Xingyi Zhou

Meta GenAI
✉ *zhouxy2017@gmail.com*
xingyizhou.xyz
Google Scholar (12k+ citations)
Github (14k+ total stars)

Employment

2025–Present **Meta GenAI**, *Bellevue, WA*,
Research Scientist.
- Multimodal Large Language Models

2022–2025 **Google DeepMind**, *Kirkland, WA*,
Research Scientist.
- Long video understanding
  - Research on Streaming video captioning.
  - Product contribution to AskPhotos in Google Photos.
- Localization in multi-modal large language models
  - Research on Pixel-aligned language models and Dense video object captioning.
  - Product contribution to Vertex Image Segmentation in Google Cloud.

Education

2017–2022 **The University of Texas at Austin**, *Austin, TX*,
Department of Computer Science, Ph.D.
Advisor: Dr. Philipp Krähenbühl
Thesis: Towards unified object recognition in the wild

2013–2017 **Fudan University**, *Shanghai, China*,
School of Computer Science, Bachelor in Science.

Research Experience

2018–2022 **Research Assistant**, *Deep Learning Lab, UT Austin*, Austin, TX,
with Dr. Philipp Krähenbühl.
Object detection; object tracking; general object representation.

2021 **Research Intern**, *Facebook AI Research*, Remote,
with Dr. Ishan Misra.
Large-vocabulary object detection.

2019–2021 **Research Intern**, *Intel Intelligent Systems Lab*, Santa Clara, CA/ Remote,
with Dr. Vladlen Koltun.
Multi-object tracking; object detection on multiple datasets.

2018 **Software Engineering Intern**, *Google Research*, Mountain View, CA,
with Tyler Zhu and Dr. Kevin Murphy.
Human pose estimation; human mesh reconstruction.

2017–2018 **Research Assistant**, *Graphics & AI Lab, UT Austin*, Austin, TX,
with Dr. Qixing Huang.
3D computer vision; pose estimation.

2014–2017 **Research Assistant**, *Institute for Media Computing, Fudan Univ.*, Shanghai, China,
with Dr. Wei Zhang.
Hand pose estimation; fine-grained image classification; decision trees.

2016 **Research Intern**, *Microsoft Research Asia (MSRA)*, Beijing, China,
with Dr. Yichen Wei.
Human/ hand pose estimation.

Selected Publications (Full list at Google Scholar)

2025 **Visual Lexicon: Rich Image Features in Language Space**, *XuDong Wang†, Xingyi Zhou, Alireza Fathi, Trevor Darrell, Cordelia Schmid (†intern hosted at Google)*, arXiv, 2024, [Paper].

2025 **Dense Video Object Captioning from Disjoint Supervision**, Xingyi Zhou***, *Anurag Arnab*, Chen Sun, Cordelia Schmid (* Equal contribution)*, ICLR (spotlight), 2025, [Paper] [Code].

2024 **Streaming Dense Video Captioning**, Xingyi Zhou***, *Anurag Arnab*, Shyamal Buch, Shen Yan, Austin Myers, Xuehan Xiong, Arsha Nagrani, Cordelia Schmid (* Equal contribution)*, CVPR, 2024, [Paper] [Code].

2024 **Pixel Aligned Language Models**, *Jiarui Xu†, Xingyi Zhou, Shen Yan, Xiuye Gu, Anurag Arnab, Chen Sun, Xiaolong Wang, Cordelia Schmid († intern hosted at Google)*, CVPR, 2024, [Paper] [Code].

2023 **How can objects help action recognition?**, Xingyi Zhou, *Anurag Arnab, Chen Sun, Cordelia Schmid*, CVPR, 2023, [Paper] [Code].

2022 **Detecting Twenty-thousand Classes using Image-level Supervision**, Xingyi Zhou, *Rohit Girdhar, Armand Joulin, Philipp Krähenbühl, Ishan Misra*, ECCV, 2022, [Paper] [Code].

2022 **Global Tracking Transformers**, Xingyi Zhou, *Tianwei Yin, Vladlen Koltun, Philipp Krähenbühl*, CVPR, 2022, [Paper][Code].

2022 **Simple multi-dataset detection**, Xingyi Zhou, *Vladlen Koltun, Philipp Krähenbühl*, CVPR, 2022, [Paper] [Code].

2021 **Probabilistic two-stage detection**, Xingyi Zhou, *Vladlen Koltun, Philipp Krähenbühl*, arXiv, 2021, [Paper] [Code].

2021 **Multimodal Virtual Point 3D Detection**, *Tianwei Yin*, Xingyi Zhou, *Philipp Krähenbühl*, NeurIPS, 2021, [Paper] [Code].

2021 **Center-based 3D Object Detection and Tracking**, *Tianwei Yin*, Xingyi Zhou, *Philipp Krähenbühl*, CVPR, 2021, [Paper] [Code].

2020 **Tracking Objects as Points**, Xingyi Zhou, *Vladlen Koltun, Philipp Krähenbühl*, ECCV, 2020 (spotlight), [Paper] [Code].

2019 **Objects as Points**, Xingyi Zhou, *Dequan Wang, Philipp Krähenbühl*, arXiv, 2019, [Paper] [Code].

2019 **Bottom-up Object Detection by Grouping Extreme and Center Points**, Xingyi Zhou, *Jiacheng Zhuo, Philipp Krähenbühl*, CVPR, 2019, [Paper] [Code].

2018 **StarMap for Category-agnostic Viewpoint and Keypoint Estimation**, Xingyi Zhou, *Arjun Karpur, Linjie Luo, Qixing Huang*, ECCV, 2018, [Paper] [Code].

2018 **Unsupervised Domain Adaptation for 3D Keypoint Estimation via View Consistency**, Xingyi Zhou, *Arjun Karpur, Chuang Gan, Linjie Luo, Qixing Huang*, ECCV, 2018, [Paper] [Code].

2017 **Towards 3D Human Pose Estimation in the Wild: A weakly-supervised Approach**, Xingyi Zhou, *Qixing Huang, Xiao Sun, Xiangyang Xue, Yichen Wei*, ICCV, 2017, [Paper] [Code].

2016 **Deep Kinematic Pose Regression**, Xingyi Zhou, *Xiao Sun, Wei Zhang, Shuang Liang, Yichen Wei*, ECCV workshops, 2016, [Paper] [Code].

2016 **Model-based Deep Hand Pose Estimation**, Xingyi Zhou, *Qingfu Wan, Wei Zhang, Xiangyang Xue, Yichen Wei*, IJCAI, 2016, [Paper] [Code].

Horners

Apr 2021 **Facebook Fellowship**.

Jan 2020 **Facebook Fellowship Finalist**.

Nov 2019 **Adobe Fellowship Finalist**.

Sep 2017 **Provost's Graduate Excellence Fellowship**, *UT Austin*.

Jun 2017 **Shanghai Outstanding Graduate**.

Aug 2016 **Award of Excellence**, *Stars of Tomorrow Internship Program, Microsoft Research Asia*.

Competition Awards

June 2021 **2nd place**, *CVPR 2021 Waymo real-time 3D detection challenge*.

Aug 2020 **1st place**, *ECCV 2020 Robust vision challenge, object detection track*.

Aug 2020 **1st place**, *ECCV 2020 Robust vision challenge, instance segmentation track*.

Jun 2020 **2nd place/ Best student submission**, *ICRA 2020 nuScenes 3D detection challenge*.

Dec 2014 **Gold medal**, *ACM International Collegiate Programming Contest (**ACM-ICPC**) Regional*.

Nov 2012 **First prize**, *National Olympiad in Informatics in Provinces (**NOIP**), Zhejiang Provinces*.

Presentations

Oct 2023 **Recognizing objects in long time and in a large-vocabulary**, *ECCV 2022 MOTComplex workshop*.

Jan 2021 **Center-based 3D object detection and tracking**, *VALSE Webinar [Video]*.

Aug 2020 **Learning a unified label space for robust object detection**, *ECCV 2020 Robust vision challenge workshop [Video]*.

Aug 2020 **CenterNet2**, *ECCV 2020 Joint COCO and LVIS Challenge workshop [Slides]*.

May 2019 **Generalized keypoint estimation**, *UT research prepaertion exam [Slides]*.

Mar 2018 **Modeling geometry in pose estimation**, *seminar presentation at UT Austin [Slides]*.

Jan 2018 **Towards 3D human pose estimation in the wild**, *GAMES Webinar [Video]*.

Professional Services

Conference Reviewer, *ECCV(2018-), CVPR(2019-), ICCV(2019-), ICLR(2020-), NeurIPS(2020-). Outstanding reviewer in CVPR2019, ECCV2020, CVPR2021, CVPR2024.*

Journal reviewer, TPAMI, TIP, IJCV, CVIU.

Teaching assistance, Neural Networks, Fall 2019.

Web master and student host, UT Forum for Artificial Intelligence.

Open-source softwares, CenterNet (7k+ stars), CenterTrack (2k+ stars), Detic (1k+ stars), ExtremeNet (1k+ stars), CenterNet2 (1k+ stars), pytorch-pose-hg-3d (500+ stars).

Supervising experiences

2024	XuDong Wang, intern at Google DeepMind, PhD student from UC Berkeley
2023	Jiarui Xu, intern at Google Research, PhD student from UC San Diego
2020	Tianwei Yin, undergraduate student from UT Austin
2018	Arjun Karpur, master student from UT Austin

Skills

Languages	Python, C/ C++, Lua
Frameworks	Jax (Codebases from scratch for Vision-Language Models, Segment Anything, and Object detection), PyTorch, Tensorflow, Torch, Caffe