

Weiming Ren

w2ren@uwaterloo.ca | +1 (548) 990-2616 | cs.uwaterloo.ca/~w2ren/ | Github: wren93

Education

- Ph.D. in Computer Science**, University of Waterloo, GPA: 4.0/4.0 Aug 2023–present
M.Sc. in Applied Computing, University of Toronto, GPA: 4.0/4.0 Sep 2021–Dec 2022
BIT-ANU 2+2 Joint Undergraduate Program Aug 2017–Jun 2021
 - **Bachelor of Advanced Computing (Honours)**, Australian National University, GPA: 6.938/7
 - **B.S. in Computer Science and Technology**, Beijing Institute of Technology, GPA: 90.28/100

Experiences

University of Waterloo **Waterloo, Canada**
Graduate Student Researcher Sep 2023–present

- AnyV2V: A Plug-and-Play Framework For Any Video-to-Video Editing Tasks** [website] [arXiv]
 - Developed a training-free video editing framework that applies a black box image editing model to produce an edited first frame and uses an image-to-video (I2V) diffusion model to generate the edited video.
 - Implemented feature injection mechanisms in the spatial and temporal layers of the I2V diffusion U-Net and obtained the initial diffusion noise through DDIM inversion to align the edited video with the source video.

01.ai **Beijing, China**
Research Intern (remote) Aug 2023–present

- ConsistI2V: Enhancing Visual Consistency for Image-to-Video Generation** [website] [arXiv]
 - Developed an I2V generation method by augmenting spatial and temporal attention operations with the first-frame latent in a video diffusion model and devised a first-frame layout-guided noise initialization method for inference.
 - Built an I2V generation evaluation framework for FVD, FID, IS and CLIPSIM metrics on UCF-101 and MSR-VTT. Collected an additional benchmark I2V-Bench consisting ~3k videos for visual quality and consistency evaluation.

Samsung AI Center Toronto **Toronto, Canada**
Research Intern May 2022–Apr 2023

- Learning Efficient Video Masked Autoencoders using Self-Attention Modelled Patch Learning**
 - Developed a patch sampling algorithm for increasing finetuning efficiency and throughput of video masked autoencoders (MAE) using reinforcement learning and self-attention scores as the supervision signal.
 - Evaluated the proposed sampling algorithm by finetuning a pretrained video MAE model on various action recognition datasets and achieved 40% fewer inference GFLOPs and GPU memory cost with competitive recognition accuracy.

- Vision-Language Models (VLMs) for Efficient Video Action Recognition**
 - Implemented a CLIP-based action recognition model containing a modified CLIP ViT encoder for frame-level feature extraction and a temporal transformer for cross-frame feature fusion.
 - Finetuned the CLIP-based action recognition model on a real-world in-house action dataset and proved that CLIP achieved comparable performance to several video transformer variants that rely on Kinetics pretraining.

- Data-Efficient Self-supervised Learning (SSL) for Video Action Recognition**
 - Aimed to use self-supervised learning to minimize the requirement of pretraining data for action recognition models.
 - Implemented a BYOL-based SSL method, which aligns the representations of two different clips from semantically similar videos to learn spatiotemporal-persistent video representations.
 - Pretrained X3D and Slow-ResNet50 models using the proposed SSL method on ActivityNet.
 - Evaluated the action recognition performance of the pretrained models on ActivityNet, UCF-101 and HMDB-51.

Awards

- David R. Cheriton Graduate Scholarship, University of Waterloo Sep 2023
Chancellor's Letter of Commendation, Australian National University May 2021
Outstanding Winner, COMAP's Mathematical Contest in Modeling (MCM) Apr 2020
The Second-class Scholarship (4 times), Beijing Institute of Technology Mar/Sep 2018, Mar/Sep 2019

Skills

Programming: Python (fluent), C/C++ (proficient), Arduino, Java, HTML/CSS/Javascript (prior experience)
Libraries & Tools: PyTorch, NumPy, Scikit-learn, Huggingface Transformers/Diffusers, Pandas, Matplotlib, Git