Yiwu Zhong

Homepage | Google Scholar | Github
yzhong52@wisc.edu

EDUCATION

09/2017 - 12/2023 **University of Wisconsin-Madison**

Ph.D. student in Computer Science | Advisor: Prof. Yin Li

09/2013-07/2017 **University of Electronic Science and Technology of China**

B.S. in Electronic Engineering | GPA: 3.82/4.00

INTERESTS

Areas Computer Vision, Natural Language Processing, Deep Learning

Topics Vision-language pretraining, Structural Visual Understanding

PUBLICATIONS

- **Learning Procedure-aware Video Representation from Instructional Videos and Their Narrations**

  Yiwu Zhong, Licheng Yu, Yang Bai, Shangwen Li, Xueting Yan*, Yin Li*

  Published in Conference on Computer Vision and Pattern Recognition (**CVPR**) 2023

- **RegionCLIP: Region-based Language-Image Pretraining**

  Yiwu Zhong, Jianwei Yang, Pengchuan Zhang, Chunyuan Li, Noel Codella, Liunian Harold Li, Luowei Zhou, Xiyang Dai, Lu Yuan, Yin Li, Jianfeng Gao

  Published in Conference on Computer Vision and Pattern Recognition (**CVPR**) 2022

- **Grounded Language-Image Pretraining**

  Liunian Harold Li, Pengchuan Zhang, Haotian Zhang, Jianwei Yang, Chunyuan Li, **Yiwu Zhong**, Lijuan Wang, Lu Yuan, Lei Zhang, Jenq-Neng Hwang, Kai-Wei Chang, Jianfeng Gao

  Published in Conference on Computer Vision and Pattern Recognition (**CVPR**) 2022

- **Learning to Generate Scene Graph from Natural Language Supervision**

  Yiwu Zhong, Jing Shi, Jianwei Yang, Chenliang Xu, Yin Li

  Published in International Conference on Computer Vision (**ICCV**) 2021

- **A Simple Baseline for Weakly-Supervised Scene Graph Generation**

  Jing Shi, **Yiwu Zhong**, Ning Xu, Yin Li, Chenliang Xu

  Published in International Conference on Computer Vision (**ICCV**) 2021

- **Comprehensive Image Captioning via Scene Graph Decomposition**

  Yiwu Zhong, Liwei Wang, Jianshu Chen, Dong Yu and Yin Li

  Published in European Conference on Computer Vision (**ECCV**) 2020

AWARDS

03/2023 CVPR 2023 Doctoral Consortium Award (**top 13%**)

06/2022 CVPR 2022 Best Paper Finalist (**top 0.4%**)

EXPERIENCE

05/2022-12/2022 **Research Intern, Meta, Menlo Park, USA**

Mentors: Xueting Yan, Licheng Yu

- Proposed a novel method for learning video representation from large-scale instructional videos, without any human annotation.
- Our method jointly learned a video encoder that captures the concepts of action steps, as well as a diffusion model that can reason about the temporal dependencies among steps.
- Established new state-of-the-art results on both step classification and forecasting tasks on COIN and EPIC-Kitchens-100 benchmarks, and enabled zero-shot step forecasting and generating diverse step predictions.
- Our work was published in **CVPR 2023**.

05/2021-11/2021 **Research Intern, Microsoft Research, Seattle, USA**

Mentors: Jianwei Yang, Jianfeng Gao

- Proposed a novel vision-language pretraining method for learning region-level visual representation.
- Our method leveraged a pretrained CLIP model to align image regions with template region descriptions, and pretrained our model to match these region-text pairs.
- Our method established new state-of-the-art on COCO and LVIS datasets when transferred to open-vocabulary object detection, and demonstrated promissing results on zero-shot inference for object detection.
- Our work was published in **CVPR 2022**.

05/2019-08/2019 **Research Intern, Tencent AI Lab, Seattle, USA**

Mentors: Liwei Wang, Dong Yu

- Proposed a novel and unified framework for image captioning. Our model simultaneously outperformed the state-of-the-art methods which were designed for diverse captioning, controllable captioning and grounded captioning, while maintaining caption quality.
- Our method decomposed image scene graph into a set of sub-graphs, identified meaningful sub-graphs and decoded the top-ranked sub-graphs into captions.
- Our work was published in **ECCV 2020**.

05/2018-08/2018

Research Intern, Tencent AI Lab, Shenzhen, China

Mentors: Jia Xu, Meng Fang

- Proposed a novel method for zero-shot image scene graph generation.
- Addressed the challenging problem of compositionality by exploiting knowledge graph, metric learning and hard examples mining.
- Our model outperformed the baselines in zero-shot scene graph generation while compared favorably to state-of-the-art methods designed for fully supervised scene graph generation.

01-05/2020, 2018

Teaching Assistant, UW-Madison, USA

- CS 540: Artificial Intelligence
- CS 559: Computer Graphics

OTHER PROJECTS

08/2020 - Present

Discovery of In-the-wild Facial Expressions from Videos

Discovered the typical patterns of facial behaviors from videos by leveraging 3D face reconstruction models and unsupervised clustering methods.

04/2020-05/2020

Instance Segmentation for Cataract Surgical Video

Applied and compared the latest image segmentation models (PointRend and Mask R-CNN) on the video dataset of cataract surgery.

09/2018-04/2019

Scene Graph Generation

Focused on image scene graph generation task, which employed object detectors (Faster-RCNN and YOLO) to detect image objects and used GCN (Graph Convolutional Network) to infer relationships between objects.

SKILLS

- **Programming Languages**

  Python, Julia, Matlab, Java

- **Deep Learning Frameworks**

  Pytorch, TensorFlow