

Zhengyuan Yang

Senior Researcher, Microsoft, Redmond, WA

☎ (585) 435-3638 | ✉ zhengyuan.yang13@gmail.com | 🏠 zhengyuan.info | 🎓 Scholar

Research Summary

My research interests involve the intersection of computer vision and natural language processing, including vision+language understanding and generation, multimodal learning, and multimodal foundation models.

Education

University of Rochester

Rochester, New York

PH.D. IN COMPUTER SCIENCE ADVISOR: PROF. JIEBO LUO

2016 - 2021

- Thesis: Visual Grounding: Building Cross-Modal Visual-Text Alignment (**ACM SIGMM Outstanding Ph.D. Thesis**)
- Thesis Committee: Jiebo Luo (advisor), Dan Gildea, Chenliang Xu, Ehsan Hoque, Zhiyao Duan

University of Science and Technology of China

Hefei, China

B.E. IN ELECTRONIC ENGINEERING AND INFORMATION SCIENCE

2012 - 2016

Selected Awards

| | |
|--|------|
| ACM SIGMM Award for Outstanding Ph.D. Thesis | 2022 |
| ECCV 2022 Outstanding Reviewer | 2022 |
| Winner of CVPR 2021 TextCaps Challenge | 2021 |
| Winner of CVPR 2021 ReferIt3D Challenge | 2021 |
| CVPR 2021 Outstanding Reviewer | 2021 |
| Twitch Research Fellowship | 2020 |
| Best Industry Related Paper Award (BIRPA) in ICPR 2018, (1/1258) | 2018 |

Services

| | | |
|----------------------------------|---|------------|
| Exhibits and Demos Chair: | IEEE International Conference on Multimedia and Expo (ICME) | 2024 |
| Area Chair: | ACM Multimedia Conference (ACMMM) | 2024 |
| Senior Program Committee: | AAAI Conference on Artificial Intelligence (AAAI) | 2023, 2024 |
| Associate Editor: | IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) | 2022-2024 |
| Guest Editor: | IEEE TMM SI on "Large Multimodal Models for Dynamic Visual Scene Understanding" | 2024 |
| | IEEE TCSVT SI on "AI-Generated Content for Multimedia" | 2023 |
| Conference Reviewer: | CVPR, ICCV, ECCV, NeurIPS, ICLR, ICML, ACL, EMNLP, ACMMM, AAAI, ACCV, WACV, ICME, ICIP. | |
| Journal Reviewer: | TPAMI, IJCV, TIP, TMM, TCybernetics, TCSVT, Pattern Recognition, Neurocomputing, TBioCAS, Access. | |

Professional Experience

Microsoft

Redmond, WA

SENIOR RESEARCHER

June 2021 - Current

- Research on multimodal understanding and generation.

Microsoft

Bellevue, WA

RESEARCH INTERN SUPERVISOR: DR. YIJUAN LU, DR. JIANFENG WANG, DR. XI YIN

May 2020 to Aug 2020

Tencent AI Lab at Bellevue

Bellevue, WA

RESEARCH INTERN SUPERVISOR: DR. BOQING GONG, DR. LIWEI WANG

Jan 2019 to Apr 2019

SnapChat Research

RESEARCH INTERN SUPERVISOR: DR. YUNCHENG LI, DR. LINJIE YANG, DR. NING ZHANG

Venice, CA

May 2018 to Aug 2018

SAIC USA

RESEARCH INTERN SUPERVISOR: DR. JERRY YU

San Jose, CA

May 2017 to Aug 2017

Selected Publications

- IJCAI 2024** Jie An, **Zhengyuan Yang**, Jianfeng Wang, Linjie Li, Zicheng Liu, Lijuan Wang, Jiebo Luo, "Bring Metric Functions into Diffusion Models," The 32nd International Joint Conference on Artificial Intelligence (IJCAI), Jeju, August 2024.
- CVPR 2024a** Chaoyi Zhang, Kevin Lin, **Zhengyuan Yang**, Jianfeng Wang, Linjie Li, Chung-Ching Lin, Zicheng Liu, Lijuan Wang, "MM-Narrator: Narrating Long-form Videos with Multimodal In-Context Learning," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, June 2024. **Highlight presentation.**
- CVPR 2024b** Tan Wang, Linjie Li, Kevin Lin, Yuanhao Zhai, Chung-Ching Lin, **Zhengyuan Yang**, Hanwang Zhang, Zicheng Liu, Lijuan Wang, "DisCo: Disentangled Control for Realistic Human Dance Generation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, June 2024.
- CVPR 2024c** Zichen Miao, Jiang Wang, Ze Wang, **Zhengyuan Yang**, Lijuan Wang, Qiang Qiu, Zicheng Liu, "Training Diffusion Models Towards Diverse Image Generation with Reinforcement Learning," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, June 2024.
- CVPR 2024d** Jieliu Qiu, Jiacheng Zhu, William Han, Aditesh Kumar, Karthik Mittal, Claire Jin, **Zhengyuan Yang**, Linjie Li, Jianfeng Wang, Ding Zhao, Bo Li, Lijuan Wang, "MMSum: A Dataset for Multimodal Summarization and Thumbnail Generation of Videos," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, June 2024. **Highlight presentation.**
- Book 2023** Chunyuan Li*, Zhe Gan*, **Zhengyuan Yang***, Jianwei Yang*, Linjie Li*, Lijuan Wang, Jianfeng Gao, "Multimodal Foundation Models: From Specialists to General-Purpose Assistants," Foundations and Trends in Computer Graphics and Vision, 2023.
- Arxiv 2023a** **Zhengyuan Yang**, Jianfeng Wang, Linjie Li, Kevin Lin, Chung-Ching Lin, Zicheng Liu, Lijuan Wang, "Idea2Img: Iterative Self-Refinement with GPT-4V(ision) for Automatic Image Design and Generation."
- Arxiv 2023b** **Zhengyuan Yang***, Linjie Li*, Kevin Lin*, Jianfeng Wang*, Chung-Ching Lin*, Zicheng Liu, Lijuan Wang*, "The Dawn of LMMs: Preliminary Explorations with GPT-4V(ision)." (**Exploratory work cataloguing use of GPT-4V**).
- Arxiv 2023c** Weihao Yu*, **Zhengyuan Yang***, Linjie Li, Jianfeng Wang, Kevin Lin, Zicheng Liu, Xinchao Wang, Lijuan Wang, "MM-Vet: Evaluating Large Multimodal Models for Integrated Capabilities."
- Arxiv 2023d** **Zhengyuan Yang**, Linjie Li, Jianfeng Wang, Kevin Lin, Ehsan Azarnasab, Faisal Ahmed, Zicheng Liu, Ce Liu, Michael Zeng, Lijuan Wang, "MM-REACT: Prompting ChatGPT for Multimodal Reasoning and Action."
- ICCV 2023a** Yushi Hu, Hang Hua, **Zhengyuan Yang**, Weijia Shi, Noah A. Smith, Jiebo Luo, "PromptCap: Prompt-Guided Task-Aware Image Captioning," IEEE International Conference on Computer Vision (ICCV), Paris, France, October 2023.
- ICCV 2023b** Tan Wang, Kevin Lin, Linjie Li, Chung-Ching Lin, **Zhengyuan Yang**, Hanwang Zhang, Zicheng Liu, Lijuan Wang, "Equivariant Similarity for Vision-Language Foundation Models," IEEE International Conference on Computer Vision (ICCV), Paris, France, October 2023. **Oral presentation.**
- ACL 2023** Shengming Yin, Chenfei Wu, Huan Yang, Jianfeng Wang, Xiaodong Wang, Minheng Ni, **Zhengyuan Yang**, Linjie Li, Shuguang Liu, Fan Yang, Jianlong Fu, Gong Ming, Lijuan Wang, Zicheng Liu, Houqiang Li, Nan Duan, "NUWA-XL: Diffusion over Diffusion for eXtremely Long Video Generation." Annual Meeting of the Association for Computational Linguistics (ACL), Toronto, Canada, July 2023. **Oral presentation.**
- IJCAI 2023** Xiaodong Wang, Chenfei Wu, Shengming Yin, Minheng Ni, Jianfeng Wang, Linjie Li, **Zhengyuan Yang**, Fan Yang, Lijuan Wang, Zicheng Liu, Yuejian Fang, Nan Duan, "Learning 3D Photography Videos via Self-supervised Diffusion on Single Images." The 32nd International Joint Conference on Artificial Intelligence (IJCAI), Macao, August 2023.
- CVPR 2023** **Zhengyuan Yang**, Jianfeng Wang, Zhe Gan, Linjie Li, Kevin Lin, Chenfei Wu, Nan Duan, Zicheng Liu, Ce Liu, Michael Zeng, Lijuan Wang, "ReCo: Region-Controlled Text-to-Image Generation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, June 2023.

| | |
|--------------------|---|
| ICLR 2023 | Chenglei Si, Zhe Gan, Zhengyuan Yang , Shuohang Wang, Jianfeng Wang, Jordan Boyd-Graber, Lijuan Wang, "Prompting GPT-3 To Be Reliable," The Eleventh International Conference on Learning Representations (ICLR), Kigali, Rwanda, May 2023. |
| ECCV 2022 | Zhengyuan Yang , Zhe Gan, Jianfeng Wang, Xiaowei Hu, Faisal Ahmed, Zicheng Liu, Yumao Lu, Lijuan Wang, "UniTAB: Unifying Text and Box Outputs for Grounded Vision-Language Modeling," European Conference on Computer Vision (ECCV), Tel Aviv, Israel, October 2022. Oral presentation (2.7%). |
| TPAMI 2023 | Jiajun Deng, Zhengyuan Yang , Daqing Liu, Tianlang Chen, Wengang Zhou, Yanyong Zhang, Houqiang Li, Wanli Ouyang, "TransVG++: End-to-End Visual Grounding with Language Conditioned Vision Transformer," IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2023. |
| TMLR 2022 | Jianfeng Wang, Zhengyuan Yang , Xiaowei Hu, Linjie Li, Kevin Lin, Zhe Gan, Zicheng Liu, Ce Liu, Lijuan Wang, "GIT: A Generative Image-to-text Transformer for Vision and Language," Transactions on Machine Learning Research (TMLR), 2022. |
| CVPR 2022 | Xiaowei Hu, Zhe Gan, Jianfeng Wang, Zhengyuan Yang , Zicheng Liu, Yumao Lu and Lijuan Wang, "Scaling Up Vision-Language Pre-training for Image Captioning," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, June 2022. |
| Arxiv 2021 | Jianfeng Wang, Xiaowei Hu, Zhe Gan, Zhengyuan Yang , Xiyang Dai, Zicheng Liu, Yumao Lu and Lijuan Wang, "UFO: A UniFied TransOrmer for Vision-Language Representation Learning." |
| AAAI 2022 | Zhengyuan Yang , Zhe Gan, Jianfeng Wang, Xiaowei Hu, Yumao Lu, Zicheng Liu, Lijuan Wang, "An Empirical Study of GPT-3 for Few-Shot Knowledge-Based VQA," The 36th AAAI Conference on Artificial Intelligence (AAAI), Vancouver, BC, February 2022. Oral presentation (4.2%). |
| ICPR 2022 | Zhengyuan Yang , Jingen Liu, Jing Huang, Xiaodong He, Tao Mei, Chenliang Xu, Jiebo Luo, "Cross-modal Contrastive Distillation for Instructional Activity Anticipation," International Conference on Pattern Recognition (ICPR), Montreal, Quebec, Canada, August 2022. |
| ICCV 2021a | Zhengyuan Yang , Songyang Zhang, Liwei Wang, Jiebo Luo, "SAT: 2D Semantics Assisted Training for 3D Visual Grounding," IEEE International Conference on Computer Vision (ICCV), Online, 2021. Oral presentation (3.4%). |
| ICCV 2021b | Jiajun Deng, Zhengyuan Yang , Tianlang Chen, Wengang Zhou, Houqiang Li, "TransVG: End-to-End Visual Grounding with Transformers," IEEE International Conference on Computer Vision (ICCV), Online, 2021. |
| CVPR 2021a | Zhengyuan Yang , Yijuan Lu, Jianfeng Wang, Xi Yin, Dinei Florencio, Lijuan Wang, Cha Zhang, Lei Zhang, Jiebo Luo, "TAP: Text-Aware Pre-training for Text-VQA and Text-Caption," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Online, June 2021. Oral presentation (4.0%). |
| CVPR 2021b | Liwei Wang, Jing Huang, Yin Li, Kun Xu, Zhengyuan Yang , Dong Yu, "Improving Weakly Supervised Visual Grounding by Contrastive Knowledge Distillation," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Online, June 2021. |
| ECCV 2020 | Zhengyuan Yang , Tianlang Chen, Liwei Wang, Jiebo Luo, "Improving One-stage Visual Grounding by Recursive Sub-query Construction," European Conference on Computer Vision (ECCV), Online, August 2020. |
| ACL 2020 | Yongjing Yin, Fandong Meng, Jinsong Su, Chulun Zhou, Zhengyuan Yang , Jie Zhou, Jiebo Luo, "A Novel Graph-based Multi-modal Fusion Encoder for Neural Machine Translation," Annual Meeting of the Association for Computational Linguistics (ACL), Online, July 2020. |
| T-CSVT 2020 | Zhengyuan Yang , Tushar Kumar, Tianlang Chen, Jingsong Su, Jiebo Luo, "Grounding-Tracking-Integration," IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT). |
| ACMMM 2020 | Huan Lin, Fandong Meng, Jinsong Su, Yongjing Yin, Zhengyuan Yang , Yubin Ge, Jie Zhou, Jiebo Luo, "Dynamic Context-guided Capsule Network for Multimodal Machine Translation," ACM Multimedia Conference, Seattle, WA, October 2020 |
| ICCV 2019 | Zhengyuan Yang , Boqing Gong, Liwei Wang, Wenbing Huang, Dong Yu, Jiebo Luo, "A Fast and Accurate One-Stage Approach to Visual Grounding," IEEE International Conference on Computer Vision (ICCV), Seoul, South Korea, 2019. Oral presentation (4.3%). |
| ICPR 2020a | Zhengyuan Yang , Yuncheng Li, Linjie Yang, Ning Zhang, Jiebo Luo, "Weakly Supervised Body Part Parsing with Pose based Part Priors," International Conference on Pattern Recognition (ICPR), Millan, Italy, January, 2020. |

| | |
|--------------------|---|
| ICPR 2020b | Zhengyuan Yang , Amanda Kay, Yuncheng Li, Wendi Cross, Jiebo Luo, "Pose-based Body Language Recognition for Emotion and Psychiatric Symptom Interpretation," International Conference on Pattern Recognition (ICPR), Millan, Italy, January, 2020. |
| CVPR 2019 | Mengshi Qi, Weijian Li, Zhengyuan Yang , Yunhong Wang, Jiebo Luo, "Attentive Relational Networks for Mapping Images to Scene Graphs," IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019. |
| ICME 2019 | Zhengyuan Yang , Yixuan Zhang, Jiebo Luo, "Human-Centered Emotion Recognition in Animated GIFs with Facial Landmarks," IEEE International Conference on Multimedia and Expo (ICME), Shanghai, China, 2019. |
| T-CSVT 2018 | Zhengyuan Yang , Yuncheng Li, Jianchao Yang, Jiebo Luo, "Action Recognition with Spatio-Temporal Visual Attention on Skeleton Image Sequences," IEEE Transactions on Circuits and Systems for Video Technology (T-CSVT). |
| ICPR 2018a | Zhengyuan Yang , Yixuan Zhang, Jerry Yu, Junjie Cai, Jiebo Luo, "End-to-end Multi-Modal Multi-Task Vehicle Control for Self-Driving Cars with Visual Perceptions," International Conference on Pattern Recognition (ICPR), Beijing, China, 2018. Best Industry Related Paper Award (BIRPA) (1/1258) . |
| ICPR 2018b | Zhengyuan Yang , Yuncheng Li, Jianchao Yang, Jiebo Luo, "Action Recognition with Visual Attention on Skeleton Images," International Conference on Pattern Recognition (ICPR), Beijing, China, 2018. |

Talks and Teaching Experience

| | |
|--|------------------------|
| Tutorial Talk – Alignments in Text-to-Image Generation | <i>Vancouver, BC</i> |
| • CVPR 2023 Tutorial on Recent Advances in Vision Foundation Models | 2023 |
| Invited Talk – Towards Cross-Modal Visual-Text Understanding and Generation | <i>Tokyo, Japan</i> |
| • ACM MM Asia 2022 Workshop on Multimedia Understanding with Pre-trained Models | 2022 |
| Tutorial Talk – Unified Image-Text Modeling | <i>New Orleans, LA</i> |
| • CVPR 2022 Tutorial on Recent Advances in Vision-and-Language Pre-training | 2022 |
| Invited Talk – SAT: 2D Semantics Assisted Training for 3D Visual Grounding | <i>Online</i> |
| • CVPR Workshop 2021 on Language for 3D Scenes | 2021 |
| Guest Lecture – Vision-and-language | <i>Rochester, NY</i> |
| • CS440 Data Mining, University of Rochester | 2021 |
| Guest Lecture – CNN and Feature Visualization | <i>Rochester, NY</i> |
| • CS440 Data Mining, University of Rochester | 2020 |
| Teaching Assistant – CS446 Machine Learning | <i>Rochester, NY</i> |
| • Dept. of Computer Science, University of Rochester | 2018 |
| Teaching Assistant – CS172 Data Structures and Algorithms | <i>Rochester, NY</i> |
| • Dept. of Computer Science, University of Rochester | 2017 |
| Teaching Assistant – CS242 Intro to Artificial Intelligence | <i>Rochester, NY</i> |
| • Dept. of Computer Science, University of Rochester | 2017 |

Selected Patents

“Weakly supervised semantic parsing”, *US Patent Number: 11,182,603* (Granted, 11/2021).