PERSONAL INFORMATION

Name:	Abdelrahman Shaker
Date of birth:	01/10/1994
Nationality:	Egyptian
Residence:	Masdar City, Abu Dhabi, UAE (Golden Visa Holder)
Phone number:	+971-527096639
Email address:	abdelrahman.m.shaker@gmail.com
Google Scholar:	https://scholar.google.com/citations?user=eEz4Wu4AAAAJ
GitHub website:	https://amshaker.github.io/
LinkedIn profile:	www.linkedin.com/in/abdel-rahman-shaker-74274b9b

PROFILE SUMMARY

I have a mixed experience between conducting pure academic research and contributing to international products in the industry. I have published over 10 research papers with ~700 citations and contributed to more than 5 international projects. My research interest is efficient Computer Vision, specially designing reliable, accurate, lightweight, and memory-efficient 2D and 3D architectures for edge devices.

EDUCATION

Mohamed bin Zayed University of Artificial Intelligence, UAE *Aug 2021 – Jun 2025*
PhD. in Computer Vision
CGPA: 3.95/4.0

Faculty of Computer Sciences, Ain-Shams University, Egypt *Aug 2017 – Jul 2020*
MSc. in Deep Learning
CGPA: 3.73/4.0

Faculty of Computer Sciences, Ain-Shams University, Egypt *Sep 2012 – Aug 2016*
BSc. in Computer Sciences
Excellent with honors grade. Department Rank: 1st (out of 149 students)

SELECTED RESEARCH PUBLICATIONS

Efficient Video Object Segmentation via Modulated Cross-Attention Memory *(WACV 2025)*

Abdelrahman Shaker, *Syed Talal , Martin Danelljan, Salman Khan, Ming-Hsuan Yang, Fahad Khan*

We propose an efficient transformer-based approach, named MAVOS, that introduces an optimized and dynamic long-term modulated cross-attention (MCA) memory to model temporal smoothness without requiring frequent memory expansion. The proposed MCA effectively encodes both local and global features at various levels of granularity while efficiently maintaining consistent speed regardless of the video length. Extensive experiments on multiple benchmarks, LVOS, Long-Time Video, and DAVIS 2017, demonstrate the effectiveness of our proposed contributions leading to real-time inference and markedly reduced memory demands without any degradation in segmentation accuracy on long videos. Compared to the best existing transformer-based approach, our MAVOS increases the speed by 7.6X, while significantly reducing the GPU memory by 87% with comparable segmentation performance on short and long video datasets.

PALO: A Polyglot Large Multimodal Model for 5B People *(WACV 2025)*

*Muhammad Maaz, Hanoona Rasheed, **Abdelrahman Shaker**, Salman Khan, Hisham Cholakkal, Rao M. Anwer, Timothy Baldwin, Michael Felsberg and Fahad Khan*

In pursuit of more inclusive Vision-Language Models (VLMs), this study introduces a Large Multilingual Multimodal Model called PALO. PALO offers visual reasoning capabilities in 10 major languages, including English, Chinese, Hindi, Spanish, French, Arabic, Bengali, Russian, Urdu, and Japanese, that span a total of ~5B people (65% of the world population).

UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation *(IEEE TMI - 2024)*

***Abdelrahman Shaker**, Muhammad Maaz, Hanoona Rasheed , Salman Khan, Ming-Hsuan Yang, Fahad Khan*

This work introduces a 3D medical image segmentation approach, named UNETR++, that offers both high-quality segmentation masks and efficiency in terms of parameters and compute cost. The core of our design is the introduction of a novel efficient paired attention (EPA) block that efficiently learns spatial and channel-wise discriminative features using a pair of inter-dependent branches based on spatial and channel attention.

GLaMM: Pixel Grounding Large Multimodal Model *(CVPR - 2024)*

*Muhammad Maaz, Hanoona Rasheed , Sahal Shaji, **Abdelrahman Shaker**, Salman Khan, Hisham Cholakkal, Rao M. Anwer, Eric Xing, Ming-Hsuan Yang, Fahad Khan*

Grounding Large Multimodal Model (GLaMM) is an end-to-end trained LMM which provides visual grounding capabilities with the flexibility to process both image and region inputs. This enables the new unified task of Grounded Conversation Generation that combines phrase grounding, referring expression segmentation, and vision-language conversations. Equipped with the capability for detailed region understanding, pixel-level groundings, and conversational abilities, GLaMM offers a versatile capability to interact with the visual inputs.

SwiftFormer: Efficient Additive Attention for Transformer-based Real-time Mobile Vision Applications *(ICCV - 2023)*

***Abdelrahman Shaker**, Muhammad Maaz, Hanoona Rasheed , Salman Khan, Ming-Hsuan Yang, Fahad Khan*

We introduce in this work a novel efficient additive attention mechanism that effectively replaces the quadratic matrix multiplication operations of the self-attention with linear element-wise multiplications. Using our proposed efficient additive attention, we build a series of models called "SwiftFormer" which achieves state-of-the-art performance in terms of both accuracy and mobile inference speed.

Arabic Mini-ClimateGPT : A Climate Change and Sustainability Tailored Arabic LLM *(EMNLP - 2023)*

**Sahal Shaji, *Abdelrahman Shaker, Omkar Thawakar, Hisham Cholakkal, Rao M. Anwer, Salman Khan, Fahad Khan*

We propose a light-weight Arabic Mini-ClimateGPT that is built on an open-source LLMs and is specifically fine-tuned on a conversational-style instruction tuning curated Arabic dataset Clima500-Instruct with over 500k instructions about climate change and sustainability. Further, our model also utilizes a vector embedding based retrieval mechanism during inference. We validate our model through quantitative and qualitative evaluations on climate-related queries.

EdgeNeXt: Efficiently Amalgamated CNN-Transformer Architecture for Mobile Vision Applications *(ECCVW- 2022)*

Muhammad Maaz, ***Abdelrahman Shaker, Hanoona Rasheed , Salman Khan, Ming-Hsuan Yang, Fahad Khan*

This work is a new hybrid architecture that effectively combine the strengths of both CNN and Transformer models. We introduce split depth-wise transpose attention (SDTA) encoder that splits input tensors into multiple channel groups and utilizes depth-wise convolution along with self-attention across channel dimensions to implicitly increase the receptive field and encode multi-scale features.

INSTA-YOLO: Real-time Instance Segmentation *(ICMLW- 2021)*

Eslam Mohamed, ***Abdelrahman Shaker, Ahmad El-Sallab, Mayada Hadhoud*

We propose Insta-YOLO, a novel one-stage end-to-end deep learning model for real-time instance segmentation. Instead of pixel-wise prediction, our model predicts instances as object contours represented by 2D points in Cartesian space. We evaluate our model on three datasets, namely, Carvana, Cityscapes and Airbus. We compare our results to the state-of-the-art models for instance segmentation.

Generalization of Convolutional Neural Networks for ECG Classification Using Generative Adversarial Networks. *(IEEE Access - 2020)*

***Abdelrahman Shaker**, Manal Tantawi , Howida Shdeed, Mohamed Fahmy Tolba*

We propose a novel data augmentation technique based on generative adversarial networks (GANs) to address data imbalance problem. We also introduce two deep learning approaches: an end-to-end method and a two-stage hierarchical method. Our results demonstrate that augmenting the original imbalanced dataset with GAN-generated heartbeats significantly enhances ECG classification performance compared to training the same techniques solely on the original dataset. Results show that our method outperforms other standard augmentation techniques.

WORK EXPERIENCE

Huawei Finland R&D *Feb 2024 – Sep 2024*

Applied Researcher / Efficient Computer Vision

I am currently engaged in a 6-month hybrid role, focusing on developing innovative and efficient solutions for Computer Vision challenges to be deployed on edge devices. My responsibilities include designing tiny architectures, applying post-training optimization, quantization, developing efficient multi-scale samplers, and implementing distillation techniques.

Valeo Egypt *Sep 2019 – Jul 2021*

Machine Learning / SW Engineer

I have contributed to the development and enhancement of advanced driver assistance systems (ADAS) and autonomous driving technologies. My role involved integrating cutting-edge deep learning models, improving system accuracy and efficiency, and collaborating cross-functionally to meet stringent customer requirements. Below are some of the key projects I worked on:

- **Driver Monitoring System (DMS):** Integral contributor to the DMS Daimler project, focusing on driver recognition and alertness within an integrity OS framework. Integrated components from ECP, Vision, and Algorithms teams, aligning with customer requirements (SRS).
- **Insta-YOLO:** Developed a novel one-stage end-to-end DL model for real-time instance segmentation, replacing YOLOv3's box regression with polygon regression in the localization head and proposing a new localization loss. The proposed model is 1.75x faster than YOLACT with comparable accuracy.
- **Seat Occupancy Detection:** Improved safety by detecting and classifying car seats to optimize airbag deployment. Increased the Average Precision (AP) by 4.0% for different versions of YOLO through a new data augmentation method based on mixing image normalization techniques.
- **Object Recognition using Radar:** Developed a prototype for classifying objects from heatmaps using TI radar sensors within a car's interior.

Valeo Egypt *Jan 2019 – Apr 2019*

Deep Learning Researcher

During the internship, I worked on enhancing the autonomous driving capabilities through the development of occupancy grid maps (OGM) based on LIDAR data. My primary focus was on optimizing the processing time for updating the OGM. By implementing a novel approach that updated only the polygon of affected points rather than the entire convex hull, I was able to significantly reduce the processing time from 3.18 milliseconds to 1.28 milliseconds. This optimization also contributed to more accurate and faster real-time environmental mapping, which is crucial for safe and effective autonomous driving.

Ain-Shams University *Mar 2018 – Jul 2021*

Teaching/Lecturer Assistant

As part of my academic involvement at Ain-Shams University, I have contributed to both teaching and course development in the fields of deep learning, computer vision, neural networks, and machine learning. My responsibilities included preparing and delivering lectures, creating and grading assignments, supervising over ten graduation projects, and assisting in exam preparation and grading.

TECHNICAL SKILLS

Artificial Intelligence: Computer Vision, Deep Learning, Machine Learning.

Programming Languages: Python and C++. **Frameworks**: PyTorch and TensorFlow.

Concepts: Object Oriented Programming, Data Structures and Algorithms, and Agile Methodologies.

EXTRA-CURRICULAR AND ACHIEVEMENTS

- Acted as a reviewer for Transformer-based related papers in WACV-2025, NeurIPS-2025, IEEE TPAMI-2024, CVPR-2024, ECCV-2024, IEEE Access.
- Teaching Assistant: *CV701*, *CV703*, *AI701* at Mohamed Bin Zayed University of AI, UAE.
- Mentor in Kaggle competitions for deep learning projects at MBZUAI and Ain-Shams University.
- Ranked 1st in Deep Learning course during Ph.D. at MBZUAI.
- First place in the Innovation Competition 2017, hosted by the Egyptian Ministry of Communications
    - The competition hosted every year for excellent graduation projects for all engineering and computer sciences colleges in Egypt. We had achieved first place in the Data Science track.

CERTIFICATES

- TensorFlow in Practice specialization from deeplearning.ai — *Coursera*
- Deep Learning specialization from deeplearning.ai — *Coursera*
- Machine Learning specialization from the University of Washington — *Coursera*

LANGUAGES

- Arabic: Native or bilingual proficiency.
- English: Professional working proficiency.

REFERENCES

Prof. Fahad Khan
Academic Advisor
Professor at MBZUAI
✉ fahad.khan@mbzuai.ac.ae

Dr. Salman Khan
Academic Advisor
Assoc. Prof.at MBZUAI
✉ salman.khan@mbzuai.ac.ae

Dr. Hesham Eraqi
Senior Scientist (L6)
Amazon Seattle, USA
✉ hesham.eraqi@amazon.com