Xiaoqian Shen

xiaoqian.shen@kaust.edu.sa https://github.com/xiaoqian-shen
xiaoqian-shen.github.io Google Scholar LinkedIn

Research Interest

- **Vision-Language:** Multi-modal Comprehension [1]–[4] / Generation [5]
- **Generative Models:** Image Generation [5] / Video Generation [6]

Education

King Abdullah University of Science and Technology, Saudi Arabia. Jan. 2024 – present
Ph.D. Computer Science, GPA: 4.0/4.0
Supervised by Prof. Mohamed Elhoseiny.

King Abdullah University of Science and Technology, Saudi Arabia. Aug. 2022 – Dec. 2023
M.Sc. Computer Science. GPA: 3.75/4.0 (M.S./Ph.D. program)
Thesis title: *Efficient Learning Algorithms for Temporally Consistent Video Synthesis*

Jilin University, China. Aug. 2018 – Jun. 2022
B.S. Computer Science. GPA: 91.1/100

Experience

- **Research Intern**, NVIDIA. Jun. 2025 - Sep. 2025, Taiwan
  Dr. Ryo Hachiuma and Dr. Min-Hung Chen
  Multimodal LLM for long video understanding
- **Research Scientist Intern**, Meta. May. 2024 - Nov. 2024, United States
  Dr. Yunyang Xiong, XR Core AI, Burlingame
  Multimodal LLM for long video understanding
- **Visiting research student**, KAUST. Dec. 2021 - Mar. 2022, Saudi Arabia
  Prof. Mohamed Elhoseiny's group
  Leverage hierarchical constructive learning for large-scale zero-shot classification
- **Research assistant**, Tsinghua University. Sep. 2020 - Mar. 2021, China
  Prof. Yongfeng Huang's group
  Medical Relation Extraction for Chinese Medicine Instructions

Publications

1. **Xiaoqian Shen**, W. Zhang, J. Chen, and M. Elhoseiny, "Vgent: Graph-based retrieval reasoning augmented generation for long video understanding," **[NeurIPS 2025],** ***Spotlight (Top 3%)***.
2. **Xiaoqian Shen**, Y. Xiong, C. Zhao, *et al.*, "Longvu: Spatiotemporal adaptive compression for long video-language understanding," **[ICML 2025],** ***Work done at Meta***.
3. **Xiaoqian Shen**, M.-H. Chen, Y.-C. F. Wang, M. Elhoseiny, and R. Hachiuma, "Zoom-zero: Reinforced coarse-to-fine video understanding via temporal zoom-in," *arXiv*, ***Work done at Nvidia***.
4. D. Zhu*, J. Chen*, **Xiaoqian Shen**, X. Li, and M. Elhoseiny, "Minigpt-4: Enhancing vision-language understanding with advanced large language models," **[ICLR 2024],** ***3k+ cites, GitHub 25k+ stars***.
5. **Xiaoqian Shen** and M. Elhoseiny, "Storygpt-v: Large language models as consistent story visualizers," **[CVPR 2025]**.

6. **Xiaoqian Shen**, X. Li, and M. Elhoseiny, "Mostgan-v: Video generation with temporal motion styles," **[CVPR 2023]**.
7. K. Ataallah, **Xiaoqian Shen**, E. Abdelrahman, *et al.*, "Goldfish: Vision-language understanding of arbitrarily long videos," **[ECCV 2024]**.
8. K. Haydarov, **Xiaoqian Shen**, A. Madasu, *et al.*, "Affective visual dialog: A large-scale benchmark for emotional reasoning based on visually grounded conversations," **[ECCV 2024]**.
9. K. Haydarov, A. Muhamed, **Xiaoqian Shen**, *et al.*, "Adversarial text to continous image generation," **[CVPR 2024]**.
10. J. Chen, D. Zhu, **Xiaoqian Shen**, *et al.*, "Minigpt-v2: Large language model as a unified interface for vision-language multi-task learning," *arXiv*, 2023.
11. E. M. Bakr, **Xiaoqian Shen***, P. Sun*, F. F. Khan*, L. E. Li, and M. Elhoseiny, "Hrs-bench: Holistic, reliable and scalable benchmark for text-to-image models," **[ICCV 2023]**.
12. D. Zhu, J. Chen, K. Haydarov, **Xiaoqian Shen**, W. Zhang, and M. Elhoseiny, "Chatgpt asks, blip-2 answers: Automatic questioning towards enriched visual descriptions," **[TMLR]**.
13. J. Zhang, S. Zhang, **Xiaoqian Shen**, T. Lukasiewicz, and Z. Xu, "Multi-condos: Multimodal contrastive domain sharing generative adversarial networks for self-supervised medical image segmentation," *IEEE Transactions on Medical Imaging*, 2023.
14. K. Yi, **Xiaoqian Shen**, Y. Gou, and M. Elhoseiny, "Exploring hierarchical graph representation for large-scale zero-shot image classification," *European Conference on Computer Vision*, pp. 116–132, 2022, **[ECCV 2022]**.
15. T. Qi, S. Qiu, **Xiaoqian Shen**, *et al.*, "Kemre: Knowledge-enhanced medical relation extraction for chinese medicine instructions," *Journal of Biomedical Informatics*, vol. 120, p. 103 834, 2021.

Academic Services

- **Conference reviewer**, CVPR, ECCV, ICCV, ICLR, SIGGRAPH Asia, AAAI, NeurIPSW
- **Journal reviewer**, IJCV, CVIU
- **Teaching Assistant**, KAUST CS 283 Deep Generative Modeling

Skills

- **Languages**: Chinese, English (TOEFL 104/120, GRE 328/340).
- **Coding**: Python, C/C++, Java, HTML5, LaTeX.
- **Software**: Photoshop, Final Cut Pro.

Awards

- KAUST Graduate Scholarship. 2022 - present
- Outstanding Undergraduate Thesis Award. 2022
- Academic Scholarship. 2019 - 2021