Abdullah Usama

ausama.bese22seecs@seecs.edu.pk — 03088404523 — LinkedIn — GitHub — Portfolio

Education

National University of Sciences and Technology (NUST), Islamabad
Bachelor of Software Engineering, SEECS - 3rd Year

- **Coursework:** Data Structures and Algorithms, Object Oriented Programming, Web Engineering, Operating Systems, Database Systems, Design and Analysis of Algorithms, Machine Learning, Embedded Systems

Experience

ML Intern - OneScreen Solutions, San Diego, California – Remote *April 2025 – May 2025*

- Worked with Vision Transformers (ViT-32) and Vision-Language Models (VLMs) like PaLI-Gemma to explore multimodal learning and representation.
- Engineered a method to reconstruct full lecture images by removing occlusions: detected persons using YOLO, segmented with SAM, and seamlessly removed them from video frames for clean image synthesis.
- Achieved higher mAP by almost 7 - 10 % by reducing label noise and improving object localization.
- Refined YOLO-format bounding boxes using SAM's multi-class segmentation masks, improving spatial precision and label accuracy.
- Developed an end-to-end pipeline to combine SAM's pixel-level masks with YOLO annotations, optimizing datasets for multi-class detection tasks.

Computer Vision Intern - Machine Vision & Intelligent Systems Lab (MachVis), SEECS

June 2024 – Aug 2024

- Engineered and optimized computer vision pipelines for real-time object detection, tracking, and motion estimation in dynamic environments.
- Designed and implemented robust feature extraction methods, including ORB and optical flow algorithms, to estimate camera motion and stabilize frame sequences.
- Developed and optimized real-time tracking systems using SORT and Kalman Filters to maintain consistent object identities across video frames.

Projects

MACHINE LEARNING, COMPUTER VISION AND AI

News AI Agent *(Website) (GitHub)*
Developed an intelligent web-based assistant to help students and CSS/PMS aspirants understand and analyze DAWN - Op & Ed, The Tribune, ParadigmShift - National & IR by offering topic - wise batch search, vocabulary insights, idioms, summaries, and article URLs through an AI-powered backend and interactive frontend.

- Built a **FastAPI-based backend API** integrated with LangChain and deployed on Render, enabling natural language interactions via Gemini 2.0 Flash.
- Implemented **article scraping and content extraction** tool to fetch DAWN editorials with their titles, URLs, and full content for specified or relative dates.
- Supported **flexible date handling** to interpret inputs like "today", "yesterday", and "last week" and convert them into valid date formats for scraping.

Skills: FastAPI, LangChain, Google Gemini 2.0 Flash, Next.js, Web Scraping, REST APIs, Natural Language Processing, Vocabulary Analysis, Idiom Extraction, Deployment (Render, Vercel), CORS, Date Parsing.

Bounding Box Refinement Pipeline *(GitHub) (My Article)* *Apr 2025 – May 2025*
Developed a data preprocessing pipeline to refine loose YOLO-format bounding boxes by integrating pixel-accurate masks from the Segment Anything Model (SAM), significantly improving object localization and detection performance.

- Leveraged SAM to extract fine-grained segmentation masks and aligned them with YOLO-detected objects to improve annotation precision.
- Merged SAM-generated masks with original YOLO labels, enabling tighter and more spatially accurate bounding boxes.
- Enhanced dataset quality for multi-class object detection tasks, resulting in improved mean Average Precision (mAP) on downstream models.

Skills: Object Detection, YOLO, Segment Anything Model (SAM), Bounding Box Optimization, Data Preprocessing, Computer Vision, Model Evaluation

Mistral-7b-Instruct Finetuning (LLM Fine-tuning) *(Hugging Face) (GitHub)* *June 2025*
Fine-tuned Mistral-7B-Instruct-v0.3 using LoRA on a custom dataset of opinion articles by Pakistani diplomat and journalist Maleeha Lodhi, enabling stylistically accurate generation of geopolitical commentary and current affairs analysis.

- Scraped and filtered 6 years of Dawn opinion articles to curate a dataset focused on Maleeha Lodhi's writing.
- Structured the dataset in instruction-response (JSONL) format and applied Parameter-Efficient Fine-Tuning (LoRA).
- Achieved consistent stylistic mimicry in model outputs, suitable for research, educational, and editorial applications.

Skills: Large Language Models, Mistral-7B, LoRA, PEFT, Data Scraping, JSONL Formatting, Prompt Engineering, NLP Fine-Tuning

Football Video Analysis *(GitHub)* *July 2024 - Aug 2024*
Developed a comprehensive football match analysis system that integrates multiple computer vision techniques to track player movement, estimate distances, and analyze ball possession in real-time.

- Engineered a **camera movement estimator** using feature detection and optical flow techniques to stabilize footage and improve tracking accuracy.
- Implemented a **player and ball detection module** leveraging YOLO and Faster R-CNN for precise object identification and SORT tracking for consistent player identity maintenance across frames.
- Engineered a **field analysis pipeline** using homography transformations, Euclidean distance metrics, and color-based clustering (K-means) to measure player movement, assign team identities, and generate a top-down tactical view of the field.

Skills: Computer Vision, Deep Learning, Object Detection, Object Tracking, Homography Estimation, Optical Flow, YOLO, Faster R-CNN, SORT Tracking, Color-Based Clustering, OpenCV, NumPy, Video Processing, Geometric Transformations, Perspective Correction.

Hand Gesture Volume Control *(GitHub)* *July 2024 - Aug 2024*
Developed a real-time system that enables hands-free volume control using dynamic hand gestures, integrating computer vision techniques for precise landmark detection and gesture recognition.

- Implemented **hand detection and tracking** using MediaPipe's pre-trained model to accurately recognize and track 21 hand landmarks in real time.
- Designed a **gesture recognition system** that computes the Euclidean distance between the thumb and index finger tips to determine volume levels dynamically.
- Integrated **system volume adjustment** via the Windows Core Audio API using Pycaw, enabling seamless and instant control of system audio.

Skills: Computer Vision, Hand Gesture Recognition, Real-Time Video Processing, MediaPipe, OpenCV, Pycaw, NumPy, System Audio Control

WEB DEVELOPMENT

Plant E-Commerce App (MERN) *(GitHub)* *Dec 2024 – Jan 2025*
Developed a plant e-commerce platform with secure authentication, multilingual support, and seamless payment integration, enhancing user experience and accessibility.

- Implemented **role-based authentication** using Clerk.com, ensuring secure and personalized user experiences with access control.
- Integrated **multi-language support** (English/Urdu) using React-i18n, allowing users to switch languages dynamically.
- Designed and integrated **Stripe-based payment processing**, enabling secure and reliable financial transactions.
- Built a **responsive frontend** with React.js, a **scalable backend** using Node.js and Express.js, leveraged **MongoDB** for efficient storage, and developed an **admin panel** for store management.

Skills: MERN Stack, Authentication & Authorization, Payment Gateway Integration, Multilingual Support, React.js, Node.js, Express.js, MongoDB, Clerk.com, Stripe, UI/UX Design, RESTful APIs.

Video-Stream App *(GitHub)* *Dec 2024 – Jan 2025*
Developed a cloud-based video streaming application utilizing microservices architecture, enabling efficient media delivery with authentication, real-time processing, and scalable cloud deployment.

- Designed a **microservices architecture** comprising authentication, video streaming, storage, monitoring, and logging services.
- Implemented **secure authentication** via Clerk.com and JWT-based authorization to control access and protect media content.
- Integrated a **controller microservice** for request validation and service coordination, and deployed
- Conducted **load testing** to assess system performance, achieving a p99 latency of 30.33 ms under peak traffic conditions.

Skills: Microservices Architecture, Video Streaming, React.js, Google Cloud Run, Clerk Authentication, JWT Authorization, Google Cloud Storage, API Gateway, Firebase Firestore, Scalable Backend Systems, System Monitoring, Load Testing.

Search Engine (Python) *March 2024 – May 2024*
Developed a scalable search engine inspired by Google's research, capable of processing and retrieving information from 100,000+ articles using advanced indexing and retrieval techniques.

- Developed a **forward and inverted indexing system** for fast document retrieval, implemented **lemmatization** for search accuracy, and designed a **dynamic lexicon** for real-time vocabulary updates.

Skills: Information Retrieval, Indexing Techniques, Forward Index, Inverted Index, Lemmatization, Search Optimization, Python, React.js, JSON Data Storage, Docker, Data Processing, Large-Scale Search Systems.

Skills

- **Programming:** JavaScript, Python, C++, SQL
- **Frameworks & Libraries:** React.js, Next.js, Node.js, Express.js, OpenCV, MediaPipe, TensorFlow, LangChain, LangSmith
- **Databases:** MongoDB, MySQL, PostgreSQL
- **Cloud & Deployment:** Docker, RESTful APIs, Vercel, Render
- **Tools:** Git, GitHub, Clerk.com, Stripe, Pycaw, React-i18n

Certificates

- **Machine Learning Specialization** *June 2024 DeepLearning.AI (Coursera)*
- **Advanced Learning Algorithms** *June 2024 Coursera*
- **Supervised Machine Learning (Regression & Classification)** *June 2024 Coursera*
- **Unsupervised Learning, Recommenders, Reinforcement Learning** *June 2024 Coursera*
- **Version Control** *October 2023 Meta (Coursera)*
- **Responsive Web Design** *September 2023 freeCodeCamp*