HANOONA ABDUL RASHEED

hanoona.bangalath@mbzuai.ac.ae, +971-50-5427243

Website: hanoonarasheed.com, Github Profile: hanoonaR, LinkedIn: Hanoona Rasheed

PERSONAL PROFILE

A Computer Vision engineer adept in research and development of deep learning driven computer vision applications. Currently working on research exploring the potentials of multimodal understanding from vision and language to build scalable general purpose vision systems that continually learn and can generalize to various domains and tasks.

RESEARCH PROJECTS

GLaMM: Pixel Grounding Large Multimodal Model ***Nov 2023***

Hanoona Rasheed, Muhammad Maaz, Sahal Shaji, Abdelrahman Shaker, Salman Khan, Hisham Cholakkal, Rao M. Anwer, Eric Xing, Ming-Hsuan Yang, Fahad S. 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 visual inputs provided by the user at multiple granularity levels (objects, object parts, attributes, relationships and holistic scene understanding).

Bridging the Gap between Object and Image-level Representations for Open-Vocabulary Detection ***May 2022 (NIPS 2022)***

Hanoona Rasheed, Muhammad Maaz, Muhammad Uzair Khattak, Salman Khan, Fahad Khan

The work solves the Open-vocabulary detection (OVD) problem using pretrained CLIP model, adapting it for object-centric local regions using region-based distillation and image-level weak supervision. The work address this problem by performing object-centric alignment of the language embeddings from the CLIP model. The proposed model seeks to minimize the gap between object and image-centric representations in the OVD setting.

Video-ChatGPT: Towards Detailed Video Understanding via Large Vision and Language Models ***Jun 2023 (Under Review)***

Muhammad Maaz, Hanoona Rasheed, Muhammad Uzair Khattak, Salman Khan, Fahad Khan

This work addresses the underexplored field of video-based conversation by introducing Video-ChatGPT. It is a multimodal model that merges a video-adapted visual encoder with a LLM. The model is capable of understanding and generating human-like conversations about videos. We introduce a new dataset of 100,000 video-instruction pairs used to train Video-ChatGPT acquired via manual and semi-automated pipeline that is easily scalable and robust to label noise. We also develop a quantiative evaluation framework for video-based dialogue models to objectively analyse the strengths and weaknesses of proposed models.

Fine-tuned CLIP models are efficient video learners ***Nov 2022 (CVPR 2023)***

Hanoona Rasheed, Muhammad Uzair Khattak, Muhammad Maaz, Salman Khan, Fahad Khan

In this work, we formulate and show the significance of an often neglected but simple baseline for transferring image-based CLIP model to video domain. ViFi-CLIP (Video Fine-tuned CLIP) shows that a simple fine-tuning of CLIP is sufficient to learn suitable video-specific

inductive biases, and can perform competitive to more complex approaches having dedicated components designed to model temporal information in videos. We introduce base-to-novel generalization benchmark for video-domain for evaluating the generalization ability of models for video action recognition.

Class-agnostic Object Detection with Multi-modal Transformer *Mar 2022 (ECCV 2022)*
Muhammad Maaz, Hanoona Rasheed, Salman Khan, Fahad Khan, Rao M. Anwer, Ming-Hsuan Yang
The work explores the potential of the recent Multi-modal Vision Transformers (MViTs) for class-agnostic object detection. Our extensive experiments across various domains and novel objects show the state-of-the-art performance of MViTs to localize generic objects in images. We also develop an efficient and flexible MViT architecture using multi-scale feature processing and deformable self-attention that can generate proposals given a specific language query.

MaPLe: Multi-modal Prompt Learning *Nov 2022 (CVPR 2023)*
Muhammad Uzair Khattak, Hanoona Rasheed, Muhammad Maaz, Salman Khan, Fahad Khan
The work proposes to learn prompts in both vision and language branches of pretrained CLIP for adapting it to different downstream tasks. Previous works only use prompting in either language or vision branch. We note that using prompting to adapt representations in a single branch of CLIP (language or vision) is sub-optimal since it does not allow the flexibility to dynamically adjust both representation spaces on a downstream task. To this end, we propose Multi-modal Prompt Learning (MaPLe) for both vision and language branches to improve alignment between the vision and language representations.

SwiftFormer: Efficient Additive Attention for Transformer-based Real-time Mobile Vision Applications *Mar 2023 (ICCV 2023)*
Abdelrahman Shaker, Muhammad Maaz, Hanoona Rasheed, Salman Khan, Ming-Hsuan Yang, Fahad Shahbaz Khan
This work introduces a novel efficient additive attention mechanism that effectively replaces the quadratic matrix multiplication operations 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.

UNETR++: Delving into Efficient and Accurate 3D Medical Image Segmentation *May 2023 (Under review)*
Abdelrahman Shaker, Muhammad Maaz, Hanoona Rasheed, Salman Khan, Ming-Hsuan Yang, Fahad Shahbaz Khan
The work proposes a 3D medical image segmentation approach, named UNETR++, that offers both high-quality segmentation masks as well as efficiency in terms of parameters and compute cost. Our extensive evaluations on three benchmarks, Synapse, BTCV and ACDC, reveal the effectiveness of the proposed contributions in terms of both efficiency and accuracy.

EDUCATION

Mohamed bin Zayed University of Artificial Intelligence, UAE *Jan 2023 - Continue*
Ph.D. in Computer Vision

Mohamed bin Zayed University of Artificial Intelligence, UAE *Dec 2020 - Dec 2022*
Research Based Masters in Computer Vision
CGPA: 4.0/4.0

APJ Abdul Kalam Technological University, India *Jun 2016 - Sept 2018*
M.Tech Signal Processing
CGPA: 9.18/10.0 **(First class with honors, Master thesis patented)**

University of Calicut, India *Mar 2012 - Apr 2016*
B.Tech Biomedical Engineering
CGPA: 8.49/10.0 **(First class with honors)**

WORK EXPERIENCE

Unique World Robotics, UAE *Oct 2019 - Apr 2020*
Software Engineer

Led a team in developing in-depth, value addition courses and curriculum on artificial intelligence and software robotics. Acted as consultant and trainer to corporations on AI in robotic process automation. Was part of research and development team in deep learning in robotic development. Key projects included vision for Alton robot and AI integrated RPA for document analysis in banking.

Robert Bosch Engineering and Business Solutions, India *Jan 2018 - Jul 2019*
Signal Processing Engineer
Worked in the Centre of Excellence for chemometrics and machine learning, innovations and incubation department. During my tenure, I harnessed advanced data preprocessing and visualization tools, while navigating the capabilities and limitations of different learning algorithms. I focused on evaluating spectral data - NIR, MIR, and Hyperspectral imaging - and advised on the appropriate machine learning tools for chemometric analysis. Two major projects that highlight my contributions include bringing the Milk Adulterant Detector from its research phase to productization and serving as the Project Lead for the Software Modelling of an Infrared Spectroscopic Simulator.

Robert Bosch Engineering and Business Solutions, India *Sep 2017 - Sep 2018*
Research Intern

Underwent intensive training conducted by senior engineers in team with assistance from a team of 'Top 85' scientists from Bosch Germany. Completed master thesis "Near Infrared Spectroscopy for Composition Analysis of Milk" and concept patented under Bosch. Developed machine learning and chemometric models and tools for fat, lactose, and protein concentration detection in milk.

WORK PROJECTS

- **Self Supervised Learning using Jigsaw Augmentation for Fine-grained classification**
- **NIR Spectroscopy for Composition Analysis of Bovine Milk (Bosch)**
- **Milk Adulterant Detector research phase 3 to the productization level (Bosch)**
- **Diabetic retinopathy classification using deep CNN and image enhancement**

TECHNICAL STRENGTHS

Computer Sciences	**Computer Vision, Deep Learning, Machine Learning**
Programming Languages	**Python, C**
ML and DL Frameworks	**PyTorch**
Softwares & Tools	**Pycharm, MATLAB**

ACHIEVEMENTS

- **Reviewer for CVPR 2023, ICCV 2023, NeurIPS 2023, ACCV 2023**
- **Reviewer for Vision Transformer Theory and Application (VTTA) Workshop, NeurIPS 2022**

- **Reviewer for Vision Transformer Theory and Application (VTTA) Workshop, ACCV 2022**
- **Reviewer for IEEE TPAMI: Special Issue, Transformer Models in Vision 2022**
- **Cat-vs-Dogs Kaggle challenge First Prize Winner MBZUAI (2020)**
- **Active Red Crescent Volunteer, Abu Dhabi, UAE (Since 2020)**
- **Calicut University Engineering Rank: 4 (2016)**

REFERENCES

Dr. Salman Khan
Academic Advisor
Associate Professor,
MBZUAI
✉ **salman.khan@mbzuai.ac.ae**

Prof. Fahad Khan
Academic Advisor
Deputy Department Chair,
Professor, MBZUAI
✉ **fahad.khan@mbzuai.ac.ae**

Prof. Timothy Baldwin
Acting Provost,
Associate Provost,
MBZUAI
✉ **timothy.baldwin@mbzuai.ac.ae**