Xiangxi Shi

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

EDUCATION

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

WORK EXPERIENCE & INTERNSHIP

Adobe, Seattle WA *Jan.2024-now*
Internship Mainly focus on llm-based document image editing
Baidu, Seattle WA *Jun.2021-Sept.2021*
Internship Mainly focus on diffusion-based text to image generation
Adobe, Seattle WA *Jun.2020-Sept.2020*
Internship Mainly focus on large-scale video representation learning
ROSE Lab / PDCL Lab, Nanyang Technological University, Singapore *Aug.2017-Sept.2020*
Officer Mainly focus on vision-to-language generation

PAPERS & WORKSHOP

Xiangxi Shi, Zhonghua Wu, Stefan Lee, Viewpoint-Aware Visual Grounding in 3D Scenes
Accepted by **IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2024)**
- Introduce viewpoint prediction as an auxiliary task in the 3D visual grounding
- Propose a novel model that learns viewpoint prediction to reduce ambiguity in spatial-relation grounding

Xiangxi Shi, Stefan Lee, Benchmarking Out-of-Distribution Detection in Visual Question Answering
Published by **IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2024)**
- Introduce an Out-of-Distribution Detection to VQA task and create a benchmark datase
- Proposed a generation-based method and examine it with other existing OOD methods in our benchmark

J Gu, **Xiangxi Shi**, J Kuen, L Qi, R Zhang ADoPD: A Large-Scale Document Page Decomposition Dataset
Published by **The Eleventh International Conference on Learning Representations (ICLR 2024)**
- Introduce ADoPD, a large-scale document page decomposition dataset for multi-tasks in document area
- Data-driven document taxonomy discovery and model-Assisted Data Annotation

Xiangxi Shi, N. Xu, S. Lee Momentum-based Video-Text Model Pretraining for Moment Localization
- Introduced momentum-based transfer strategy to boost performance on the zero-shot video retrieval task
- Proposed a two-stage framework consisting of a post-tuned video retrieval model and a weight-lighted Score Refinement Network for moment localization adaptation

Xiangxi Shi, et al. Remember What You have drawn: Semantic Image Manipulation with Memory
- Disentangle the image features into texture and structure parts and encoded it with a set of latent memories
- 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 (ICCV2021)**
- First propose a set of learnable embedding to learning meta-class information for few-shot semantic image segmentation
- Quality Measurement Module (QMM) is proposed to measure the quality of all the support images

Xiangxi Shi, X. Yang, J. Gu, et. al. Finding It at Another Side: A Viewpoint-Adapted Matching Encoder for Change Captioning
Published by **16th European Conference on Computer Vision (ECCV2020)**
- propose a image encoder that explicitly distinguishes semantic changes from the viewpoint changes
- propose a reinforcement learning module that helps the model focus on the semantic change regions

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 novel bi-directional video encoder based on the selected keyframe to capture the key idea of a video clip
- A reinforcement learning is proposed to pick out the better key frame of a video to represent it

Xiangxi Shi, J. Cai, S. Joty, J. Gui. Video Captioning with Boundary-Aware Hierarchical Language Decoding and Joint Video Prediction
Published in **Neuralcomputing**
- Introduce a binary gate into the hierarchical GRU language decoder to generate captions at phrase level
- Introduce the video and language reconstruction to learn the better representation for both sides

PROGRAMMING & SKILLS

Python, C/C++, AWS, Matlab, PyTorch, Pytorch3D, Huggingface, Accelerate, Gradio, Pytorch Lightning, OpenCV, Vim, Unity3D, VirtualBox, Unix/Linux, Git
Github:https://github.com/Sxx1995