

Vamsi Krishna Kocherla

Deep Learning | Machine Learning | Computer Vision | Robotics | GenAI | Open Source Contributor
United States | koche156@umn.edu | +1-763-900-1422 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Education

University of Minnesota - Twin Cities
Master of Science in Robotics

Minneapolis, USA
Aug 2024 – Aug 2026 (Expected)

Relevant Coursework: Computer Vision, Intelligent Robotic Systems, Image Processing & Application, Natural Language Processing, Deep Learning

GITAM University

Hyderabad, India

Bachelor of Technology in Electronics & Communication Engineering

Jun 2017 – Jun 2021

Relevant Coursework: Signals & Systems, Digital Signal Processing, Probability Theory and Random Process

Skills & Interests

Programming Languages: Python, C++, Golang, CUDA, NodeJS

Machine Learning & Deep Learning: PyTorch, TensorFlow, OpenCV, Scikit-learn, Transformers, CNNs, GANs, TensorRT, ONNX

Computer Vision: Object Detection, Semantic/Instance Segmentation, Stereo Vision, SLAM, 3D Reconstruction, Feature Matching, Depth/Disparity Maps, Multi view 3D Geometry, ViT, Faster-RCNN, Mask-RCN, YOLO,

Robotics & Embedded Systems: ROS/ROS2, Isaac Sim (Omniverse), Gazebo, OpenAI Gym, Ardupilot, PX4, MQTT

Edge & IoT Hardware: Jetson Orin Nano, Jetson AGX, Raspberry Pi, Arduino, Sensor Fusion, Sensor Integration

Cloud & Deployment: Kubernetes, KNative, Docker, KServe, Triton Inference Server, AWS, GCP, Azure, REST APIs, gRPC, WebSockets

Databases & Tools: MongoDB, MySQL, PostgreSQL, Redis, ClickHouse, RabbitMQ, Apache Kafka, Prometheus, Grafana

Tools & Technologies: Git, DVC, Ubuntu, Linux, Linux File System, Flask, Fast API, Gitlab CI/CD, Matplotlib

Work Experience

Upjao Agrotech

Senior Computer Vision Scientist (Core Team)

Ahmedabad, India

(Jan 2019 – Jul 2024)

- Developed and implemented novel deep learning models based on Visual Transformers, Faster-RCNN, Mask-RCN, YOLO incorporating custom loss functions, activation functions, and training techniques to achieve upto 20% improvement in accuracy on Classification, Object Detection and Segmentation tasks.
- Designed and implemented a high-performance, serverless inference platform on Kubernetes, utilizing NVIDIA Triton Inference Server, KServe, KNative, and gRPC for serving multiple deep learning models with low latency (<500ms). Optimized the platform for scalability and reliability through microservices architecture and containerization.
- Optimized and deployed deep learning models on resource-constrained edge devices (Raspberry Pi, Jetson Boards) and mobile devices, achieving a 40% reduction in inference time and a 20% reduction in model size through techniques such as model quantization, pruning, and knowledge distillation.
- Improved inference performance by 50% and reduced infrastructure costs by 60% through multithreading, vectorization, and structural pipelining. Further optimized inference speed by utilizing FP16 and INT8 precision models, leveraging the capabilities of NVIDIA A100 and Tesla T4 GPUs.
- Leveraged cutting-edge deep learning models using PyTorch and classical computer vision algorithms using OpenCV to develop robust and accurate computer vision solutions, including object detection, segmentation, and classification, while optimizing model efficiency and training time.
- Established automated testing and deployment procedures for machine learning and deep learning models through the implementation of CI/CD pipelines, incorporating multi-stage pipelines to guarantee thorough validation of the developed models.
- Developed and implemented a model caching algorithm that reduced model scale-up time from 15-20 minutes to 3-4 seconds, significantly improving application responsiveness and resource utilization.

Tata Consultancy Services

AI / ML Developer

Hyderabad, India

(Jun 2021 – Apr 2023)

- Selected into TCS Rapid Labs Cohort, an elite R&D team focused on cutting-edge technologies including **AI and Robotics**.
- Designed and developed multiple NLP-based proof-of-concept projects within three months, utilizing BERT & Transformer based models. These projects involved building core AI algorithms, backend infrastructure using Flask, and user interfaces with React & Vite.
- Developed a novel algorithm to identify and remove complex repeated data from PDFs, Word Docs and images as part of the data cleaning process. This algorithm improved data processing efficiency by 20%, resulting in faster data filtration in downstream pipelines, and reduced human errors significantly.

Projects

Autonomous Drone Navigation with Stereo SLAM | Python, ROS 2, ArduPilot, C++, SLAM, Custom Stereo Vision, Camera Calibration

- Developed an autonomous drone system utilizing a **custom-built stereo vision system with Raspberry Pi cameras. Performed manual camera calibration to achieve accurate depth perception and 3D scene reconstruction.**
- Developed a robust drone navigation system using **Stereo Visual-Inertial SLAM (ORB-SLAM3/RTAB-Map)** for real-time 3D mapping and localization in GPS-denied environments.
- Integrated the drone's sensor suite, **Jetson Orin Nano compute board, and Pixhawk flight controller using ROS 2**, enabling seamless communication and coordination between various drone modules.

- Developed and tested the system in a SLAM-capable simulation environment before deploying it on a real drone, focusing on mapping accuracy and localization.
- Addressed challenges of autonomous navigation in complex indoor and outdoor spaces, enabling the drone to operate safely and independently.
- Emphasized risk management through drone safety training, careful handling of electronic components, and maintaining consistency across software versions.

Image-Based Visual Servoing (IBVS) for Robotic Arm Control | Python, ROS, OpenCV, PyTorch, Stereo Camera, Open3D

- Engineered a custom stereo camera system utilizing Raspberry Pi cameras, **performing comprehensive camera calibration and stereo rectification to determine intrinsic, extrinsic, and baseline parameters.**
- **Generated 3D point clouds and depth maps** from 2D disparity maps leveraging the Open3D library.
- Implemented the "[Stereo Anything](#)" research paper for disparity map generation.
- Optimized the disparity map generation model using TensorRT, achieving real-time performance with inference times under 5 milliseconds.
- Integrated the YOLOv8 segmentation model to delineate object boundaries and subsequently estimated the distance between the robotic arm and the target object using the generated depth maps..
- Implemented an Image-Based Visual Servoing (IBVS) controller for a robotic arm, utilizing real-time visual feedback to achieve accurate object grasping and manipulation tasks.
- Enabled dynamic, real-time adjustment of the robotic arm's position and orientation based on visual feedback, significantly enhancing its object handling and relocation accuracy.

Drone (Quadcopter) | Pixhawk, BLDCs, F450 drone frame, 40 Amp ESCs, Telemetry device, 6200 mAh battery, GPS & Compass Module, Altimeter

- Acquired an in-depth understanding of key components, such as **brushless motors and ESCs**, essential for enhancing the flight time and telemetry range of quadcopters.
- Utilized the **Pixhawk open-source board**, renowned for **autonomous vehicles**, as a foundational framework to construct a highly capable **quadcopter with advanced functionalities.**
- Designed and constructed a **quadcopter** that could fly higher than a seven-story building, incorporating a **power distribution board, ESCs, brushless motors, and a Pixhawk control board.**
- Successfully configured the Pixhawk control board to achieve a **flying range of 2.8 kilometres**, constant **telemetry input to the ground control station**, and a flight time of 20 minutes (**6200mAh battery**) with a **payload capacity of 6 kilograms.**

Image Super Sampling | Python, PyTorch, OpenCV, CUDA

- Implemented a Deep Learning (DL) Model grounded in the **Super Resolution Generative Adversarial Network (SRGAN) research paper**, achieving notable success in upscaling a 720p movie to a high-resolution 4K format.
- **Optimized inference memory usage**, reducing GPU memory consumption by 40% compared to baseline.
- Enabled efficient processing of large videos without encountering OOM errors.
- Implemented **dynamic memory allocation** mechanisms to seamlessly transfer data **between CPU and GPU memory** during inference.

Real-Time Noise Cancellation in Audio Signals using Deep Learning | Python, PyTorch, TorchAudio, Librosa

- Utilized Mozilla's Common Voice Dataset for clean speech and the UrbanSound8k dataset as real-life noise to develop an algorithm employing Short-Time Fourier Transform (STFT) for intelligent mixing of the two, resulting in a noise-added speech signal.
- Implemented a **Wave-UNet-based DL model**, incorporating **One Cycle Learning Rate (LR)** to expedite training and achieve a global minimum faster. The trained model demonstrated impressive results, yielding an **MSE loss of approximately 0.2**

Awards, Achievements & Extra Curriculars

- Awarded **Best Team Award** in the TCS Rapid Labs Cohort, demonstrating exceptional technical and collaborative skills.
- Served as a **juror on the 2023 LaEleuvia Hackathon Event**, attended by over 80 students, showcasing technical expertise and leadership potential.
- **Organizer and Mentor** in Innovation Centre Club.
- Actively involved and organized AI & robotics education and outreach programs, showcasing robotics projects at inter-university events, organizing workshops, and mentoring students

Publications & Patents

- Kumar, Ch Praveen, K. Praveen Kumar, KSS Vamsi Krishna, and T. Shwetha. "Noise Cancellation and Speech Enhancement Using Multi-Layer Convolutional Neural Network." Design Engineering (2021): 921-926.
- System for Quality Assessment of Agricultural Product(436541)
- A method and system for encoding and decoding the data by using markers (420903)
- An Adjustable Apparatus for Uniform Image Acquisition. (Provisional - 202221022993)
- Generic Classification using Very Less Data (filed)
- Cattle Identification using Face Recognition. (filed)

Certifications

- [Deep Learning Specialization](#), [Convolutional Neural Networks](#), [Sequence Models](#), [Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization](#), [Structuring Machine Learning Projects](#)