

# Amir Mazaheri

COMPUTER VISION RESEARCH SCIENTIST • PHD

📞 (321) 240-3601 | ✉ amirmazaheri1990@gmail.com | 🌐 amirmazaheri1990 | in amirmazaheri1990 | Berkeley, CA

## Summary

Dr. Amir Mazaheri is a Computer Vision scientist with deep expertise in large-scale video understanding, Vision-Language Models (VLMs), and multimodal AI systems. He has pioneered state-of-the-art solutions across video temporal reasoning, content moderation, in-video search, and multimodal representation learning. Dr. Mazaheri holds a PhD from UCF's CRCV, has authored top-tier publications at **CVPR**, **ICCV**, **ECCV**, **EMNLP**, and **AAAI**, and holds multiple US patents. He brings both foundational research depth and production-scale delivery for real-world AI systems.

## Experience

### Warner Bros. Discovery (HBO)

San Francisco, CA

STAFF MACHINE LEARNING ENGINEER – COMPUTER VISION

July 2025 – Present

Leading large-scale video understanding and content moderation systems, including automated harmful content detection and multi-modal safety classifiers for HBO's streaming platform. Also developing temporal segmentation, scene/shot boundary detection, VLM-powered video indexing, and LLM-enhanced metadata generation for fine-grained video search and discovery across Warner Bros.' catalog. Leveraging state-of-the-art Vision-Language Models and video transformers for both content understanding and safety enforcement at scale.

### Tubi

San Francisco, CA

SENIOR MACHINE LEARNING ENGINEER - COMPUTER VISION

September 2021 - July 2025

Developing AI-driven content analysis and personalization solutions using Vision-Language Models, including brand identification and logo detection, scene segmentation, personalized trailers and content artwork generation through user preference modeling, and LLM-powered recommendation enhancements.

### Aibee U.S. Corporation

Palo Alto, CA

ALGORITHM SCIENTIST

July 2020 - September 2021

Designed and implemented CV models for human tracking in the crowd, person re-identification, and AI-based benchmarking tools for tracking.

### Netflix

Los Gatos, CA

RESEARCH SCIENTIST INTERN

May 2018 - August 2018

Developing a semi-supervised activity detection system for identifying events in movies without human annotations and training an attentive deep neural network to optimize movie billboard design based on visual patterns.

### Nielsen

Orlando, FL

RESEARCH FELLOWSHIP

June 2017 - Aug. 2018

Developed automated methods for TV advertisement analysis by focusing on classification, tagging, and descriptive analysis of broadcast ads.

### Center for Research in Computer Vision (CRCV)

Orlando, FL

GRADUATE RESEARCH ASSISTANT

August 2013 - April 2020

CRCV is a leading research group in Computer Vision and Machine Learning to promote foundational AI research [Link].

## Education

### Ph.D. in Computer Science

FL, USA

UNIVERSITY OF CENTRAL FLORIDA

April 2020

- **PhD Advisor:** Prof. Mubarak Shah, **Center for Research in Computer Vision (CRCV)**
- **Dissertation Title:** Video Content Understanding Using Text.

### M.Sc. in Computer Science

FL, USA

UNIVERSITY OF CENTRAL FLORIDA

May 2016

### BS in Computer Science

Tehran, Iran

SHARIF UNIVERSITY OF TECHNOLOGY

July 2013

## Skills

---

<b>Programming Languages</b>	Proficient in Python, C/C++, SQL, and MATLAB; familiar with SCALA and Java.
<b>Machine Learning Frameworks</b>	Experienced in developing and deploying models using TensorFlow, Keras, and PyTorch.
<b>Vision-Language Models (VLMs) and LLMs</b>	Expert in state-of-the-art VLMs and LLMs; experienced in instruction tuning, and multimodal fine-tuning for video/content understanding tasks.
<b>Cloud Computing &amp; Big Data Tools</b>	Skilled in cloud services (AWS), big data processing (Apache Spark, PySpark).
<b>Containerization &amp; DevOps</b>	Proficient in Docker for containerization; experienced with CI/CD pipelines for streamlined development.

## Patents

---

- **Machine learning techniques for advanced frequency management** - (U.S. Patent - US20240314371A1). [Link]
- **Multimedia scene break detection** - (U.S. Patent - US20240357217A1). [Link]

## Publications

---

- **WoundNet: A Domain-Adaptable Few-Shot Classification Framework for Wound Healing Assessment** - (ISBI' 2023). [Link]  
WoundNet introduces a few-shot framework for assessing wound healing to distinguish “healer” from “non-healer” wounds through temporal image analysis. Experimental results validate the framework’s effectiveness in improving early intervention in wound care and demonstrate up to 92% accuracy. Through this work, I **co-advised** two MS students from UC Santa Cruz.
- **Context-aware analysis of group submissions for group anomaly detection and performance prediction**. - (AAAI' 2023). [Link]  
This research integrates student-generated materials and applies a context-aware deep learning framework to monitor students’ performance in active learning exercises in large-scale courses.
- **Video Generation from Text Employing Latent Path Construction for Temporal Modeling** - (ICPR' 2022). [Link]  
This paper introduces a pioneering approach to **text-to-video generation**, the first to utilize realistic datasets such as A2D and UCF101. It regresses latent representations of initial and final frames, employs context-aware interpolation for intermediate frames, and addresses the challenge of visualizing natural language descriptions as a coherent video.
- **MMFT-BERT: Multimodal Fusion Transformer with BERT Encodings for Visual Question Answering** - (Findings in EMNLP' 2020). [Link]  
We introduced **MMFT-BERT** (Multi-Modal Fusion Transformer with BERT) for Visual Question Answering, utilizing separate and combined multi-modal (video and text) processing with BERT encodings and a novel fusion method. We achieved SOTA results on the TVQA dataset and created TVQA-Visual, a diagnostic subset to analyze the model’s handling of both modalities.
- **Deep Photo Cropper and Enhancement** - (ICIP' 2020). [Link]  
We developed an innovative image enhancement solution by utilizing dual deep networks for precise cropping and quality enhancement of embedded images using spatial transformers and super-resolution techniques. We also introduced and utilized a novel dataset through this task to demonstrate qualitative and quantitative improvements in image processing.
- **Pay attention! - Robustifying a Deep Visuomotor Policy through Task-Focused Attention** - (CVPR' 2019). [Link]  
This project enhances robotic manipulation with Task-Focused visual Attention (TFA) to improve task execution and recovery from disturbances using natural language instructions. The approach increases resilience against visual and physical disruptions. Our framework mimics human-like selective attention, which leads to increased robustness and adaptability in robots.
- **Visual Text Correction** - (ECCV' 2018). [Link]  
This paper introduces *Visual Text Correction (VTC)*, a novel task focused on correcting inaccuracies in video descriptions. The method detects errors and suggests accurate replacements by understanding video-text semantics and word relations. Through this work, we demonstrate the ability to automate text correction for video content.
- **Video Fill In the Blank using LR/RL LSTMs with Spatial-Temporal Attentions** - (ICCV' 2017). [Link]  
This paper tackles the Video-Fill-In-the-Blank (VFIB) challenge by proposing a framework that uses dual LSTMs for textual encoding of sentence fragments, integrated with an *external memory*, and employs spatial and temporal attention models for visual encoding. The approach effectively selects discriminative visual features to accurately predict missing words in video descriptions.
- **Learning a multi-concept video retrieval model with multiple latent variables** - (ACM-TOMM' 2018). [Link]  
This work aims to provide a foundational model for computing the ranking scores of a video in response to one or more concepts, where the concepts may be supplied directly by users or inferred by the system from user queries.