Sheng Liu

Phone: (716) 907-3924
Email: sliu66@buffalo.edu
Website: https://shengliu23.github.io

EDUCATION

State University of New York at Buffalo, Buffalo, NY, USA 09/2018 - 06/2022
Ph.D. in Computer Science
• Advisor: Prof. Junsong Yuan
• Awarded Chair's Fellowship

Xi'an Jiaotong University, Xi'an, Shaanxi, China 08/2013 - 06/2017
B.E. in Automation
• Member of the Special Class for the Gifted Young

RESEARCH INTERESTS

Vision and Language [**P3, P6, P7, P8**], Image and Video Generation [**P1**], 3D Vision [**P2**], Neural Rendering [**P4, P5**], Natural Language Processing (NLP),

I tackle **2D** and **3D** vision problems that include: multi-modal pre-training [**P3**], image and video captioning [**P6, P7**], visual question answering [**P8**], image and video compositing [**P1**], text-to-image generation, structure-from-motion [**P2**], neural human radiance fields [**P4**], kinematic formula learning [**P5**].

I possess hands-on experience in developing models for various NLP problems, including machine translation and sentiment analysis. Moreover, I have solid knowledge of, and practical experience with, **large language model (LLM)**, such as aligning LLM with human preferences via direct policy optimization (DPO) and retrieval augmented generation (RAG).

WORK EXPERIENCES

Applied Scientist II @ Amazon Prime Video, Seattle, WA, USA 07/2022 - Present

1. Self-supervised Pre-training for Image and Video Harmoinzation [**CVPR'23**]
   • Proposed a label efficient self-supervised harmonization method, effectively reducing annotated data requirements by 50% without any drop in performance.
   • Our method achieved a 1.0 PSNR improvement on iHarmony4 dataset.
   • Partnered with VFX artists to seamlessly incorporate our method into their workflow, resulting in a 30% reduction of their compositing time.
   • Filed a patent as the primary inventor.
2. Real-time Virtual Product Placement
   • Served as co-leader and main contributor.
   • Proposed a marker-based solution that won A/B testing.
   • Successfully deployed our solution in production via collaboration with product managers, VFX artists and software engineers.
   • See how our solution worked in **real** Twitch streams [stream 1]. The "GARTH BROOKS" poster, a **virtual** object inserted using our solution, harmoniously blended in the environments.
   • Filed a patent as the primary inventor.

Research Assistant @ University at Buffalo, Buffalo, NY, USA 08/2020 - 05/2022

1. Multi-modal Content Understanding [**TPAMI'21, P8**]
   • Proposed Sibling Convolutional Encoder (SibNet), a novel video captioning model which was trained via multi-task learning.
   • SibNet achieved top performance on two widely used benchmarks, *i.e.*, MSR-VTT and MSVD, in 2020.

• Proposed Question-Dependent Prompt Generation (QDPG) in 2021. QDPG enabled us to formulate visual question answering as a fill-in-the-blank problem.
• Our novel formulation enabled vision-and-language pre-trained models to perform zero-shot and few-shot question answering.

2. Kinematic Formula Learning [**MM'22**]
• Proposed a novel framework which leveraged neural radiance fields (NeRF) to learn kinematic formulas from multi-view videos without supervision. We only assumed knowledge of camera parameters.
• Demonstrated that our framework effectively learned kinematic of explosion, large angle pendulum, free fall, and was readily applicable to animation tasks.

3. Neural Clothed Human Model [**VCIP'22**]
• Proposed an animatable neural clothed human model which leveraged NeRF to represent both 3D geometry and appearance of a clothed human.
• Curated a dataset and demonstrated the effectiveness of our proposed neural clothed human model on our dataset.

Applied Scientist Intern @ Amazon Prime Video, Seattle, WA, USA 07/2021 - 10/2021
Structure-from-Motion for Cinematic Contents [**CVPR'22**][demo video][Amazon Science blog]
- Identified that limited camera motion, a distinctive feature of cinematic contents, is the reason why existing Structure-from-Motion methods perform poorly on cinematic contents.
- Curated a dataset featuring cinematic contents with limited camera motion.
- Proposed Depth-Guided Structure-from-Motion (DepthSfM), efficiently addressing the unique challenge posed by the limited camera motion.
- DepthSfM outperformed existing methods by more than 15.0% across three metrics.
- Filed a patent as the primary inventor.

Research Intern @ Microsoft Research, Seattle, WA, USA 05/2020 - 08/2020
Open-Vocabulary Visual Instance Search [**AAAI'22**][demo video]
- Introduced Open-Vocabulary Visual Instance Search (OVIS), *i.e.*, a novel task which aims to localize visual instances within a large image repository given an arbitrary textual query.
- Proposed a Visual-Semantic Aligned (ViSA) pre-training method, a vision-and-language pre-training method tailored for OVIS.
- Curated a dataset featuring over 1,600 textual queries paired with their corresponding visual instances for evaluation.
- ViSA demonstrated a 6.0% improvement in mAP over existing vision-and-language pre-trained models.

Project Officer @ Nanyang Technological University, Singapore 08/2017 - 07/2018
Multi-modal Content Understanding [**MM'18**]
- Proposed Sibling Convolutional Encoder (SibNet), a novel video captioning model which was trained via multi-task learning.
- SibNet achieved top performance on two widely used benchmarks, *i.e.*, MSR-VTT and MSVD, in 2018.

SELECTED PUBLICATIONS

[**P1**] LEMaRT: Label Efficient Masked Region Transform for Image Harmonization
Sheng Liu, Cong Phuoc Huynh, Cong Chen, Maxim Arap and Raffay Hamid
IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2023

[**P2**] Depth-Guided Sparse Structure-from-Motion for Movies and TV Shows
Sheng Liu, Xiaohan Nie and Raffay Hamid
IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022

[**P3**] OVIS: Open-Vocabulary Visual Instance Search via Visual-Semantic Aligned Pre-Training
Sheng Liu, Kevin Lin, Lijuan Wang, Junsong Yuan and Zicheng Liu
AAAI Conference on Artificial Intelligence (**AAAI**), 2022

[P4] NeCH: Neural Clothed Human Model
Sheng Liu*, Liangchen Song*, Yi Xu and Junsong Yuan
Visual Communications and Image Processing (**VCIP**), 2022
* indicates equal contribution

[P5] Learning Kinematic Formulas from Multiple View Videos
Liangchen Song*, **Sheng Liu***, Zhong Li, Yuqi Ding, Yi Xu and Junsong Yuan
ACM Conference on Multimedia (**ACM MM**), 2022
* indicates equal contribution

[P6] SibNet: Sibling Convolutional Encoder for Visual Captioning
Sheng Liu, Zhou Ren and Junsong Yuan
IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2021

[P7] SibNet: Sibling Convolutional Encoder for Video Captioning
Sheng Liu, Zhou Ren and Junsong Yuan
ACM Conference on Multimedia (**ACM MM**), 2018 (**Oral**)

[P8] Rethinking Visual Question Answering as Fill-in-the-Blank Question
Sheng Liu and Junsong Yuan
Tech report

HONORS AND AWARDS

Academic Awards

1. **Chair's Fellowship**, State University of New York at Buffalo 2018
2. **Special Prize**, Electronic Design Competition of Shaanxi Province (1/65) 2016
3. **Outstanding Student**, Xi'an Jiaotong University (Top 10%) 2015'16

Sports Awards

1. **Third Place**, College Student Volleyball Competition of Shaanxi Province 2015
2. **Second Place**, College Student Volleyball Competition of Shaanxi Province 2014

PROFESSIONAL SERVICES

Conference Reviewer

1. Conference on Neural Information Processing Systems (NeurIPS'22'23)
2. IEEE Conference on Computer Vision and Pattern Recognition (CVPR'19'20'21'22)
3. IEEE International Conference on Computer Vision (ICCV'19'21'23)
4. IEEE International Conference on Computer Vision (AAAI'20'21'22)

Journal Reviewer

1. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
2. IEEE Transactions on Image Processing (TIP)
3. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)