Xiangxi Shi

+1 (360)-593-3228| shixia@oregonstate.edu| sxx1995.github.io

EDUCATION

Oregon State University, United States *Sept 2020-present*
Ph.D. in Computer Science
University of Science and Technology of China, Hefei, China *Sept 2013 - Jun 2017*
Bachelor of Engineering in Automation

WORK EXPERIENCE

Adobe *Seattle, WA*
Research Scientist Intern *Jan 2024 - Nov 2024*

- Key developer of **OIDA-QA**, a large-scale, multimodal benchmark for document-based QA.
    - Built scalable pipelines to process and extract metadata from **400K+** documents
    - Designed and implemented a **GPT-based question generator**, generating **3M+ high-quality** Q&A pairs
    - Fine-tuned **Large Language Model (LLM)** to enhance their performance specifically on document-based QA tasks.
- Co-developed ADOPD-INSTRUCT, a **large-scale, multimodal dataset** for document editing.
    - Constructed **181K multimodal dataset** to support advanced editing model development.
    - Built **annotation tools** and **led human curation processes** to ensure the dataset's labeling quality and consistency.

Baidu USA *Seattle, WA*
Research Scientist Intern *Jun 2022 - Sept 2022*

- Developed a mask-based image editing system requiring no training.
- Delivered high-quality visuals and strong **semantic alignment** (CLIP score 34.7).

Adobe *Seattle, WA*
Research Scientist Intern *Jun 2021 - Sep 2021*

- Developed and deployed a video search model now serving live traffic on **Adobe Stock**.
- Boosting zero-shot performance by 10%, improving search relevance for users.
- Proposed a **two-stage video localization framework** outperforms **SoTA** methods by **22.5%** in first recall rate.

ROSE Lab, Nanyang Technological University *Singapore*
Research Assistant *Aug 2017 - Sep 2020*

- Conducted research on **image/video captioning**; co-authored **5 top-tier** conference papers (**ICCV**, **ECCV**, **ACM-MM**).

PUBLICATIONS & RESEARCH

*Published in **top-tier AI conferences**, including **CVPR, ICCV, ECCV, ICLR, ACM-MM** and **WACV***

*Xuan Shen, Y. Wang, **Xiangxi Shi**, et. al. Efficient Reasoning with Hidden Thinking (Under review)*

- **Featured in the TLDR AI newsletter (Feb 2025)**, reaching over **650K readers**.
- Introduced a novel framework that encodes **Chain-of-Thought reasoning** into **latent representations**, effectively **optimizing** computational resource usage.
- Achieved improved reasoning efficiency and competitive **zero-shot** accuracy across multiple **Multimodal Large Language Model (MLLM)** benchmarks.

3D Visual Grounding without Human-Annotated Queries(Under Review,First author)

- Proposed a novel **3D visual grounding task** designed to reduce dependence on manually provided queries.
- Achieved a **6% performance improvement** across major benchmarks, including ScanRefer and Nr3D.

Xiangxi Shi,** Z. Wu, S. Lee, Viewpoint-Aware Visual Grounding in 3D Scenes**(CVPR 2024)

- Developed a viewpoint-adaptive method for precise **3D language-to-object grounding**.
- Outperformed **state-of-the-art (SoTA)** methods by over **2%**.

*J. Gu, **Xiangxi Shi**, et. al. ADoPD: A Large-Scale Document Page Decomposition Dataset**(ICLR 2024)***

- Introduced a comprehensive dataset comprising **120K documents** for **multi-task applications** in document analysis.
- Developed a **model-assisted** data collection pipeline, **reducing labeling costs by 70%.**
- Proposed a **data-driven method** for discovering document taxonomy using **GPT-4 and CLIP.**

Z.Wu, **Xiangxi Shi*(equal contribution)**, et. al. Learning Meta-class Memory for Few-shot Semantic Segmentation**(ICCV2021)***

- Implement an accurate segmentation application of untrained objects using limited same-category reference examples.
- **First** to propose **learnable meta-class embeddings** for few-shot semantic image segmentation.
- Surpassed **SoTA** performance by **1%** (1-shot) **and 1.5%** (5-shot) on the PASCAL-5i benchmark.

Xiangxi Shi**, X. Yang, et. al. Finding It at Another Side: A Viewpoint-Adapted Matching Encoder for Change Captioning**(ECCV2020)

- Developed a **captioning model** distinguishing real changes from viewpoint shifts
- Proposed a **reinforcement learning** method to effectively guide the model's attention toward regions with semantic changes.
- **Surpassed SoTA performance by 8 points (+23.5%)** in CIDEr

*Z. Yang, **Xiangxi Shi,** et. al. Hijacking Vision-and-Language Navigation Agents with Adversarial Environmental Attacks **(WACV2025)***

- **First proposed** an **adversarial attacks** in Vision-and-Language Navigation (VLN) that manipulate 3D objects mesh and build up a simulation platform to enable the differentiable mesh manipulation
- Introduce a **novel sequential attack task** to guide the attacked agent following the **predefined path** with sequential actions through a **manipulated 3D object**.

Xiangxi Shi**, S. Lee, Benchmarking Out-of-Distribution Detection in Visual Question Answering **(WACV 2024)

- Collected **300K+ data** to construct an **Out-of-Distribution Detection (OOD) dataset** for the Visual Question Answering (**VQA)** task.
- Designed and implemented **19 model-score configurations** to systematically evaluate OOD performance across models.
- Proposed a **generative approach** for detecting OOD samples by synthesizing relevant questions for given images.

Xiangxi Shi**, J. Cai, S. Joty, J. Gui. Watch It Twice: Video Captioning with a Refocused Video Encoder **(ACM-MM19)

- Introduced **a novel model** for generating video captions based on detected keyframes.
- Achieved a 6.4-point CIDEr **improvement over SoTA methods** on the MSVD benchmark.

X. Shen, **Xiangxi Shi***(equal contribution), et. al. OIDA-QA: A Multimodal Benchmark for Analyzing the Opioid Industry Document Archive*

work completed during internship at Adobe

- Built scalable pipelines to process and extract metadata from **400K+** documents
- Designed and implemented a **GPT-based question generator**, generating **3M+ high-quality** Q&A pairs
- Fine-tuned **Large Language Model (LLM)** to enhance their performance specifically on document-based QA tasks.

*W. Zhu, **Xiangxi Shi**, et. al. ADOPD-INSTRUCT: A Large-Scale Multimodal Dataset for Document Editing*

work completed during internship at Adobe

- Constructed **181K multimodal dataset** to support advanced editing model development.
- Built **annotation tools** and **led human curation processes** to ensure the dataset's labeling quality and consistency.