Chaoyu Li

Tempe, AZ | 917-327-7930 | chaoyuli@asu.edu | LinkedIn | GitHub | Website

EDUCATION

Arizona State University   **Aug 2023 - May 2028 (Expected)**
Doctor of Science, Computer Science, **GPA: 4.0/4.0**   Tempe, AZ, U.S.

University of Southern California   **Aug 2021 - May 2023**
Master of Science, Applied Data Science, **GPA: 4.0/4.0**   Los Angeles, CA, U.S.

Chongqing University of Posts and Telecommunication   **Aug 2016 - May 2020**
Bachelor of Science in Internet of Things, **GPA: 3.5/4.0 (Rank 3/141)**   Chongqing, China

PUBLICATIONS (*Equal Contribution)

[1] Chaoyu Li, Eun Woo Im, Pooyan Fazli. **VidHalluc: Evaluating Temporal Hallucinations in Multimodal Large Language Models for Video Understanding.** (Under submission to CVPR, Project Link)

[2] Chaoyu Li, Sid Padmanabhuni, Maryam S Cheema, Hasti Seifi, Pooyan Fazli. **VideoA11y: Method and Dataset for Accessible Video Description.** (Under submission to CHI)

[3] Xu Chen*, Chenqiang Gao*, Chaoyu Li, Yi Yang, Deyu Meng. **Infrared Action Detection in the Dark via Cross-Stream Attention Mechanism.** IEEE Transactions on Multimedia (TMM), vol. 24, pp. 288-300, 2022, doi: 10.1109/TMM.2021.3050069.

[4] Fengshun Zhou*, Chenqiang Gao*, Fang Chen, Chaoyu Li, Xindou Li, Feng Yang, Yue Zhao. **Face Anti-Spoofing Based on Multi-layer Domain Adaptation.** 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2019, pp. 192-197, doi: 10.1109/ICMEW.2019.00-88.

RESEARCH EXPERIENCE

People and Robots Lab | Arizona State University   **Aug 2023 - Present**
Research Assistant (Advisor: Prof. Pooyan Fazli)   AZ, U.S.

- Designed and implemented VidHalluc, the largest benchmark to date for **multimodal large language models (MLLMs)** hallucination evaluation on **video understanding** tasks, containing over **5,000** videos and **9,000** questions.
- Developed DINO-HEAL, a novel training-free algorithm leveraging saliency information from DINOv2 to re-weight visual features during inference, achieving a **3.02%** average boost across five models in **mitigating hallucinations**.
- Created VideoA11y, a novel MLLM-based method for **generating video descriptions** tailored to blind and low vision (BLV) individuals with 42 curated audio description (AD) guidelines.
- Released VideoA11y-40K, the first large-scale video description models training **dataset** with **40,000** videos for BLV users and introduced a new **video accessibility benchmark** based on the dataset.

IRIS Computer Vision Lab | University of Southern California   **May 2022 - May 2023**
Research Assistant (Advisor: Prof. Ram Nevatia)   CA, U.S.

- Innovated an approach formulating patch defense problems as **image segmentation** tasks and applied a task-agnostic defense, accomplishing SOTA robust accuracy without any degradation in benign performance.
- Fine-tuned the architecture to detection tasks on CARLA dataset (under DapricotPatch attack and Adversarial-Patch attack), increased adversarial pixel segmentation recall to **99.46%**, and boosted (avg) mAP by **15%** across tasks.

TRECVID 2019 ActEV, held by NIST | Computer Vision Lab of CQUPT   **Dec 2018 - May 2021**
Research Assistant (Advisor: Chenqiang Gao)   Chongqing, China

- Created an end-to-end pipeline to recognize action categories of people and vehicles in traffic surveillance videos on VIRAT dataset and drove team collaboration, ranking at the **top 20%** (Paper, Team name: NTT_CQUPT).
- Fine-tuned Mask R-CNN on VIRAT dataset to lift small **objects recognition** recall to **81%** and further applied Multiple-scale Sliding Window to get proposals from original labeled videos, achieving a recall of **83.5%**.
- Utilized the Non-Maximum Suppression Algorithm to merge proposals from the classifier that overlap in the time domain, increasing accuracy by **18%**.

Social Security Big Data Genetic Infrastructure Service Technology Research **July 2018 - Aug 2018**

Research Assistant (Advisor: Prof. Rui Feng) Shanghai, China

- Built an OCR model with CTPN+CRNN to **recognize information** on ID Cards at **60%** accuracy.
- Optimized the positioning part in CTPN according to the Chinese characters size, enhancing the recognition rate by **30%**.
- Fine-tuned CRNN on Synthetic Chinese String dataset, further lifting the Chinese recognition rate by **11%**.

PROFESSIONAL EXPERIENCE

Microsoft Asia-Pacific Research and Development **July 2019 - Aug 2019**

Machine Learning Research Assistant (Microsoft Bing)

- Optimized an end-to-end Tech News **search engine**, which automatically retrieves, tags and ranks news daily from 20+ technology websites and blogs (e.g., TechCrunch.com, Wired.com), and auto-pushes news to targeted customers.
- Built a **recognition model** with MTCNN+RESNET18 and DREAM module to identify celebrities in news images, attaining a recognition rate of **69%**, and delivered model architecture to cross-functional teams and leadership.
- Led collaboration with UI designers to establish a portal (HTML) connected to a MongoDB database in JavaScript.

AWARDS & HONORS

- Silver Award, 2016 ACM-ICPC Asia EC-FINAL Contest
- Silver Award, 2018 ACM-ICPC Asia, Qingdao Division
- National Scholarship 2017-2018 (top 0.2%)

TECHNICAL SKILLS

Programming Languages: C, Python, C++, JavaScript, SQL, HTML
Machine Learning: PyTorch, TensorFlow, Transformers, Keras, Scikit-learn, Spark, Hadoop, Computer Vision, NLP
DevOps Tools: Git, Linux, Shell, AWS, CI/CD, Weight & Biases