

ZHUOWAN LI

<https://lizw14.github.io>

410-900-2393 ◊ lizhuowan14@gmail.com

RESEARCH INTERESTS

My research interests lie in computer vision and natural language processing, including multimodal models, vision-and-language, compositional reasoning, model robustness, etc. I have experience in pretrained LLMs/VLMs. I am always excited to explore topics like generative AI, embodied AI and foundation models.

EDUCATION

Johns Hopkins University 2018 - 2024

Ph.D. in Computer Science

Advisors: Alan Yuille, Benjamin Van Durme

Thesis: On the Diagnosis and Generalization of Compositional Visual Reasoning. 2024.

Tsinghua University 2014 - 2018

B.E. in Electronic Information Science and Technology

Second major: Journalism and Communication

RESEARCH EXPERIENCE

Google DeepMind, Mountain View, CA Feb 2024 - Present

Software Engineer

- Gemini post-training, with a focus on personalization.
- Retrieval-augmented generation, model self-improvement, etc.

Amazon AWS, Santa Clara, CA May 2023 - Nov 2023

Applied scientist intern. Mentors: Bhavan Jasani, Peng Tang, Shabnam Ghadar

- Document and chart visual question answering using LLMs, synthetic data and compositional reasoning (in CVPR 2024).

Facebook, Menlo Park, CA May 2021 - September 2021

Research intern. Mentors: Dhruv Mahajan

- Unsupervised domain adaptation with guidance of pretrained models.

Adobe Research, San Jose, CA May 2019 - Nov 2019

Research intern. Mentors: Quan Tran, Long Mai, Zhe Lin

- Propose new datasets and methods for a new task: context-aware group captioning (in CVPR 2020).

Sensetime, Beijing, China Oct 2017 - March 2018

Part-time research intern. Mentor: Shuai Yi

- Learn discriminative human representations by reconstructing multi-view images using GAN (in NeruIPS 2018).

Tsinghua University, Beijing, China Nov 2016 - May 2017

Research Assistant. Advisor: Prof. Shengjin Wang

- Improve person re-identification with fine-grained attributes and pose information.

PUBLICATIONS

Gemini 2.5: Pushing the Frontier with Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities

Google Gemini Team.

Technical Report, 2025, abs/2507.06261

Retrieval Augmented Generation or Long-Context LLMs? A Comprehensive Study and Hybrid Approach

Zhuowan Li, Cheng Li, Mingyang Zhang, Qiaozhu Mei, Michael Bendersky.

EMNLP 2024 (Industry Track)

ExoViP: Step-by-step Verification and Exploration with Exoskeleton Modules for Compositional Visual Reasoning

Yuxuan Wang, Alan Yuille, **Zhuowan Li***, Zilong Zheng*.

COLM 2024

Synthesize Step-by-Step: Tools, Templates and LLMs as Data Generators for Reasoning-Based Chart VQA.

Zhuowan Li*, Bhavan Jasani*, Peng Tang, Shabnam Ghadar.

CVPR 2024

Causal-CoG: A Causal-Effect Look at Context Generation for Boosting Multi-modal Language Models

Shitian Zhao, **Zhuowan Li**, Yadong Lu, Alan Yuille, Yan Wang.

*CVPR 2024 **Highlight***

Localization vs. Semantics: Visual Representations in Unimodal and Multimodal Models

Zhuowan Li, Cihang Xie, Benjamin Van Durme, Alan Yuille.

EACL 2024

3D-Aware Visual Question Answering about Parts, Poses and Occlusions.

Xingrui Wang, Wufei Ma, **Zhuowan Li**, Adam Kortylewski, Alan Yuille.

NeurIPS 2023

Super-CLEVR: A Virtual Benchmark to Diagnose Domain Robustness in Visual Reasoning.

Zhuowan Li, Xingrui Wang, Elias Stengel-Eskin, Adam Kortylewski, Wufei Ma, Benjamin Van Durme, Alan Yuille.

*CVPR 2023 **Highlight***

Visual Commonsense in Pretrained Unimodal and Multimodal Models.

Chenyu Zhang, Benjamin Van Durme, **Zhuowan Li***, Elias Stengel-Eskin*.

*NAACL 2022 **Oral***

SwapMix: Diagnosing and Regularizing the Over-Reliance on Visual Context in Visual Question Answering.

Vipul Gupta, **Zhuowan Li**, Adam Kortylewski, Chenyu Zhang, Yingwei Li, Alan Yuille.

CVPR 2022

Calibrating Concepts and Operations: Towards Symbolic Reasoning on Real Images.

Zhuowan Li, Elias Stengel-Eskin, Yixiao Zhang, Cihang Xie, Quan Tran, Benjamin Van Durme, Alan Yuille.

ICCV 2021

Context-Aware Group Captioning via Self-Attention and Contrastive Features.

Zhuowan Li, Quan Tran, Long Mai, Zhe Lin, Alan Yuille.

CVPR 2020.

FD-GAN: Pose-guided Feature Distilling GAN for Robust Person Re-identification.

Yixiao Ge*, **Zhuowan Li***, Haiyu Zhao, Guojun Yin, Shuai Yi, Xiaogang Wang, Hongsheng Li.

NeurIPS 2018

PATENTS

Contrastive Captioning for Image Groups. Patent no. US 20240037939A1, 2024/2/1

Quan Tran, Long Mai, Zhe Lin, **Zhuowan Li**.

TALKS

Towards Generalizable Visual Reasoning

- Google Research	Nov 2023
- Horizon Robotics	Oct 2023
- Tiktok	Oct 2023
- UC Santa Cruz	Sep 2023
- MIT: Computational Cognitive Science Group (Josh Tenenbaum's group)	May 2023

AWARDS

CVPR 2023 Doctoral Consortium

MENTORING

Vipul Gupta	visiting intern, now PhD at Penn State Univeristy
Chenyu Heidi Zhang	JHU undergrad, now master at Stanford University
Chenyu Zhang	JHU master, now PhD at University of Trento
Xingrui Wang	USC master, now PhD at JHU
Varun Iyer	JHU master
Shitian Zhao	visiting undergrad from ECNU
Yijiang Li	JHU master student
Yuxuan Wang	visiting master student from Peking University

TEACHING

EN.601.461/661. Computer Vision	Johns Hopkins University
Role: Teaching Assistant	Spring 2022
Instructor: Kapil Katyal	

SERVICE

Reviewer for CVPR, ICCV, ECCV, NeurIPS, ICML, ICLR.

SKILLS

Programming Languages	Python, MATLAB, C++, C, \LaTeX
Deep Learning Tools	Pytorch, Tensorflow, Torch, Caffe