

# SIJIN CHEN

Fudan University | H: +86 187 1792 9716 | [csjch3cook@gmail.com](mailto:csjch3cook@gmail.com)  
[Homepage](#) | [GitHub](#) | [Google Scholar](#)

## ABOUT ME

I am currently a final-year Master's student in Artificial Intelligence at Fudan University (Sep. 2021 - Jun. 2024, expected), where Prof. [Tao Chen](#) is my advisor. I am also fortunate to work closely with Dr. [Hongyuan Zhu](#) from A\*STAR, Singapore, and Dr. [Gang Yu](#), Dr. [Xin Chen](#), and Dr. [Chi Zhang](#) from Tencent. Before this, I obtained my Bachelor's degree in Data Science and Big Data Technology also from Fudan University (Sep. 2017 - Jun. 2021).

My long-term research goal is to develop vision-language systems that possess the capacity to comprehend, reason, and envision the physical world. Outside my research, I love sports and music.

## RESEARCH INTERESTS

Multi-modal Learning, Vision and Language, Large Language Models, and 3D Generation.

## EDUCATION

<b>Masters in Artificial Intelligence</b> (GPA 3.56/4.00) Fudan University. Supervised by Prof. <a href="#">Tao Chen</a> .	Sep. 2021 - Jun. 2024 Shanghai, China
<b>Bachelor in Data Science and Big Data Technology</b> Fudan University.	Sep. 2017 - Jun. 2021 Shanghai, China

## PUBLICATIONS AND PRE-PRINTS ([GOOGLE SCHOLAR](#))

- LL3DA: Visual Interactive Instruction Tuning for Omni-3D Understanding, Reasoning, and Planning.  
[Sijin Chen](#), Xin Chen, Chi Zhang, Mingsheng Li, Gang Yu, Hao Fei, Hongyuan Zhu, Jiayuan Fan, Tao Chen.  
[[CVPR 2024](#) | [project](#) | [paper](#) | [github](#)]  
[Summary]: 3D-LLMs respond to visual and text interactions in complex 3D scenes.
- Vote2Cap-DETR++: Decoupling Localization and Describing for End-to-End 3D Dense Captioning.  
[Sijin Chen](#), Hongyuan Zhu, Mingsheng Li, Xin Chen, Peng Guo, Yinjie Lei, Gang Yu, Taihao Li, Tao Chen.  
[[Submitted to T-PAMI \(minor revision\)](#) | [paper](#) | [github](#)]  
[Summary]: Decoupled feature extraction for localizing and describing objects in 3D scenes.
- End-to-End 3D Dense Captioning with Vote2Cap-DETR.  
[Sijin Chen](#), Hongyuan Zhu, Xin Chen, Yinjie Lei, Gang Yu, Tao Chen.  
[[CVPR 2023](#) | [paper](#) | [github](#) | [youtube](#)]  
[Summary]: Addressing 3D dense captioning as a set prediction problem with parallel decoding.
- M3DBench: Let's Instruct Large Models with Multi-modal 3D Prompts.  
Mingsheng Li, Xin Chen, Chi Zhang, [Sijin Chen](#), Hongyuan Zhu, Fukun Yin, Gang Yu, Tao Chen.  
[[Under Review](#) | [project](#) | [paper](#) | [github](#)]  
[Summary]: A large scale dataset querying 3D LLMs with text, 2D, and 3D prompts.
- WI3D: Weakly Incremental 3D Detection via Visual Prompts.  
Mingsheng Li, [Sijin Chen](#), Shengji Tang, Hongyuan Zhu, Xin Chen, Fukun Yin, Tao Chen.  
[[Under Review](#) | [paper](#)]  
[Summary]: Introducing new categories to 3D detectors with 2D foundation models.

## PROJECTS

- **Language Modelling for 3D Mesh Generation.** Jan. 2024 - Jun. 2024  
Tuning LLMs for auto-regressive triangular mesh generation.
- **Language for 3D Scenes.** Aug. 2021 - Mar. 2024  
Proposed a set-to-set solution for localizing and describing objects in 3D scenes, accepted to [CVPR 2023](#) and won the Scan2Cap challenge at [ICCV 2023](#). Submitted an advanced **Vote2Cap-DETR++** to [T-PAMI with minor revision](#). Presented **LL3DA**, a large language 3D assistant responding to both text and visual interactions with complex 3D environments, accepted to [CVPR 2024](#). Put forward **M3DBench**,

a large-scale 3D vision language dataset covering 327k lines of annotations for 10 tasks related to 3D perception, understanding, reasoning, and planning, [under review](#).

- **Class-Incremental 3D Detection.** Apr. 2023 - Dec. 2023  
Proposed WI3D, introducing new categories to 3D detectors using images, [under review](#).
- **Earlier Projects.** Before Sep. 2021
  - Self-Supervised Pre-training on 3D Point Clouds.** Developed a self-supervised learning algorithm that learns global- and patch-level contrastive representations for 3D point clouds.
  - A Smart Advertisement Display System.** Developed a human perception system that detects faces, recognizes facial expressions, estimates eye gaze, age, and gender for advertisement recommendation.

## SCHOLARSHIPS AND AWARDS

---

First place winner of the Scan2Cap Challenge at <a href="#">ICCV 2023</a> .	Oct. 2023
National Scholarship (rank 1/46).	Sep. 2023
Second prize of the Scholarship for Outstanding Students at Fudan University.	Sep. 2022
Award for the Scholarship for Outstanding Students at Fudan University.	Sep. 2021
Second prize of the Scholarship for Outstanding Students at Fudan University.	Jun. 2021

## RESEARCH INTERN

---

**Tencent.** Jan. 2024 - Jun. 2024 expected  
**Research Intern.** Supervised by Dr. [Xin Chen](#) and Dr. [Gang Yu](#), working on 3D generation.

## INVITED TALKS

---

Winner presentation of <b>the Scan2Cap Challenge</b> at <a href="#">ICCV 2023</a> . [ <a href="#">slides</a>   <a href="#">youtube</a> ]	Oct. 2023
Paper presentation at <b>the workshop for advances in 3D vision</b> , <a href="#">VALSE 2023</a> .	Apr. 2023

## SKILLS

---

Languages:	Chinese (native), English (proficient), Shanghai dialect
Programming:	Python, R, C, Matlab, SQL
Tools:	PyTorch, Blender, Visual Studio, Spyder, Jupyter Notebook

## REFERENCES

---

Prof. <a href="#">Tao Chen</a>	Advisor	FUDAN UNIVERSITY	<a href="mailto:eetchen@fudan.edu.cn">eetchen@fudan.edu.cn</a>
--------------------------------	---------	------------------	--