SUBHANSHU SETHI

sethisubhanshu@gmail.com · github.com/subhanshusethi · linkedin.com/subhanshusethi
+91 9958777303

WORK EXPERIENCE

Novus Hi-Tech *Computer Vision Intern*
Jun 2024 – Present

- Implemented deep learning-based **object detection and tracking systems** using **YOLO, ByteTrack, and Kalman filters**, resulting in a **20% accuracy improvement** on NVIDIA edge devices.
- Accelerated **pose estimation pipelines** using **RANSAC** and **TensorRT**, achieving real-time performance.
- Engineered **Dockerized, scalable solutions** for deployment on autonomous pallet picker robots.

EDUCATION

B.Tech in Electrical Engineering, Delhi Technological University (DTU) *2021 – 2025*
Relevant Coursework: Computer Vision, Machine Learning, Deep Learning Specialization (Andrew Ng)

SKILLS

Programming Languages: Python, C++, SQL
Frameworks & Libraries: PyTorch, TensorFlow, Transformers (HuggingFace), OpenCV, TensorRT
Tools & MLOps: Docker, GitHub, MLflow, Weights & Biases (WandB)
Hardware Platforms: Jetson Nano, Intel NUC, Raspberry Pi
Machine Learning Techniques: CNNs, LSTMs, NLP, LLM finetuning, Quantization

PROJECTS

Image Captioning *(Paper Under Review)* *Dec 2023 – Nov 2024*
GitHub

- Achieved a **BLEU-4 score of 54** by developing a multimodal Encoder-Decoder model combining **CLIP vision encoder** and **GPT-2 decoder**, enhanced with a novel attention mechanism and **Contrastive Learning** for improved semantic alignment.
- Designed and optimized deep learning pipelines for **image-to-text generation**, evolving from **CNN-LSTM baselines** to **Transformer-based architectures**, enabling superior contextual understanding and sequence fluency.

Grammar Scoring Engine *Mar 2025 – Apr 2025*
GitHub

- Developed a pipeline by **fine-tuning Gemma-3 and LLaMA-3 models**, and integrating OpenAI's **Whisper**, achieving a **Pearson correlation score of 0.82** for grammar evaluation from speech.
- **Reduced inferencing time by 50%** through model quantization.

Multimodal Classification *(Paper Accepted)* *Feb 2025 – Apr 2025*
GitHub

- Designed a **multimodal classification model** by integrating image and text modalities, employing a **Mixture of Experts** module and an enhanced **Attention Mechanism** to capture both local and global contextual cues.
- **Achieved state-of-the-art performance with a reduced parameter count** by integrating a frozen **CLIP image encoder** and a **ModernBERT text encoder**.

ACHIEVEMENTS

- **3rd Place** – International Micro Aerial Vehicle Competition, Aachen, Germany (*Nov 2023*)
  Created a **lane-following drone pipeline** with onboard **mapping** for autonomous **search and rescue**.
- **Top 10 / 19,000+ teams** – Flipkart Grid Robotics Challenge, Bengaluru (*Jan 2025*)
  Built a **production-ready feature extractor for FMCG products** using **VLMs like Qwen**, and implemented **object counting and segmentation** pipelines for **zero-shot retail imagery analysis**.

PUBLICATIONS

- **ContXCLIP: Contextual Attention for Vision-Language Understanding**
  Under review at *Signal Processing: Image Communication (Elsevier)*
  Proposed **dual-attention** and **context-preserving** modules integrating **CLIP + GPT-2**, enhanced via **contrastive learning**.
- **GEXSent: Gated Experts for Robust Sentiment Analysis Across Modalities**
  Accepted at *IJCNN*, Rome, Italy, 2025
  Introduced a novel **Mixture of Experts** based module and **Gated Attention** to combine **CLIP-based visual features** and **MordernBERT based textual features** for robust sentiment analysis.