

Mustansar Fiaz

Computer Vision • Vision-Language Models • Multimodal AI • Reinforcement Learning • AI Agents

Staff Research Scientist, IBM Research | Abu Dhabi, UAE

mustansar.fiaz@gmail.com | mustansar.fiaz@ibm.com | mustansarfiaz.github.io | LinkedIn | GitHub | Google Scholar

Profile Summary

Staff Research Scientist at IBM Research with a PhD in computer science and 5+ years of post-PhD research in **computer vision, vision-language models (VLMs), and multimodal AI**. My work spans **post-training of foundation models** (SFT, LoRA, reinforcement learning with GRPO), **self-evolving models, AI agents and tool use, efficient inference**, and core vision tasks including object detection, segmentation, tracking, and change detection, with applications in **Earth observation, climate, and medical imaging**. Published at **ICML, ECCV, CVPR, WACV, and BMVC**, and in ACM Computing Surveys, IEEE TGRS, and Pattern Recognition. Winner, **IBM watsonx Challenge 2025** (GeoAIAgent); Finalist, **UAE AI Award 2025** (EarthDial). Contributed to research proposals securing USD 900K in funding and co-supervised 3 PhD and 4 master's students.

Experience

IBM Research

Oct. 2023 – Present

Staff Research Scientist | Abu Dhabi, UAE

- **Reinforcement fine-tuning of VLMs (GeoVLM-R1)**: Took the project from problem definition to publication, designing a two-stage pipeline of supervised fine-tuning followed by GRPO-based reinforcement learning with format-compliance and task-aware rewards, improving structured visual reasoning across **28 downstream datasets**.
- **Self-evolving VLMs (TerraEvo)**: Developed a LoRA-based framework that adapts VLMs using unlabeled imagery, optimizing a Solver and a Proposer through reinforcement learning (GRPO and REINFORCE, respectively). Designed label-free rewards with a tool-free judge, skill-diversity controls, and adaptive KL regularization to maintain skill coverage and reduce catastrophic forgetting.
- **Multi-sensor foundation model (EarthDial)**: Co-developed a conversational vision-language foundation model and an instruction-tuning dataset of **11.11M instruction pairs** across optical, SAR, multispectral, and temporal imagery. Evaluated performance and generalization across **44 downstream datasets** through supervised and zero-shot evaluation. Recognized as a **Finalist in the Scientific Research category at the UAE AI Award 2025**.
- **Parallel Instance Decoding (PID)**: Co-developed a generative localization framework for VLMs that separates instance identification from spatial localization, then decodes all bounding boxes in a single parallel pass using an instance-isolated attention mask. Improved F1@0.50 by **8.3/7.3 points** on DIOR/DOTAv2 while cutting response latency by **22.4%/43.6%** over autoregressive decoding.
- **Agentic geospatial AI (GeoAIAgent)**: Member of the team that built GeoAIAgent, a multi-agent system that lets users ask natural-language questions about geospatial data (e.g., detecting methane plumes or mapping floods). A central orchestrator coordinates five specialized agents that search data catalogs, connect to external sources such as NASA and ESA, and run IBM, NASA, and ESA geospatial foundation models. **Winner of the Research category, IBM watsonx Challenge 2025**, among more than **150,000 IBMers**.
- **Agentic AI (OpenEarthAgent)**: Co-developed a tool-augmented agent framework for multi-step reasoning, planning, and tool use, trained with SFT on tool trajectories (**14,538 training / 1,169 evaluation instances**). Standardized tool calls through a unified JSON interface and validated trajectories through deterministic replay.
- **Urban heat islands (UHI)**: Worked on a joint IBM-MBZUAI project to identify and mitigate urban heat islands in the UAE, using remote sensing data and AI to support sustainable solutions against the risks of global warming in the region.

Mohamed bin Zayed University of Artificial Intelligence

Aug. 2021 – Oct. 2023

Research Associate | Abu Dhabi, UAE

- Carried out research on **remote sensing change detection, person search, semantic segmentation, domain adaptation, and medical image segmentation**, published at WACV, BMVC, IGARSS, and IEEE TGRS, and filed a U.S. patent application.
- Wrote research proposals: the MBZUAI-WIS Joint Program for AI Research proposal (with the guidance of my postdoctoral supervisor), awarded **USD 900K**, and a proposal on large language models for climate change, awarded a **Google Research Gift Award (2023)**.
- Co-supervised **3 PhD and 4 master's students**; designed and taught the **CV701** computer vision laboratories and served as teaching assistant for the course.

Tricubics

Mar. 2021 – Jul. 2021

Senior AI Software Engineer / AI Team Lead | Seoul, South Korea

- Led the AI team in **computer vision model development** for unmanned retail stores, translating application requirements into model design and technical priorities.

Education

PhD, Computer Science and Engineering

Mar. 2016 – Feb. 2021

Kyungpook National University, South Korea

Thesis: Robust Object Tracking and Segmentation Using Siamese Networks. *KINGS KNU Scholarship*.

Master of Engineering

Mar. 2014 – Feb. 2016

Sejong University, South Korea

Sejong International Scholarship.

BS, Computer and Information Sciences

Mar. 2007 – Aug. 2011

Pakistan Institute of Engineering and Applied Sciences (PIEAS), Pakistan

BS Fellowship.

Research Funding

MBZUAI–WIS Joint Program for AI Research, Project WIS P008

USD 900K

Wrote the proposal with the guidance of my postdoctoral supervisor. The project developed deep-learning methods for cell segmentation and tracking in live embryo imaging, leading to SA2-Net (BMVC 2023, Oral) and GA2-Net (Pattern Recognition 2024) publications.

Teaching Experience

- **CV701 Laboratory Lead, MBZUAI** (graduate): designed the lab material, taught the lab sessions, and created and marked lab assignments in Python and PyTorch.
- **Teaching Assistant, CV701, MBZUAI** (graduate): ran help sessions, answered student questions, and supported marking.
- **Guest Lectures** (while at IBM Research): delivered lectures to graduate students at MBZUAI and undergraduate students at United Arab Emirates University (UAEU).

Supervision

Co-supervised **3 PhD and 4 master's students** at MBZUAI, guiding them in choosing research problems, designing experiments, and publishing their work.

Technical Skills

Research Areas: Computer vision; vision-language models; multimodal learning; generative AI; reinforcement learning; AI agents; efficient deep learning; video understanding; remote sensing and Earth observation; medical image analysis.

Methods: Supervised fine-tuning (SFT); LoRA/PEFT; DPO; GRPO and REINFORCE; reward design; knowledge distillation; instruction tuning; domain adaptation; transformers; CNNs; state space models; object detection, segmentation, tracking, and change detection.

Tools: Python; PyTorch; Hugging Face Transformers; DeepSpeed; vLLM; distributed multi-GPU training; IBM watsonx Orchestrate; R; Git; Weights & Biases; TensorBoard.

Leadership: Research planning; proposal writing; mentoring and co-supervising PhD and master's students; teaching; cross-team and international collaboration.

Professional Service

Conference Reviewer: ICLR (2027), CVPR (2026), ECCV (2026), NeurIPS (2026), WACV (2025), BMVC (2025).

Journal Reviewer: IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI); IEEE Transactions on Geoscience and Remote Sensing; IEEE Transactions on Multimedia; Pattern Recognition; Knowledge-Based Systems.

Volunteer: vision-transformer workshops at ICCV 2023, NeurIPS 2022, and ACCV 2022; IW-FCV 2021.

Awards & Recognition

Winner, Research category, IBM watsonx Challenge 2025 (GeoAI Agent; 150,000+ IBMers participated); **Finalist, Scientific Research category**, UAE AI Award 2025 (EarthDial); **Best Paper & Best Presentation**, IW-FCV 2022; **Outstanding Research CSE Thesis**, 2021; **Best Student Paper**, IW-FCV 2020. Oral presentations: BMVC 2023, IGARSS 2024/2025, WACV 2025. Scholarships: KINGS KNU Scholarship (PhD), Sejong International Scholarship (MS), BS Fellowship (BS).

Selected Papers

- **Mustansar Fiaz**, H. Debary, P. Fraccaro, et al. *TerraEvo: Towards Reliable Self-Evolving for Earth Observation Vision-Language Models*. ICLR 2027 (submitted).
- S. Patle, M. Z. Zaheer, **Mustansar Fiaz**, H. A. Rasheed, F. Khan, and S. Khan. *Parallel Instance Decoding for Language-Guided Localization in Aerial Images*. ICLR 2027 (submitted).
- **Mustansar Fiaz**, H. Debary, P. Fraccaro, D. Paudel, L. V. Gool, F. Khan, and S. Khan. *GeoVLM-R1: Reinforcement Fine-Tuning for Improved Remote Sensing Reasoning*. arXiv:2509.25026, 2025. [Project Page] [pdf]
- A. Shabbir, M. U. Sheikh, M. A. Munir, H. Debary, **Mustansar Fiaz**, et al. *OpenEarthAgent: A Unified Framework for Tool-Augmented Geospatial Agents*. ECCV 2026. [Project Page] [pdf]
- M. Javaid, M. Noman, A. Hannan, S. Nawaz, **Mustansar Fiaz**, and S. Ghuffar. *Towards Effective Waste Segmentation for Automated Waste Recycling*. ICML 2026. [Project Page] [pdf]
- D. N. Abdul Kareem, A. Hannan, M. Noman, J. Lahoud, **Mustansar Fiaz**, and H. Cholakal. *InceptionMamba: Efficient Multi-Stage Feature Enhancement with Selective State Space Model for 2D Medical Image Segmentation*. MICCAI 2026 Workshop (EMA). [OpenReview] [arXiv]
- L. P. Tizzei, N. Gabrielle, L. Klein, I. Khabibrakhmanov, M. Zortea, H. Debary, **Mustansar Fiaz**, et al. *GeoAI Agent: Agentic Workflow for Automated Geospatial Data Management*. International Workshop on Agentic Engineering, pp. 68–72, 2026. [pdf]
- S. Sagar, A. Dudhane, H. Debary, **Mustansar Fiaz**, et al. *EarthDial: Turning Multi-Sensory Earth Observations to Interactive Dialogues*. CVPR 2025, pp. 14303–14313. [Project Page] [pdf]
- **Mustansar Fiaz**, M. Noman, H. Debary, K. Ali, and H. Cholakal. *HyRet-Change: A Hybrid Retentive Network for Remote Sensing Change Detection*. IEEE IGARSS 2025 – Oral, pp. 7209–7213. [Project Page] [pdf]
- M. Ali, M. Javaid, M. Noman, **Mustansar Fiaz**, and S. Khan. *COSNet: A Novel Semantic Segmentation Network using Enhanced Boundaries in Cluttered Scenes*. IEEE/CVF WACV 2025 – Oral, pp. 1363–1372. [Project Page] [pdf]
- M. Noman*, **Mustansar Fiaz***, H. Cholakal, S. Narayan, R. M. Anwer, S. Khan, and F. S. Khan. *Remote Sensing Change Detection With Transformers Trained from Scratch*. IEEE Transactions on Geoscience and Remote Sensing, 2024. [Project Page] [pdf]
- A. M. Khaleed*, **Mustansar Fiaz***, and H. Cholakal. *DDAM-PS: Diligent Domain Adaptive Mixer for Person Search*. IEEE/CVF WACV 2024, pp. 6674–6683. [Project Page] [pdf]
- M. Noman, **Mustansar Fiaz**, and H. Cholakal. *ChangeBind: A Hybrid Change Encoder for Remote Sensing Change Detection*. IEEE IGARSS 2024 – Oral, pp. 8417–8422. [Project Page] [pdf]
- **Mustansar Fiaz**, et al. *Guided-Attention and Gated-Aggregation Network for Medical Image Segmentation*. Pattern Recognition, 2024. [Project Page] [pdf]
- **Mustansar Fiaz**, M. Heidari, R. M. Anwer, and H. Cholakal. *SA2-Net: Scale-aware Attention Network for Microscopic Image Segmentation*. BMVC 2023 – Oral. [Project Page] [pdf]
- **Mustansar Fiaz**, H. Cholakal, R. M. Anwer, and F. S. Khan. *SAT: Scale-Augmented Transformer for Person Search*. IEEE/CVF WACV 2023, pp. 4809–4818. [Project Page] [pdf]
- **Mustansar Fiaz**, A. Mahmood, S. Javed, and S. K. Jung. *Handcrafted and Deep Trackers: Recent Visual Object Tracking Approaches and Trends*. ACM Computing Surveys, 2019.

Full publication list: Google Scholar.

Book Chapter

M. Fiaz, A. Mahmood, and S. K. Jung. *Deep Siamese Networks toward Robust Visual Tracking*. In Visual Object Tracking with Deep Neural Networks, IntechOpen. DOI: 10.5772/intechopen.86235.

Patent Application

M. Fiaz, H. Cholakal, S. Narayan, R. M. Anwer, and F. S. Khan. *System and Method for Attention-Aware Relation Mixer for Person Search*. U.S. patent application 17/983,741; published May 9, 2024.