

# Zongxin Yang, Ph.D.

✉ zongxinyang1996@gmail.com

in Zongxin Yang

🌐 <https://z-x-yang.github.io/>

## Experience

- 2024/08 – Now     **Postdocral Researcher.** Harvard Medical School, Harvard University. Boston, MA, US.
- 2021/11 – 2024/07     **Postdocral Researcher.** CCAI, Zhejiang University. Hangzhou, China.
- 2019/09 – 2021/09     **Research Intern.** Baidu Research. Beijing, China.  
**Outstanding Intern of 2021.**

## Education

- 2018/08 – 2021/09     **Ph.D. in Computer Science**  
**University of Technology Sydney**  
Advisor: Prof. Yi Yang  
Thesis title: *Sequence Modelling with Deep Learning for Visual Content Generation and Understanding.*
- 2014/09 – 2018/06     **B.E. in Electronic and Information Engineering**  
**University of Science and Technology of China**  
GPA: 3.99/4.3     Ranking: 3/95     **Guo Moruo Scholarship (Highest Scholarship)**

## Research Publications

### Preprints and Journal Articles

- 1 D. Zhou, Y. Li, F. Ma, \*, Z. Yang, and Y. Yang, "Migc++: Advanced multi-instance generation controller for image synthesis," *TPAMI*, 2025.
- 2 **Z. Yang**, X. Wang, J. Miao, Y. Wei, W. Wang, and Y. Yang, "Scalable video object segmentation with identification mechanism," *TPAMI*, 2024.
- 3 **Z. Yang**, Y. Wei, and Y. Yang, "Collaborative video object segmentation by multi-scale foreground-background integration," *TPAMI*, 2021.
- 4 J. Ma\*, **Z. Yang\***, S. Kim, *et al.*, "Medsam2: Segment anything in 3d medical images and videos," *arXiv preprint arXiv:2504.03600*, 2025.
- 5 Y. Xu, **Z. Yang**, and Y. Yang, "Photorealistic text-to-3d avatar generation with constraints for decoupled geometry and appearance," *ToMM*, 2025.
- 6 X. Pan, **Z. Yang\***, S. Bai, and Y. Yang, "Gd<sup>2</sup>-nerf: Generative detail compensation via gan and diffusion for one-shot generalizable neural radiance fields," *ToMM*, 2025.
- 7 J. Liu and **Z. Yang**, "Test-time adaptation for real-world video adverse weather restoration with meta batch normalization," *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.
- 8 Y. Zhu, K. Li, and **Z. Yang**, "Exploiting efficientsam and temporal coherence for audio-visual segmentation," *IEEE Transactions on Multimedia*, 2025.
- 9 K. Li, T. Jiang, **Z. Yang**, Y. Yang\*, Y. Zhuang, and J. Xiao, "Explore synergistic interaction across frames for interactive video object segmentation," *TCSVT*, 2024.
- 10 C. Liang, L. Zhu, **Z. Yang**, W. Chen, and Y. Yang, "Noise-tolerant hybrid prototypical learning with noisy web data," *ToMM*, vol. 20, no. 10, pp. 1–19, 2024.
- 11 M. Li, J. Tao, **Z. Yang**, and Y. Yang, "Human101: Training 100+ fps human gaussians in 100s from 1 view," *Preprint*, 2024.

- 12 W. Song, H. Jiang, Z. Yang, R. Quan, and Y. Yang, "Insert anything: Image insertion via in-context editing in dit," *arXiv preprint arXiv:2504.15009*, 2025.
- 13 Y. Lu, F. Ni, H. Wang, *et al.*, "Show me a video: A large-scale narrated video dataset for coherent story illustration," *TMM*, 2023.
- 14 S. Jiao, V. Goel, S. Navasardyan, *et al.*, "Collaborative content-dependent modeling: A return to the roots of salient object detection," *TIP*, 2023.

## Conference Proceedings

- 1 K. Li, **Z. Yang\***, L. Chen, Y. Yang, and J. Xiao, "Catr: Combinatorial-dependence audio-queried transformer for audio-visual video segmentation," in *ACM MM (Best Paper Award, Corresponding author)*, 2023.
- 2 **Z. Yang**, G. Chen, X. Li, W. Wang, and Y. Yang, "Doraemongpt: Toward understanding dynamic scenes with large language models (exemplified as a video agent)," in *ICML*, 2024.
- 3 **Z. Yang** and Y. Yang, "Decoupling features in hierarchical propagation for video object segmentation," in *NeurIPS*, 2022.
- 4 **Z. Yang**, Y. Wei, and Y. Yang, "Associating objects with transformers for video object segmentation," in *NeurIPS*, 2021.
- 5 **Z. Yang**, X. Yu, and Y. Yang, "Dsc-posenet: Learning 6dof object pose estimation via dual-scale consistency," in *CVPR*, 2021.
- 6 **Z. Yang**, Y. Wei, and Y. Yang, "Collaborative video object segmentation by foreground-background integration," in *ECCV*, 2020.
- 7 **Z. Yang**, L. Zhu, Y. Wu, and Y. Yang, "Gated channel transformation for visual recognition," in *CVPR*, 2020.
- 8 **Z. Yang**, J. Dong, P. Liu, Y. Yang, and S. Yan, "Very long natural scenery image prediction by outpainting," in *ICCV*, 2019.
- 9 H. Xiong\*, **Z. Yang\***, J. Yu, *et al.*, "Streaming video understanding and multi-round interaction with memory-enhanced knowledge," in *ICLR*, 2025.
- 10 D. Zhou\*, J. Xie\*, **Z. Yang\***, and Y. Yang, "3dis: Depth-driven decoupled instance synthesis for text-to-image generation," in *ICLR*, 2025.
- 11 Z. Zhang, **Z. Yang\***, and Y. Yang, "Sifu: Side-view conditioned implicit function for real-world usable clothed human reconstruction," in *CVPR*, 2024.
- 12 X. Shen, J. Ma, C. Zhou, and **Z. Yang\***, "Controllable 3d face generation with conditional style code diffusion," in *AAAI*, 2024.
- 13 Z. Zhang, J. Xie, Y. Lu, **Z. Yang**, and Y. Yang, "In-context edit: Enabling instructional image editing with in-context generation in large scale diffusion transformer," in *NeurIPS*, 2025.
- 14 D. Zhou, M. Li, **Z. Yang**, and Y. Yang, "Dreamrenderer: Taming multi-instance attribute control in large-scale text-to-image models," in *ICCV*, 2025.
- 15 Y. Xu, **Z. Yang**, and Y. Yang, "Skdream: Controllable multi-view and 3d generation with arbitrary skeletons," in *CVPR*, 2025, pp. 314–325.
- 16 J. Xiu, M. Li, **Z. Yang**, W. Ji, Y. Yin, and R. Zimmermann, "Few-shot incremental learning via foreground aggregation and knowledge transfer for audio-visual semantic segmentation," in *AAAI*, vol. 39, 2025, pp. 8788–8796.
- 17 Z. Zhang, L. Sun, **Z. Yang**, L. Chen, and Y. Yang, "Global-correlated 3d-decoupling transformer for clothed avatar reconstruction," in *NeurIPS*, 2023.
- 18 X. Pan, **Z. Yang**, J. Ma, C. Zhou, and Y. Yang, "Transhuman: A transformer-based human representation for generalizable neural human rendering," in *ICCV*, 2023.

- 19 J. Li, **Z. Yang**, X. Wang, J. Ma, C. Zhou, and Y. Yang, "Jotr: 3d joint contrastive learning with transformers for occluded human mesh recovery," in *ICCV*, 2023.
- 20 Y. Gan, **Z. Yang**, X. Yue, L. Sun, and Y. Yang, "Efficient emotional adaptation for audio-driven talking-head generation," in *ICCV*, 2023.
- 21 Y. Xu, **Z. Yang**, and Y. Yang, "Integrating boxes and masks: A multi-object framework for unified visual tracking and segmentation," in *ICCV*, 2023.
- 22 S. Huang, J. Jia, **Z. Yang**, *et al.*, "Shuffled autoregression for motion interpolation," in *ICASSP*, IEEE, 2023, pp. 1–5.
- 23 X. Shen, **Z. Yang**, X. Wang, J. Ma, C. Zhou, and Y. Yang, "Global-to-local modeling for video-based 3d human pose and shape estimation," in *CVPR*, 2023.
- 24 J. Miao, **Z. Yang**, L. Fan, and Y. Yang, "Fedseg: Class-heterogeneous federated learning for semantic segmentation," in *CVPR*, 2023.
- 25 T. Ma, Y. Sun, **Z. Yang**, and Y. Yang, "Prod: Prompting-to-disentangle domain knowledge for cross-domain few-shot image classification," in *CVPR*, 2023.
- 26 G. Li, Y. Sun, **Z. Yang**, and Y. Yang, "Decompose to generalize: Species-generalized animal pose estimation," in *ICLR*, 2023.
- 27 D. Zhou, **Z. Yang**, and Y. Yang, "Pyramid diffusion models for low-light image enhancement," in *IJCAI*, 2023.
- 28 Y. Xu, **Z. Yang**, and Y. Yang, "Video object segmentation in panoptic wild scenes," in *IJCAI*, 2023.
- 29 S. Huang, **Z. Yang**, L. Li, Y. Yang, and J. Jia, "Avatarfusion: Zero-shot generation of clothing-decoupled 3d avatars using 2d diffusion," in *ACM MM*, 2023.
- 30 F. Zhu, **Z. Yang**, X. Yu, Y. Yang, and Y. Wei, "Instance as identity: A generic online paradigm for video instance segmentation," in *ECCV*, 2022.
- 31 M. Kristan, A. Leonardis, J. Matas, *et al.*, "The tenth visual object tracking vot2022 challenge results," in *ECCV*, 2022.
- 32 Y. Xu, Y. Sun, **Z. Yang**, J. Miao, and Y. Yang, "H2fa r-cnn: Holistic and hierarchical feature alignment for cross-domain weakly supervised object detection," in *CVPR*, 2022.
- 33 X. Pan, P. Li, **Z. Yang**, *et al.*, "In-n-out generative learning for dense unsupervised video segmentation," in *ACM MM*, 2022.

## Technical Reports

- 1 Y. Cheng, L. Li, Y. Xu, *et al.*, "Segment and track anything," in [github.com/z-x-yang/Segment-and-Track-Anything](https://github.com/z-x-yang/Segment-and-Track-Anything) (**Project leader**, **2.4k stars**), 2023.
- 2 **Z. Yang**, P. Li, Q. Feng, Y. Wei, and Y. Yang, "Going deeper into embedding learning for video object segmentation," in *ICCV Workshops*, 2019.
- 3 **Z. Yang**, Y. Ding, Y. Wei, and Y. Yang, "Cfbi+: Collaborative video object segmentation by multi-scale foreground-background integration," in *CVPR Workshops*, 2020.
- 4 **Z. Yang**, J. Zhang, W. Wang, *et al.*, "Towards multi-object association from foreground-background integration," in *CVPR Workshops*, 2021.
- 5 Y. Xu, J. Li, **Z. Yang**, Y. Yang, and Y. Zhuang, "Zju reler submission for epic-kitchen challenge 2023: Trek-150 single object tracking," in *CVPR Workshops*, 2023.
- 6 J. Li, Y. Xu, **Z. Yang**, Y. Yang, and Y. Zhuang, "Zju reler submission for epic-kitchen challenge 2023: Semi-supervised video object segmentation," in *CVPR Workshops*, 2023.
- 7 W. Wang, Y. Sun, **Z. Yang**, and Y. Yang, "Leveraging vision and vision-language models into large-scale product retrieval," in *CVPR Workshops*, 2022.

- 8 C. Liang, Y. Wu, T. Zhou, *et al.*, “Rethinking cross-modal interaction from a top-down perspective for referring video object segmentation,” in *CVPR Workshops*, 2021.
- 9 C. Liang, **Z. Yang**, J. Miao, Y. Wei, and Y. Yang, “Memory aggregated cfbi+ for interactive video object segmentation,” in *CVPR Workshops*, 2020.
- 10 Q. Feng, **Z. Yang**, P. Li, Y. Wei, and Y. Yang, “Dual embedding learning for video instance segmentation,” in *ICCV Workshops*, 2019.

## Awards (Selected)

---

- 2023
  - 📖 [ACM MM 2023] **Best Paper Award**.
  - 📖 [ICCV 2023] **1st** in the VOTS 2023 challenge.
  - 📖 [CVPR 2023] **1st** in Semi-Supervised Video Object Segmentation of EPIC-Kitchens Challenges.
  - 📖 [CVPR 2023] **1st** in TREK-150 Object Tracking of EPIC-Kitchens Challenges.
- 2022
  - 📖 [ECCV 2022] **1st** in the VOT 2022 real-time segmentation tracking challenge.
  - 📖 [ECCV 2022] **1st** in the VOT 2022 short-term segmentation tracking challenge.
  - 📖 [CVPR 2022] **1st** in eBay eProduct Visual Search Challenge.
- 2021
  - 📖 [CVPR 2021] **1st** in the 3rd Large-scale Video Object Segmentation Challenge.
  - 📖 [CVPR 2021] **1st** in the 3rd Large-scale Referring Video Object Segmentation Challenge.
  - 📖 **Outstanding Intern**. Awarded by Baidu Inc.
- 2018
  - 📖 **Guo Moruo Scholarship**. Awarded by USTC (**Highest Scholarship, Top 1.5%**).
  - 📖 Outstanding Graduates. Awarded by USTC.

## Invited Talks (Selected)





---

- 2025
  - 📖 **Cross-modal Controllable Generation by Post-training**  
Valse Webinar, China, Jun 2025
  - 📖 **Cross-modal Controllable Image and 3D Generation**  
Shanghai AI Lab, Shanghai, China, Jan 2025
- 2024
  - 📖 **Object-centric Intelligence in Dynamic Scenes**  
Beijing Jiaotong University, China, Jun 2024
- 2023
  - 📖 **Decoupled Memory of Associating Objects with Transformers**  
1st at the VOTS2023 challenge, ICCV 2023, Paris, France, Oct 2023

## Invited Talks (Selected) (continued)

-  **Segment and Track Anything**  
VALSE Webinar, China, Aug 2023
-  **Digital Human Generation and Reconstruction Driven by Knowledge and Data**  
Zhejiang University, China, Mar 2023
- 2022  **Decoupling Features in Hierarchical Propagation for Video Object Segmentation**  
**Spotlight** presentation, NeurIPS 2022, New Orleans, USA, Nov 2022
-  **Associating Objects with Multi-scale Transformers for Video Object Segmentation**  
**Winner** presentation, VOT2022 Workshop, ECCV 2022, Tel-Aviv, Israel, Oct 2022
-  **Associating Multi-object Segmentation in Video Understanding**  
VALSE Webinar, China, Mar 2022
- 2021  **Towards Multi-Object Association from Foreground-Background Integration**  
**1st** at the 3rd Large-scale Video Object Segmentation Challenge, CVPR 2021, Online, June 2021
-  **Associating Objects with Transformers for Video Object Segmentation**  
**Tutorial:** Data-Efficient Learning in An Imperfect World, CVPR 2021, Online, June 2021
- 2020  **Collaborative Video Object Segmentation by Foreground-background Integration**  
**Spotlight** presentation, ECCV 2020, Online, Aug 2020
-  **Collaborative Video Object Segmentation by Multi-Scale Foreground-Background Integration**  
**2nd** at the 2020 DAVIS Challenge, CVPR 2020, Online, Jun 2020
- 2019  **Going Deeper into Embedding Learning for Video Object Segmentation**  
The 2nd Large-scale Video Object Segmentation Challenge, ICCV 2019, Seoul, Korea, Oct 2019

## Professional Services

- Area Chair  **CVPR 2026**
- Workshop Organizer  **The 4th Pixel-level Video Understanding in the Wild Challenge** at CVPR 2025.  
**The 3rd Pixel-level Video Understanding in the Wild Challenge** at CVPR 2024.  
**The 2nd Pixel-level Video Understanding in the Wild Challenge** at CVPR 2023.
- Journal Reviewer  **TPAMI, IJCV, TIP, TCSVT, TIST, KBS, FCS, etc.**
- Conference Reviewer  **NeurIPS, ICML, CVPR, ICCV, SIGGRAPH Asia, AAAI, IJCAI, etc.**  
**Outstanding Reviewer of ICCV 2023.**