

## 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](mailto: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](https://www.linkedin.com/in/abdel-rahman-shaker-74274b9b)



## PROFILE SUMMARY

I am a researcher and engineer with a strong blend of academic and industry experience, specializing in efficient Computer Vision and Multi-Modal AI. My work focuses on designing reliable, accurate, and memory-efficient 2D and 3D architectures optimized for edge devices. I have published over 10 research papers in top-tier venues with more than 1,400 citations and have contributed to 5+ international projects.

## EDUCATION

**Mohamed bin Zayed University of Artificial Intelligence, UAE** *Aug 2021 – Jul 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: 1<sup>st</sup> (out of 149 students)

## SELECTED RESEARCH PUBLICATIONS

**Mobile-VideoGPT: Fast and Accurate Baseline for Mobile Video Understanding** [Under Review]

*Abdelrahman Shaker, Muhammad Maaz, Chenhui Gou, Hamid Rezaatofghi, Salman Khan, Fahad Khan*

Current video understanding models often struggle with deployment on edge devices due to their substantial number of parameters and high computational demands. To address this challenge, I am currently working on Mobile-VideoGPT, an efficient multimodal model specifically designed for edge devices with fewer than a billion parameters across all encoders, a practical choice for edge deployment. Mobile-VideoGPT is based on lightweight projectors and encoders to efficiently balance spatial and temporal context and overcomes the limitations of traditional video large multimodal models (LMMs), which either excel in spatial detail capture without temporal awareness or struggle with spatial understanding due to computational constraints.

**GroupMamba: Parameter-Efficient and Accurate Group Visual State Space Model** [CVPR – 2025]

*Abdelrahman Shaker, Syed Talal, Salman Khan, Jurgen Gall, Fahad Khan*

We introduce GroupMamba to enhance the computational efficiency and interaction within state-space models and address the challenges of scaling SSM-based models for computer vision, particularly the instability and inefficiency of large model sizes. To address this, we introduce a Modulated Group Mamba layer which divides the input channels into groups and applies our proposed SSM-based efficient Visual Single Selective Scanning block independently to each group.

## 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 MCA effectively encodes local and global features at various levels of granularity while efficiently maintaining consistent speed regardless of the video length. Extensive experiments on multiple benchmarks demonstrate the effectiveness of the proposed contributions leading to real-time inference and less memory demands. Compared to the best existing transformer-based approach, our method increases the speed by 7.6X, while significantly reducing the GPU memory by 87% with comparable segmentation performance.

## 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]

*Hanoona Rasheed, Muhammad Maaz, 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.

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

We introduce a GAN-based data augmentation technique to tackle data imbalance in ECG classification. Additionally, we propose two deep learning approaches: an end-to-end method and a two-stage hierarchical method. Our results show that GAN-generated heartbeats significantly improve classification performance, outperforming standard augmentation techniques.

## WORK EXPERIENCE

---

### **MBZUAI**

*May 2025 – Now*

*Postdoctoral Researcher Accosiate I | Computer Vision Department*

### **Huawei Finland R&D**

*Feb 2024 – Aug 2024*

*Hybrid Applied Researcher Intern | Efficient Computer Vision*

I am currently focusing on developing innovative and efficient solutions for Computer Vision challenges to optimized for edge devices. My work involves designing tiny architectures, applying post-training optimization, quantization, developing efficient multi-scale samplers, and implementing distillation techniques. I received the President Award for Outstanding Individual Contribution in 2024.

### **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. 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.

### **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

---

- **Huawei FiRC President Award 2024:** Awarded for Outstanding Individual Contribution in 2024
- **PhD Research Achievement Award for Class of 2025 at MBZUAI, Computer Vision Department:** Recognized for exceptional research performance and impactful contributions during the PhD program
- **Research Rockstar Award for Class of 2025 at MBZUAI, Computer Vision Department:** Awarded for outstanding innovation, publication excellence, and leadership in research activities
- Over two million downloads for [EdgeNeXt](#) and [SwiftFormer](#) on HuggingFace.
- Acted as a reviewer for Transformer-based related papers in CVPR-2025, WACV-2025, NeurIPS-2025, IEEE TPAMI-2024, CVPR-2024, ECCV-2024, IEEE Access.
- Teaching Assistant: [CV701](#), [CV703](#), [AI701](#) at MBZUAI, UAE.
- Mentor in Kaggle competitions for deep learning projects at MBZUAI and Ain-Shams University.
- Ranked 1<sup>st</sup> in Deep Learning course during Ph.D. at MBZUAI.
- First place in the Innovation Competition 2017, hosted by the Egyptian Ministry of Communications.

## CERTIFICATES

---

- |   |                 |
|---|-----------------|
| • TensorFlow in Practice specialization from deeplearning.ai        | <i>Coursera</i> |
| • Deep Learning specialization from deeplearning.ai                 | <i>Coursera</i> |
| • Machine Learning specialization from the University of Washington | <i>Coursera</i> |

## LANGUAGES

---

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

## REFERENCES

---

[Prof. Ming-Hsuan Yang](#)

External PhD Supervisor

Google DeepMind; University of California at Merced

✉ [mhyang@ucmerced.edu](mailto:mhyang@ucmerced.edu)

[Prof. Fahad Khan](#)

PhD Supervisor

Professor, MBZUAI

✉ [fahad.khan@mbzuai.ac.ae](mailto:fahad.khan@mbzuai.ac.ae)

[Dr. Hesham Eraqi](#)

Former Supervisor

Senior Scientist (L6), Amazon, Seattle, USA

✉ [heraqi@amazon.com](mailto:heraqi@amazon.com)