Neet Mehulkumar Mehta

India | neetmehta19@gmail.com | +91 9484508208 | GitHub: https://github.com/neetmehta
Website: neetmehta.github.io

MACHINE LEARNING ENGINEER

Perception AI Engineer with 2+ years of experience building **production-grade autonomous driving perception systems**, specializing in **multi-sensor fusion, BEV-based 3D object detection, and embedded deployment**. Proven track record of improving model accuracy and latency through **geometry-aware learning, CUDA/TensorRT optimization, and end-to-end MLOps pipelines**, and shipping perception stacks to vehicles under **SIL/HIL and V-model processes**. Recently applied **image-conditioned generative vision models** as an extension of perception expertise, focusing on **spatial alignment, consistency, and controllable generation**, without shifting away from core perception and robotics fundamentals.

EDUCATION

Worcester Polytechnic Institute (WPI)   **Worcester, MA**
Master of Science- Robotics Engineering, GPA- 3.85/4.00   Dec 2022
Nirma University   **Ahmedabad, India**
Bachelor's in mechanical engineering, GPA- 7.8/10.00   May 2020

KEY SKILLS

- **Programming Skills**:   C++, CUDA, Python, MATLAB
- **Tools and Libraries**:   ROS2, Pytorch, TensorRT, ONNX, CARLA simulator, AWS, PCL (Point Cloud Library), OpenCV, Docker, Git, ComfyUI, Blender 3D, Solidworks.
- **Languages:** English, Hindi, Gujarati

RELEVANT WORK EXPERIENCE

Aavaran   **Bhuj, India**
Applied Generative Vision Engineer, Self Employed   Mar 2025 – Present
Pytorch, ComfyUI, Docker, Vastai

Applied generative vision techniques as an extension of my perception and geometry background, focusing on spatial alignment, scale consistency, and deployment efficiency.

- Designed and deployed **image-conditioned generative vision pipelines** for interior visualization and virtual try-on, emphasizing **geometry consistency, spatial alignment, and controllable generation**.
- Built **text- and image-conditioned workflows** using **Qwen 2.5.11 Image Edit** and **FLUX.2-dev**, training **LoRA adapters** and applying **ControlNet** for identity, pose, and composition control.
- Optimized inference using **quantized GGUF / FP8 models** on **RTX 4090 (Vast.ai)** and consumer GPUs via **custom ComfyUI workflows**, extending pipelines toward **short-form generative video (WAN 2.2)**.

TORC Robotics   **Austin, TX**
ML Engineer II   Feb 2023 – Feb 2025
C++, CUDA, ROS2, Python, Pytorch, AWS, TensorRT
Learned Sensor Fusion

- Defined **clear upstream and subsystem interface requirements** for image (ISP), LiDAR, and depth inputs, specifying **latency, synchronization, and data quality constraints**, and translating system-level requirements into well-scoped perception components.
- Evaluated **learned multi-sensor BEV object detectors** (including **BEVFusion**), shortlisting BEVFusion as the baseline and extending it with **depth-aware BEV unpooling** to generate dense 3D camera features for improved fusion quality.
- Designed and implemented an **attention-based spatio-temporal feature aligner** to jointly align **LiDAR BEV and dense camera features across sensors and time**, achieving a **+2.5% mAP improvement** over vanilla BEVFusion on proprietary datasets.
- Built a **production-grade inference pipeline** converting models **PyTorch → ONNX → TensorRT**, applying **FP16 post-training quantization** to reduce end-to-end latency by **~50%** on embedded platforms.
- Implemented **custom TensorRT plugins and CUDA kernels** for performance-critical operations, including **BEV pooling (~1M camera points), voxelization, NMS, and bounding box decoding**, addressing key runtime bottlenecks.
- Integrated the optimized model into a **proprietary inference engine** and architected a **sensor-abstracted ROS2 perception node** supporting arbitrary numbers of cameras and LiDARs, extensible to radar, with standardized ICP-defined outputs.
- Delivered the full perception stack to a **production truck**, following a **V-model development process** with **unit testing, SIL, and HIL validation**, supported by rich visualization tools in **RViz and Foxglove**.

- Built a **fully automated end-to-end MLOps pipeline** for model training, versioning, conversion, validation, and deployment using **Comet + CI/CD**, enabling seamless rollout of new model versions via a single configuration update.

Unsupervised Domain Adaptation & Monocular Depth Estimation
- Implemented a **POC for geometric Unsupervised Domain Adaptation (GUDA)** to address sim-to-real gaps in autonomous driving perception, leveraging **large-scale synthetic data from Parallel Domain** without labeled real-world supervision.
- Designed a **geometry-guided adaptation pipeline** for **semantic segmentation and monocular depth estimation**, enforcing cross-domain consistency using **depth, camera pose, and geometric alignment constraints**.
- Developed an **unsupervised monocular depth estimation model** using **view synthesis and photometric reconstruction**, jointly learning depth and ego-motion from monocular video sequences via **differentiable warping and multi-scale losses**.
- Incorporated **geometric priors** including edge-aware smoothness, occlusion masking, temporal consistency, and scale-invariant normalization to improve depth stability in dynamic driving environments.
- Evaluated adapted models on real-world driving data, demonstrating **improved generalization over non-adapted baselines** for both **segmentation and depth estimation**, validating the effectiveness of GUDA-based sim-to-real transfer.
- Served as **SCRUM Master for the perception team**, driving sprint planning, backlog grooming, cross-team coordination, and delivery tracking, while ensuring alignment between research, systems, and deployment milestones for a complex multi-sensor autonomy stack.

TORC Robotics
Perception Engineer – Co-Op
Blacksburg, VA
Jan 2022 – Aug 2022

C++, Python, Pytorch, AWS, TensorRT
- Worked on Data extraction and data postprocessing for deep learning architectures. Established extendable pipeline to generate detailed metrics report for each Deep learning model.
- Developed automated hyperparameter tuning stage in AWS Sagemaker. Used Bayesian search to find optimal hyperparameters.

PROJECTS

Synthetic Data Generation for Self Driving Cars using CARLA simulator
Aug 2022 – Present

Python, Pytorch, CARLA
- Built a **scalable synthetic data generation toolkit using CARLA**, enabling unlimited multi-task datasets for autonomous driving perception (depth, optical flow, segmentation, 2D/3D detection).
- Implemented a **highly configurable multi-sensor simulation framework**, supporting dynamic control over **sensor placement, intrinsics/extrinsics, scene complexity, traffic, pedestrians, and capture frequency**.
- Generated **time-synchronized, multi-modal annotations** with a modular, extensible architecture to support **sensor fusion, self-supervised learning, domain adaptation, and sim-to-real research**.

Self-Supervised Monocular Depth Estimation (Monodepth2) from scratch
June 2022 – July 2022

Python, Pytorch
- Implemented **Monodepth2 from scratch in PyTorch**, reproducing the full self-supervised training pipeline including **pose estimation, differentiable view synthesis, automasking, and multi-scale photometric losses**.
- Validated correctness through **qualitative depth visualization and quantitative evaluation** on standard driving datasets, matching reported Monodepth2 behavior.

3D Object Detection in Point Cloud using Voxelnet
Sept 2021 – Dec 2021

Python, Pytorch, OpenCV
- Implement a 3D detection network (VoxelNet) on the KITTI vision (Point Cloud) benchmark dataset to unify feature extraction and bounding box prediction into a single-stage, end-to-end trainable deep network.