Weiming Ren

Email	w2ren@uwaterloo.ca	**Homepage**	cs.uwaterloo.ca/~w2ren/
LinkedIn	linkedin.com/in/wmren/	**GitHub**	github.com/wren93

Research Interests

Generative Modelling: developing novel algorithms based on diffusion models for controllable video generation, image and video editing, and image restoration.

Multimodal Understanding: designing efficient model architectures and data curation pipelines to enhance large multimodal models (LMMs) for image and video understanding.

Education

Sep 2023 - Present **Ph.D. in Computer Science** - University of Waterloo
GPA: 4.0/4.0, Advisor: Prof. Wenhu Chen
Related Courses: Neural Networks, Software Engineering for Big Data and AI

Sep 2021 - Dec 2022 **MSc. in Applied Computing** - University of Toronto
GPA: 4.0/4.0
Related Courses: Machine Learning for Healthcare, Neural Networks and Deep Learning, Machine Learning for Machine Vision as Inverse Graphics

Jul 2019 - Jun 2021 **Bachelor of Advanced Computing (Honours)** - The Australian National University
GPA: 6.938/7.0, Graduated with First-class Honours

Aug 2017 - Jun 2021 **B.S. in Computer Science and Technology** - Beijing Institute of Technology
GPA: 90.28/100

Publications

VISTA: Enhancing Long-Duration and High-Resolution Video Understanding by Video Spatiotemporal Augmentation
Weiming Ren, Huan Yang, Jie Min, Cong Wei, Wenhu Chen
IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) 2025.

OmniEdit: Building Image Editing Generalist Models Through Specialist Supervision
Cong Wei*, Zheyang Xiong*, Weiming Ren, Xinrun Du, Ge Zhang, Wenhu Chen
International Conference on Learning Representations (ICLR), 2025.

MMLU-Pro: A More Robust and Challenging Multi-Task Language Understanding Benchmark
Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, Tianle Li, Max Ku, Kai Wang, Alex Zhuang, Rongqi Fan, Xiang Yue, Wenhu Chen
NeurIPS Dataset and Benchmark Track Spotlight, 2024.

Video Diffusion Models: A Survey
Andrew Melnik, Michal Ljubljanac, Cong Lu, Qi Yan, Weiming Ren, Helge Ritter
Transactions on Machine Learning Research (TMLR), 2024.

AnyV2V: A Plug-and-Play Framework for Any Video-to-Video Editing Tasks
Max Ku*, Cong Wei*, Weiming Ren*, Huan Yang, Wenhu Chen
**Equal First-Authors. Transactions on Machine Learning Research (TMLR), 2024.*

StructLM: Towards Building Generalist Models for Structured Knowledge Grounding

Alex Zhuang, Ge Zhang, Tianyu Zheng, Xinrun Du, Junjie Wang, Weiming Ren, Stephen W Huang, Jie Fu, Xiang Yue, Wenhu Chen

Conference on Language Modeling (COLM), 2024.

ConsistI2V: Enhancing Visual Consistency for Image-to-Video Generation

Weiming Ren, Huan Yang, Ge Zhang, Cong Wei, Xinrun Du, Wenhao Huang, Wenhu Chen

Transactions on Machine Learning Research (TMLR), 2024.

MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI

Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, Wenhu Chen

IEEE / CVF Computer Vision and Pattern Recognition Conference (CVPR) Oral, Best Paper Finalist, 2024.

HiCu: Leveraging Hierarchy for Curriculum Learning in Automated ICD Coding

Weiming Ren, Ruijing Zeng, Tongzi Wu, Tianshu Zhu, Rahul G. Krishnan

Machine Learning for Healthcare Conference (MLHC), 2022.

Experiences

Sep 2024 - Present Samsung AI Centre Toronto, Toronto, Canada
Research Intern

Developed a semantic-aware image super-resolution (SR) model by injecting fine-grained semantic representations (e.g. DINO/SAM features) into a pretrained diffusion-based SR model.

Aug 2023 - Nov 2024 **01.AI**, Beijing, China
Research Intern

- Proposed a video spatiotemporal augmentation method and curated a synthetic long-duration and high-resolution video instruction-following dataset. Conducted supervised finetuning on the proposed dataset to enhance the video understanding capabilities of video LMMs.
- Developed a training-free video editing framework (AnyV2V) based on pretrained image editing models and image-to-video (I2V) diffusion models. Implemented feature injection mechanisms in the I2V U-Net and DDIM inversion to align the edited video with the source video.
- Designed and implemented an image-to-video generation method (ConsistI2V) 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.

May 2022 - Apr 2023 **Samsung AI Centre Toronto**, Toronto, Canada
Research Intern

- Developed a patch sampling algorithm for increasing finetuning efficiency and throughput of video masked autoencoders (MAE) using reinforcement learning and attention scores as supervision signals. Achieved 40% fewer inference GFLOPs and GPU memory costs.
- Leveraged various vision-language foundational models (e.g. CLIP/BLIP) and self-supervised learning methods (e.g. MAE/BYOL) and designed data-efficient video action recognition algorithms for real-world in-house action recognition.

Awards

Sep 2023 David R. Cheriton Graduate Scholarship, University of Waterloo

May 2021 Chancellor's Letter of Commendation, Australian National University

Apr 2020 Outstanding Winner, COMAP's Mathematical Contest in Modeling (MCM)

Mar/Sep 2018, Mar/Sep 2019 The Second-class Scholarship (4 times), Beijing Institute of Technology

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