

YIZHI (DAVID) SONG

☎ 765-430-3961

✉ songcx2211@gmail.com

🌐 [linkedin.com/yizhi](https://www.linkedin.com/in/yizhi)

🏠 [My personal website](#)

Professional

Research Scientist at Intelligent Creation, Bytedance/Tiktok

Aug. 2025 – Present

Developing unified image generation and editing models

Bellevue, WA

Education

CGV Lab, Department of Computer Science, Purdue University

Sep. 2019 – Apr. 2025

Ph.D. in Computer Science - Generative AI

West Lafayette, IN

College of Computer Science, Zhejiang University

Sep. 2015 – Jul. 2019

B.E. in Computer Science & Technology

Hangzhou, China

Publications and Patents [[Google Scholar](#)]

- Tang, Y., ..., **Song, Y.**, ... (2025). Caption Anything in Video: Fine-grained Object-centric Captioning via Spatiotemporal Multimodal Prompting ([Project Page](#)).
- Hua, H.*, Zeng, Z.*, **Song, Y.***, ... (2025). MMIG-Bench: Towards Comprehensive and Explainable Evaluation of Multi-Modal Image Generation Models. *NeurIPS D&B 2025* ([Project Page](#)).
- He, L., Zeng, X., Chen, A., **Song, Y.**, ... (2024). Advancing Vision Language Models by Large-scale Synthetic Dataset Generation (under review).
- Xiong, Z., Xiong, W., Shi, J., Zhang, H., **Song, Y.**, ... (2024). GroundingBooth: Grounding Text-to-Image Customization ([Project Page](#)).
- Song, Y.**, He, L., ... & Aliaga, D. (2024). Refine-by-Align: Reference-Guided Artifacts Refinement through Semantic Alignment. *ICLR 2025* ([Project Page](#)).
- He, L., **Song, Y.**, ... (2024). Kubrick: Multimodal Agent Collaborations for Video Generation. *CVPR 2025 AI4CC* ([Project Page](#)).
- Tarrés, G. C., Lin, Z., Zhang, Z., Zhang, J., **Song, Y.**, ... & Kim, S. Y. (2024). Thinking Outside the BBox: Unconstrained Generative Object Compositing. *ECCV 2024* ([PDF](#)).
- Song, Y.**, Zhang, Z., ... & Aliaga, D. (2024). IMPRINT: Generative Object Compositing by Learning Identity-Preserving Representation. *CVPR 2024* ([Project Page](#)) ([Productized](#) and appeared in [Adobe Max Sneak](#)).
- Song, Y.**, Zhang, Z., Lin, Z., Cohen, S., Price, B., ... & Aliaga, D. (2023). ObjectStitch: Object Compositing With Diffusion Model. *CVPR 2023* ([PDF](#)) ([Reposted by AK](#)).
- Song, Y.**, Fan, R., Huang, S., Zhu, Z., & Tong, R. (2019). A Three-stage Real-time Detector for Traffic Signs in Large Panoramas. *CVM 2019 oral* ([PDF](#)).
- Song, Y.**, Zhang, Z., ... & Kim, S. Y. Systems and Methods for Image Compositing. *US Patent*: US20250022099A1.

Working & Internship Experiences

Object-Centric Image Editing with MLLM & Diffusion

Adobe Research, Jun. 2024 – Aug. 2024

Research Scientist Intern

San Jose, CA

- Design an image editing model to move/insert/remove objects following captions, leveraging **VLM**'s reasoning ability.
- Trained a 5B **DiT Diffusion** using distributed training as an image editing engine, which preserves object identity.
- Collect a paired object-centric image editing dataset with captions describing compositionality and object relationship.

ObjectStitchv2: Image Editing with ID-Preserving Representation

Adobe, May 2023 – Aug. 2023

Research Scientist Intern

San Jose, CA

- Jointly trained **DINOv2** and **Diffusion** for **ID-preserving representation**, greatly improved detail preservation.
- Improved self-supervised training by using large scale multi-view datasets and introducing harmonization augmentation.
- Introduced shape-guided generation, allowing edits such as **novel view synthesis** and **non-rigid transformations**.

ObjectStitch: Generative Object Compositing with Diffusion

Adobe, Jun. 2022 – Sep. 2022

Research Scientist Intern

Remote

- Developed the **first diffusion model-based** unified framework for generative object compositing that handles view synthesis, geometry correction, harmonization and shadow generation at the same time while preserving appearance.
- Designed a content adaptor based on **ViT** and **CLIP** that produces multi-modal embedding from the inputs.
- Proposed a fully **self-supervised** training scheme without any manual annotations and data augmentation techniques.

Depth-Based Image Inpainting

Qualcomm, Inc., Jun. 2021 – Aug. 2021

Interim Engineering Intern

Remote

- Developed a scene **depth-aware inpainting** model, and integrated it in an interactive **image editing application**.

- The application supported zooming and moving of various foreground objects while filling the revealed **irregular holes**.
- Designed a new training scheme, generated a **synthetic RGBD dataset** to train the network with **partial conv**.
- The trained model **outperformed** the traditional inpainting models on RGB-D images captured by mobile phone.

Real-time Traffic Sign Detection

Tsinghua University, Aug. 2018 – Sep. 2018

Instructor: Prof. Shimin Hu

Beijing, China

- Proposed a novel traffic sign detection framework (based on **Faster RCNN**) for autonomous driving which achieved both the fastest speed (more than **100fps**) and state-of-the-art detection accuracy (**0.92**) on TT100k benchmark.

Technical Skills

DiT, MLLM, Multi-node distributed training, Pytorch, Diffusers, OpenCV, OpenGL, Git, Qt, Linux, Python, C, C++.