Zhengyuan Yang

Senior Researcher, Microsoft, Bellevue, 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, such as vision+language and multimodal learning.

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*

Professional Experience

Microsoft — *Bellevue, WA*

Senior Researcher — *June 2021 - Current*

- Research on vision-language understanding and generation.

Microsoft — *Bellevue, WA*

Research Intern **Supervisor**: Dr. Yijuan Lu, Dr. Jianfeng Wang, Dr. Xi Yin — *May 2020 to Aug 2020*

- TextVQA and TextCaption (CVPR 2021a).

Tencent AI Lab at Bellevue — *Bellevue, WA*

Research Intern **Supervisor**: Dr. Boqing Gong, Dr. Liwei Wang — *Jan 2019 to Apr 2019*

- Real time image-text grounding (ICCV 2019).

SnapChat Research — *Venice, CA*

Research Intern **Supervisor**: Dr. Yuncheng Li, Dr. Linjie Yang, Dr. Ning Zhang — *May 2018 to Aug 2018*

- Weakly-supervised human part parsing (ICPR 2020a).

SAIC USA — *San Jose, CA*

Research Intern **Supervisor**: Dr. Jerry Yu — *May 2017 to Aug 2017*

- Camera-based E2E self-driving (Best Industry Related Paper Award in ICPR 2018).

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

Associate Editor:	IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)	2022-2024
Senior Program Committee:	AAAI Conference on Artificial Intelligence (AAAI)	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.	

Selected Publications

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%)**.

Arxiv 2022a 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," Under Review.

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 TransfOrmer for Vision-Language Representation Learning," Under Review.

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

Invited Talk – Towards Cross-Modal Visual-Text Understanding and Generation *Tokyo, Japan*
- ACM MM Asia 2022 Workshop on Multimedia Understanding with Pre-trained Models 2022

Tutorial Talk – Unified Image-Text Modeling *New Orleans, LA*
- CVPR 2022 Tutorial on Recent Advances in Vision-and-Language Pre-training 2022

Invited Talk – SAT: 2D Semantics Assisted Training for 3D Visual Grounding *Online*
- CVPR Workshop 2021 on Language for 3D Scenes 2021

Guest Lecture – Vision-and-language *Rochester, NY*
- CS440 Data Mining, University of Rochester 2021

Guest Lecture – CNN and Feature Visualization *Rochester, NY*
- CS440 Data Mining, University of Rochester 2020

Teaching Assistant – CS446 Machine Learning *Rochester, NY*
- Dept. of Computer Science, University of Rochester 2018

Teaching Assistant – CS172 Data Structures and Algorithms *Rochester, NY*
- Dept. of Computer Science, University of Rochester 2017

Teaching Assistant – CS242 Intro to Artificial Intelligence *Rochester, NY*
- Dept. of Computer Science, University of Rochester 2017

Selected Patents

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