K HARI KRISHNA

COMPUTER VISION ENGINEER

+91 9177595499

smarth326@gmail.com

Mahbubnagar, Telangana, India

SUMMARY

Efficient and hands-on AI Systems Engineer with strong expertise in **AI model deploymen**t, **TensorRT-based optimization**, and **image/video processing** for real-time applications. Skilled in integrating pretrained deep learning models, **accelerating inference pipelines,** and **reducing system latency through lightweight engineering solutions.** Experienced in transforming traditional pipelines into faster, resource-efficient systems using tools like Docker, Pillow, and multiprocessing, even without deep customization of deep learning frameworks.

EDUCATION

Bachelor of Commerce (B.Com) in Computer Applications

Government Degree College, Khairatabad, India
2019-2022

TECHNICAL SKILLS

- **Languages:** Python (Strong), Basic C++.
- **Model Integration:** TensorRT, YOLOv7 (Pretrained)
- **Libraries/Tools:** OpenCV, Pillow, FFmpeg, MediaMTX, PyTorch (basic usage).
- **Systems & Tools:** Docker, RabbitMQ, Redis, Git, Label Studio, Django (API-level)
- **Performance Optimization:** Parallel Processing (Multithreading, Multiprocessing), Real-Time Pipeline Efficiency

ACHIEVEMENTS

- Awarded "**Innovator of the Month**" for successfully optimizing platform components and contributing impact R&D solutions.
- Played a key role in bench-marking and transitioning major workloads to GPU.

PROFESSIONAL EXPERIENCE

Computer Vision Intern → Computer Vision Engineer

ZestIoT Technologies Pvt. Ltd., Hyderabad, Telangana, India.
Feb 2023 – Present

- Started as a **Computer Vision Intern** in Feb 2023 and converted to a full-time **Computer Vision Engineer** in Sep 2023 after consistently contributing to high-impact R&D and production solutions.
- Integrated **YOLOv7-based object detection** and **segmentation** models into the company's AI platform.
- Performed **TensorRT optimization** for object detection and segmentation models**,** significantly improving inference speed and reducing hardware utilization.
- Converted pretrained models from **PyTorch → ONNX → TensorRT** and deployed them in real-time systems.
- Replaced OpenCV with Pillow to accelerate image processing, achieving a **2× performance boost** and lower memory usage.
- Built and deployed an **image augmentation pipeline** as a **Dockerized microservice** integrated with the platform.
- Resolved threading and multiprocessing issues to enable efficient, parallel execution in real-time pipelines.
- **Resolved library and package conflicts to successfully** run detection and segmentation models in a single production container, reducing memory usage, startup time, and deployment complexity.