
Master Of Science in Robotics | Electrical & Computer Engineering Concentration
Coursework: Robot Sensing, Autonomous Field Robotics, Advanced Computer Vision

Publication (Link)

- *B. Maxwell, A. Patel," Logarithmic Lenses: Exploring Log RGB Data for Image Classification", CVPR 2024*

Experience

Research Student – Computer Vision Lab **May-Dec 2023**
Northeastern University, Boston

- Researched raw *log RGB* data's impact on deep networks like *ResNet-18*, improving classification performance and robustness to intensity and color variations, on *novel RAW10 dataset* (*10k DNG & JPG images each, 10 categories*)

Medtronic **Jan 2024-Apr 2025**
Surgical R&T Machine Learning Engineer, Boston

- Ground Truth (G.T.) Generation (Python) with Camera Calibration

- *Built an end-to-end SLAM pipeline with DROID-SLAM for dense depth estimation in surgical videos, optimizing camera trajectory using GTSAM and refining 3D reconstruction with Bundle Adjustment* and *LightGlue feature matching*
- Developed a real-time *Ground Truth pose estimation* pipeline using *OptiTrack* camera capture and *robot kinematics* with *PnP* and *ROMA feature detection* for training deep learning models on *instrument articulation*
- Developed a *VTK-based* application for *real-time manual pose articulation* of the instruments of the surgical robot
- Automated *Endoscope camera calibration* using *Zhang's* method on a Charuco board, cutting *calibration time* by *40%*
- Implemented a *custom pipeline* to generate *synthetic* images with varying *focal lengths* and *distortions*, enabling *regression* of camera intrinsic parameters using a *pretrained Transformer* model

- Deep Learning (PyTorch)

- Performed *Semantic Segmentation* for*, Robot-Assisted Surgery*, on *10,000* medical images from *S3 bucket*, containing both *mask* and *line* annotations, to segment hernia using the *Swin Base Transformer*
- Utilized PyTorch *Distributed Data Parallel (DDP)* for multi-*GPU* training, reducing training time by *30%*
- Developed a *YOLOv8*-based pipeline for precise detection of surgical *instrument tips* from medical images in real-time
- Applied *Monocular Depth Estimation* to get *metric* distance between two instruments from an image by *Depth Anything*
- Implemented a *PyTorch wrapper* with *Optical Flow* on *FAST API* using *Unimatch*, converting models to *ONNX* and *TensorRT* for *10x* reduction in real time annotation of medical image frames with *1-second latency*

Kisan Drip Irrigation Pvt Ltd **Aug 2020-Aug 2022**
Artificial Intelligence Engineer, India

- Designed a *3D defect detection system* using *Intel RealSense D455* cameras and *Open3D*, combining *voxel downsampling*, *RANSAC plane segmentation*, *DBSCAN clustering*, and *surface curvature analysis* to detect missing holes, dents, and deformations in pipe-extruder and sprinkler products
- Experimented with *PointNet*-based deep learning models for *point cloud classification* to enhance complex defect identification, achieving a *30% improvement* over traditional 2D vision methods
- Deployed the *3D vision pipeline* as a containerized *FastAPI service* integrated into on-premises manufacturing workflows

Monocular Visual SLAM with G2O optimization in CARLA (Python) (GitHub) **Aug-Sept 2024**

- Built *Monocular Visual SLAM* with *pose estimation* and *global bundle adjustment* with *G2O* in *CARLA* simulator

Structure from Motion (SFM) from scratch with GTSAM (Python) (GitHub) **Dec-Jan 2024**

- Implemented a full *Structure from Motion* pipeline for *3D reconstruction* using SIFT-based feature matching, *pose estimation*, *triangulation*, and *GTSAM*-based *optimization* to minimize *reprojection error*

Perception System for Autonomous Navigation using ROS2 (Python, C++) **Jan-Feb 2025**

- Designed a *Graph SLAM*-based *Perception* System for *mapping*, *obstacle detection*, and *localization* using *slam_toolbox* and *nav2* libraries in *ROS2*, validated in *Gazebo* with custom environment

3D Object Detection & Tracking with Late Sensor Fusion on Waymo Dataset (PyTorch) (GitHub) **Feb-Mar 2025**

- Built a *3D object detection* and *multi-target tracking LiDAR-Camera Late Fusion* pipeline using *FPN-ResNet* with *Bird's Eye View (BeV)* from *LiDAR*, along with Extended Kalman Filter (EKF) and nearest-neighbor data association
- Generated a globally consistent *semantic 3D map* by fusing *LIDAR* point cloud data with *vehicle pose transformations*
- Augmented *Llama 3* (via *Ollama*) to generate real-time spoken *natural language descriptions* of the 3D environment
- Designed *interactive Three.js* based *web visualizer* for *scene reconstruction*, *LLM messages* for *frame-by-frame playback*

CUDA: Flash Attention in C++ (currently implementing other algorithms in CUDA) (GitHub) **Jan 2025-Present**

- Optimized forward pass of *Flash Attention* with *CUDA* reducing memory usage and increasing speed significantly