Xiangxi Shi

Mobile: +1 (360)-593-3228| Email: shixia@oregonstate.edu

EDUCATION

Oregon State University, United States *Sept.2020-present*
Ph.D. in Computer Science
Nanyang Technological University, Singapore *Apr.2020-Sep.2020*
Project Officer of Parallel and Distributed Computing Lab (PDCL)
Nanyang Technological University, Singapore *Aug.2017-Apr.2020*
Project Officer of Rapid-Rich Object Search (ROSE) Lab
University of Science and Technology of China, Hefei, China *Sept.2013-June.2017*
Bachelor of Engineering in Automation

RESEARCH INTERESTS

I am currently interested in the areas of Computer Vision and Natural Language Processing, including

- Visual Language Navigation
- Out-of-Distribution Detection
- 3D Visual-Language Learning

•

Work Experience & Internship

Baidu Inc. (Seattle) *Jun.2022-Sep.2022*
Internship Mainly focus on diffusion-based text to image generation
Adobe Inc. *Jun.2021-Sep.2021*
Internship Mainly focus on large-scale video representation learning
ROSE Lab, NTU *Aug.2017-Apr.2020*
Officer Mainly focus on vision to language generation

PAPERS & WORKSHOP

Xiangxi Shi, Stefan Lee, Benchmarking Out-of-Distribution Detection in Visual Question Answering
Accepted by WACV2024

- Introduce an Out-of-Distribution Detection to VQA task.
- Create a benchmark dataset with existing VQA datasets.
- Proposed a generation-based method and examine it with other existing OOD methods in our benchmark.

Xiangxi Shi, N. Xu, S. Lee Momentum-based Video-Text Model Pretraining for Moment Localization

- Propose a two-stage framework consisting of a post-tuned compatibility video retrieval model and a weight-lighted Score Refinement Network (SRN) trained separately on different datasets for moment localization adaptation.
- Post-tuned with the proposed Momentum-based transfer strategy, the video retrieval model achieves powerful video-text comprehension, leading to SoTA performance on the zero-shot video retrieval and moment localization tasks.

Xiangxi Shi, et al. Remember What You have drawn: Semantic Image Manipulation with Memory

- Proposed to disentangle the image features into texture and structure parts and introduce a set of latent memories to represent the texture information.
- Moreover, we further introduced a memory-level adversarial training loss to keep the memories robust and prominent.

Z.Wu, **Xiangxi Shi**, G. Lin, J. Cai. Learning Meta-class Memory for Few-shot Semantic Segmentation
Published in the IEEE/CVF International Conference on Computer Vision 2021

- First propose a set of learnable embedding to learning meta-class information for few-shot semantic image segmentation.
- For k-shot scenarios, a Quality Measurement Module (QMM) is proposed to measure the quality of all the support images to effectively fuse all the support features.
- Extensive experiments on PASCAL-5 and COCO datasets show that our proposed method performs the best in all settings.

Xiangxi Shi, X. Yang, J. Gu, S. Joty, and J. Cai. Finding It at Another Side: A Viewpoint-Adapted Matching Encoder for Change Captioning
Published by 16th European Conference on Computer Vision (ECCV2020)

- State-of-the-art among the current proposed change captioning methods.
- propose a novel image encoder that explicitly distinguishes semantic changes from the viewpoint changes by predicting the changed and unchanged regions in the feature space.
- propose a novel Reinforcement learning module that helps the model focus on the semantic change regions so as to generate better change captions.

Xiangxi Shi, J. Cai, S. Joty, J. Gui. Watch It Twice: Video Captioning with a Refocused Video Encoder
Published in the *27th ACM International Conference on Multimedia (ACMMM19)*

- Introduce a reinforcement learning based keyframe selection method to pick out the better key frame of a video to represent it.
- Introduce a novel bi-directional video encoder based on the selected keyframe.
- Train the selection model without labeled data by the weakly supervised reward calculated from generated captions.

Xiangxi Shi, J. Cai, S. Joty, J. Gui. Video Captioning with Boundary-Aware Hierarchical Language Decoding and Joint Video Prediction

Published in *Neural Computing*

- Introduce a binary gate into the low-level GRU language decoder to detect the language boundaries and generate captions at phases level with a hierarchical language decoder.
- Introduce the video and language reconstruction to learn the better representation for both sides.

Bastan M, **Shi X**, Gu J, et al. NTU ROSE Lab at TRECVID 2018: Ad-hoc Video Search and Video to Text[J]. 2018.

- Re-implemented the CST-captioning model and enhanced it with multiple additional data resources, including static frame features, motion features and audio features.
- Achieved the 3rd place in caption generation task and 5th place at retrieval task in TRECVID supported by NIST.

Shi X, Kang K, Cao Y. An iterative method for optical flow estimation with motion blur[C]

//2016 Visual Communications and Image Processing (VCIP). IEEE, 2016: 1-4. Present a method for estimating the optical flow of image sequences while considering the blur effect

- Performed two steps until convergence after an initial optical flow, 1) the blur kernel is estimated using the information from optical flow; 2) the optical flow is estimated considering the blur kernel.
- Achieved Average of Endpoint Error (AEE) of 0.79795.

OTHER WORKS

Few-Shot Recognition for Indian Food, ROSE Lab, NT *Nov.2018 - Sept. 2020*

- Implement few-shot recognition to realize food recognition on multiple datasets with limited data
- Improve the few-shot learning network with a distance prediction network
- Achieve accuracy of 71.28% for base classes, 74.56% for novel classes and 60.44% for all classes, better than the initial CVPR2018 paper claimed

Dispersion Detection Algorithm in Anomaly Detection Project, ROSE Lab, NTU *Aug.2017- Oct.2018*

- Implemented an algorithm to detect the dispersion event in videos as a clue of the video anomaly detection
- Implemented a threshold-based dispersion detection based on the dense of crossover points of different humans' tracks

Fire rescue training agent, summer research in University of Newcastle, Australia *Aug. 2016-Oct. 2016*

- Built a VR system for fire rescuing training, including a VR environment, intelligent agents and hardware using Unity
- Implemented a VR environment for test using Unity and C#
- Search the escape route using greedy algorithm

Automatic Navigation of Four-rotor UAV, research training program in USTC *Jun. 2015-Oct. 2015*

- Implemented the computer vision system for an UAV to avoid the carriers and fly safely during the trip based on
- Achieve 3D ground plane region and scene depth estimation based on monocular image.
- Apply fusion of image defocus, image saturation and dark channel prior to estimate the relative depth map of scene.
- Highest score for National Undergraduate Training Programs for Innovation and Entrepreneurship

PROGRAMMING & SKILLS

C++(***NOIP Fujian 2010 First Prize***), Python, Matlab, PyTorch, OpenCV, Vim, Unity3D, VirtualBox, Unix/Linux, Git