Gheffari Youcef Soufiane

+213552598134 | gheffari.youcef.soufiane@gmail.com | linkedin.com/in/gheffari-youcef-soufiane | github.com/youcefgheffari

Summary

Postgraduate student in Artificial Intelligence with hands-on experience in deep learning, computer vision, and natural language processing. Proficient in building and deploying end-to-end machine learning systems using Python, PyTorch, and TensorFlow. Strong analytical thinker with a research-oriented mindset, solid software engineering foundations, and a proven ability to work in multidisciplinary teams. Actively seeking opportunities to contribute to impactful AI research and real-world applications across academic and industrial settings.

Education

University Center of Naama — Naama, Algeria
Bachelor of Science in Computer Science — *Oct 2020 – May 2023*
Core focus on algorithms, data structures, software engineering, and mathematical foundations.

Completed projects in systems programming, databases, and web development.

University of Science and Technology of Oran (USTOMB) — Oran, Algeria
Master of Science in Artificial Intelligence and Its Applications — *Oct 2023 – May 2025*
Specialized in deep learning, computer vision, and natural language processing.

Completed advanced coursework in neural networks, reinforcement learning, and probabilistic models.

Thesis: Arabic Speech Emotion Recognition System (Wav2Vec2, CNN, Transformer, LSTM).

Master's Thesis Project

Arabic Speech Emotion Recognition System
Master's Thesis — MSc in Artificial Intelligence and Its Applications — *Oct 2023 – May 2025*

Designed and implemented a deep learning system to classify emotions in Arabic speech using both acoustic and contextual cues.

Leveraged the **EYASE (Egyptian Arabic Speech Emotion)** dataset of labeled Modern Standard Arabic utterances.

Extracted audio representations using **Wav2Vec2** and trained multiple deep architectures:

* CNN-based classifier — **75.2% Accuracy**
* CNN + BiLSTM with attention — **84.6% Accuracy**
* CNN + Transformer — **93.4% Accuracy**

Performed comparative analysis of models in terms of generalization, F1-score, and inference speed.

Demonstrated that Transformer-based architecture significantly outperformed alternatives in both accuracy and robustness.

Research Projects

DrivingRecon 4D Scene Reconstruction
4D Reconstruction and Neural Rendering — *Python, PyTorch, OpenCV, Panda3D*

Implemented a simplified version of the DrivingRecon paper using multi-view driving video.

Built a 4D Gaussian rendering pipeline with depth estimation, 3D encoding, and temporal attention.

Achieved real-time scene reconstruction and novel view synthesis from synthetic or real data.

4D Scene Prediction and Trajectory Reasoning
Predictive Computer Vision Project — *YOLOv8, GRU, Panda3D, GPT-4, PyTorch*

Built a predictive 4D scene understanding system combining object tracking, camera pose alignment, and GRU-based future motion estimation.

Converted 2D tracks into real-world 3D trajectories using TUM RGB-D data and depth backprojection.

Trained a GRU model to forecast object motion 5 steps into the future and visualized results in 3D using Panda3D.

Integrated GPT-4 for natural language explanation of predicted behaviors such as crossing, stopping, or accelerating.

Text-to-3D Scene Synthesis via Diffusion-Guided NeRF
Generative Vision Project *Stable Diffusion, NeRF, Score Distillation Sampling, PyTorch*

- Implemented a system to generate 3D scenes from textual prompts using Score Distillation Sampling (SDS) and Stable Diffusion.
- Integrated NeRF rendering pipeline with view consistency to ensure multi-view coherence across novel perspectives.
- Trained a Gaussian-based 3D volume using optimization guided by a frozen text-to-image diffusion model.
- Achieved high-fidelity, coherent 3D scenes from single-sentence text prompts using gradient backpropagation from rendered views.
- Inspired by research such as DreamFusion, Magic3D, and Progressive Gaussian Splatting.

Semi-Supervised Semantic Segmentation with Self-Correction (CVPR 2020 Reproduction)
Computer Vision Paper Reproduction *Python, PyTorch, Scikit-learn, CNN*

- Re-implemented and trained dual-network segmentation system using PyTorch.
- Integrated ancillary and primary segmentation networks with self-correction (linear & CNN-based).
- Achieved segmentation performance competitive with supervised baselines.
- Applied on Pascal VOC 2012 dataset using weak and full annotations.

LLM-based Visual Grounding in 3D Driving Scenes
AI Research Project — Self-Initiated *Python, PyTorch, Detectron2, GPT-4, Panda3D*

- Built a complete 3D scene understanding pipeline using RGB-D frames from the TUM dataset, leveraging object detection, depth backprojection, and visual reasoning.
- Constructed dynamic scene graphs capturing spatial relationships and queried them using GPT-4 to identify objects based on free-form natural language input.
- Mapped GPT-identified objects back to 3D point clouds and visualized them interactively in real-time using Panda3D.
- Exported point clouds and scene graphs for demo-ready `.ply` and `.json` formats, enabling portfolio visualization and reproducibility.
- Achieved real-time 3D semantic understanding and language-based scene exploration in autonomous driving contexts.

Engineering Projects

Brain Tumor Classification System
Deep Learning Project *Python, PyTorch, Scikit-learn, CNN, Autoencoder*

- Designed and implemented a pipeline for classifying brain tumors using MRI scans.
- Applied CNNs and autoencoders for feature extraction and tumor classification.
- Combined traditional descriptors (HOG, LBP, SIFT, ORB) with deep learning features.
- Used PCA to reduce dimensionality and improve classification performance.
- Achieved high accuracy in distinguishing glioma, meningioma, pituitary tumors, and healthy images.

Group Activity Recognition
Deep Learning Project *Python, PyTorch, Scikit-learn, CNN, LSTM*

- Designed and implemented a hierarchical deep learning pipeline for recognizing group activities in videos.
- Applied CNNs for person-level feature extraction and LSTMs for modeling group-level temporal dynamics.
- Trained and evaluated on the Collective Activity Dataset (Crossing, Waiting, Queueing, Walking, Talking).
- Achieved accurate group activity classification through a two-stage CNN+LSTM architecture.

Multi-Object Tracking + 3D Reconstruction in Driving Scenes
Advanced Computer Vision Project *Python, PyTorch, YOLOv8, DeepSORT, TUM RGB-D, TensoRF*

- Built a full object-centric 3D reconstruction pipeline combining detection, tracking, pose fusion, and volume optimization.
- Used **YOLOv8 (COCO)** for object detection and **DeepSORT** for real-time multi-object tracking.
- Cropped tracked object images and associated them with **groundtruth camera poses** from the TUM freiburg1_desk dataset.
- Trained a dummy **TensoRF-style 3D tensor volume** per object using image supervision and pose alignment.
- Visualized learned volumetric slices and exported intermediate `tensor.pt` files for downstream 4D reconstruction.

Multimodal Instruction-Following Agent in 3D Environments
AI Engineering Project *Habitat-Sim, CLIP, BERT, PyTorch, Python*

Engineered a real-time multimodal agent capable of executing 3D navigation tasks based on natural language instructions.

Integrated CLIP for vision encoding and BERT for language understanding to ground instructions in the agent's 3D visual context.

Simulated indoor environments with Habitat-Sim, enabling realistic physics-based navigation and rendering.

Designed and trained a policy network that maps language + visual embeddings to movement commands using supervision from expert demonstrations.

Achieved accurate and generalizable goal-directed navigation in unseen rooms with instruction-conditioned behavior.

Technical Proficiencies

Programming Languages: Python, SQL, JavaScript, C

Development Tools & Platforms: Git, Jupyter, Google Colab, VS Code, PyCharm, Linux (Ubuntu/Debian)

Software Engineering: Object-Oriented Programming, Git Version Control, Debugging

AI and Machine Learning Toolkits

Deep Learning Frameworks: PyTorch, TensorFlow, Keras

Machine Learning Libraries: Scikit-learn, NumPy, Pandas, Seaborn, Matplotlib

Computer Vision Tools: OpenCV, YOLOv8, Panda3D

Natural Language Processing (NLP): NLTK, Hugging Face Transformers, Wav2Vec2, LangChain

3D and Visual Computing: Gaussian Splatting, Open3D (basic), Panda3D

Experiment Tracking & Visualization: Weights & Biases (W&B)

Professional Competencies

Analytical Thinking: Able to decompose complex AI problems and design scalable, research-backed solutions.

Scientific Communication: Proficient in presenting ideas clearly through technical writing, research documentation, and visualizations.

Collaboration & Teamwork: Experienced in cross-functional teamwork across academic projects, AI engineering tasks, and open-source contributions.

Self-Directed Learning: Continuously stay up-to-date with state-of-the-art research and implement cutting-edge tools independently.

Project & Time Management: Skilled at managing multiple deadlines across research and engineering workflows using agile practices.

Attention to Reproducibility: Prioritize code quality, reproducible results, and structured experimentation in all AI/ML projects.

Certifications

Machine Learning Process A-Z, 365 Data Science

Machine Learning Algorithms A-Z, 365 Data Science

Machine Learning Specialization, Coursera — Andrew Ng (DeepLearning.AI)

Data Science & Machine Learning Foundations, Maven Analytics

Core Research Domains

Computer Vision and 3D Perception: Representation learning from images, videos, and depth data; neural rendering; geometric deep learning; object detection, segmentation, and 3D scene reconstruction.

Multimodal Learning: Joint modeling of vision, language, and audio to enhance cross-modal understanding, robustness, and generalization in complex environments.

Natural Language Processing (NLP): Focus on semantic understanding, contextual embeddings, and Arabic NLP in low-resource settings.

Speech Processing: Emotion recognition and automatic speech recognition (ASR), integrated as supporting modalities in multimodal systems.

Applied AI Focus Areas

Vision-Language Models (VLMs): Leveraging models like CLIP, GPT-4V, and Flamingo for visual grounding, image captioning, referring expression comprehension, and multimodal reasoning.

Explainable AI (XAI): Designing interpretable and trustworthy AI systems for sensitive domains such as healthcare, autonomous driving, and safety-critical applications.

Real-Time Perception Systems: Efficient CV + NLP pipelines optimized for deployment in autonomous systems, including object-centric 3D tracking and multimodal scene understanding.

Research Vision

Advance the development of **multimodal foundation models** capable of reasoning jointly over vision, language, and audio with minimal supervision and enhanced generalization.

Explore **4D dynamic scene understanding** and spatiotemporal perception for real-time applications in autonomous driving, robotics, and AR/VR.

Design **interpretable and reliable AI systems** tailored for high-stakes domains such as healthcare, autonomous systems, and AI-assisted decision making.

Bridge the gap between **vision-language understanding and embodied AI**, enabling agents to perceive, communicate, and act intelligently in real-world 3D environments.

Promote inclusive AI through research in **low-resource languages** and culturally diverse multimodal datasets, with a focus on Arabic NLP and multilingual scene understanding.

Languages

Arabic — Native

French — Fluent

English — Professional Working Proficiency

Publications

G. Y. Soufiane, "Arabic Speech Emotion Recognition using Deep Learning", *Manuscript in Preparation*, Target: September 2025 Submission.