

# Yizhe Zhu

Email: [zyzfflying@gmail.com](mailto:zyzfflying@gmail.com)

Linkedin: [yizhe-ethan-zhu](https://www.linkedin.com/in/yizhe-ethan-zhu)

Homepage: <https://yzzhu0.github.io>

## SUMMARY

---

- I am a leading researcher in Diffusion Image/Video Synthesis at Bytedance. My work spans from large-scale text-guided image/video generation and editing models to innovative AIGC filters, which I have successfully integrated into AIGC products on TikTok and Douyin.
- With 8+ years of combined research and engineering experience, I possess deep expertise in generative models, computer vision, and machine learning. I have in-depth knowledge of multi-modality issues (language and vision), disentangled representation learning, and self-supervised learning.

## WORKING EXPERIENCE

---

### ByteDance, Intelligent Creation Lab

Senior Research Scientist

Feb. 2022 - present

Research Scientist

Aug. 2020 - Feb. 2022

#### · **Image personalization:**

a) Developed a pioneering finetune-free human-portrait personalization model that turned into viral filters: [AI portrait](#) (18M videos created), [AI Xianxia](#) (4M videos created).

b) Developed [MoMA](#), a SOTA general-object personalization model. As a tuning-free plug-and-play module, MoMA requires only a single reference image and generates images with high detail fidelity, enhanced identity preservation and prompt faithfulness by leveraging **Multimodal LLM**.

#### · **Image Generation:**

Developed a pioneering large-scale text-to-image model, trained on 1000+ GPUs using billions of text-image pairs. The improved version was integrated in [DouBao AI Agent](#).

#### · **Video Generation:**

a) Developed a landmark-driven face animation algorithm that accurately animates static face images. It turned into a viral effect [“Mayiyahei”](#) (12M videos created).

b) Developed [MagicPose](#) a diffusion-based model for 2D human pose and facial expression retargeting, with two disentangled ControlNets to enhance the pose accuracy and identity-preservation.

c) Developed [MagicVideo](#), an efficient text-to-video generation framework based on latent diffusion models, designed to generate high-resolution photo-realistic videos.

### NEC Labs America, Inc.

Jun. 2019 - Dec. 2019

Research Intern

- Invented a Controllable Video Generation model with a Sequential VAE by disentangling appearance and motion representation in a self-supervised manner.

### Hikvision Research Institute

Jun. 2018 - Sep. 2018

Research Intern

- Developed a conditional latent variable model to generate desirable image features based on class-level attributes, and an alternating back-propagation algorithm to optimize the model.
- Developed a multi-attention localization model for the object-part detection in a weakly-supervised manner, and provided enhanced visual features for Zero-Shot Learning.

## EDUCATION

---

### Rutgers University

Ph.D. in Computer Science

Specialties: Computer Vision, Generative model, Machine Learning

### University of Missouri

Master of Science in Electronic & Computer Engineering

Specialties: Computer Vision, Image Compression

### Shanghai University

Bachelor of Science in Electronics & Communication Engineering

Jan. 2015 - May 2020

Advisor: Prof. Ahmed Elgammal

Sep. 2013 - Dec. 2014

Advisor: Prof. Zhihai He

Sep. 2009 - Aug. 2013

## SELECTED RESEARCH PUBLICATION

---

According to [Google Scholar](#), I have authored 19 peer-reviewed publications with a total of 1800+ citations and an h-index of 16.

1. Kunpeng Song, **Yizhe Zhu**, Bingchen Liu, Qing Yan, Xiao Yang, Ahmed Elgammal. MoMA: Multimodal LLM Adapter for Fast Personalized Image Generation. (ECCV under review) 2024 [[Project Page](#)]
2. Di Chang, Yichun Shi, Quankai Gao, Jessica Fu, Hongyi Xu, Guoxian Song, Qing Yan, **Yizhe Zhu**, Xiao Yang, Mohammad Soleymani. MagicPose: Realistic Human Poses and Facial Expressions Retargeting with Identity-aware Diffusion. (arXiv:2311.12052) 2023 [[Project Page](#)]
3. Daquan Zhou, Weimin Wang, Hanshu Yan, Weiwei Lv, **Yizhe Zhu**, Jiashi Feng. MagicVideo: Efficient Video Generation With Latent Diffusion Models. (arXiv:2211.11018) 2022 [[Project Page](#)]
4. Yufan Zhou, Bingchen Liu, **Yizhe Zhu**, Xiao Yang, Changyou Chen, Jinhui Xu. Shifted Diffusion for Text-to-image Generation. *International Conference on Computer Vision and Pattern Recognition (CVPR)* 2022
5. Bingchen Liu, **Yizhe Zhu**, Kunpeng Song, Ahmed Elgammal. Towards Faster and Stabilized GAN Training for High-fidelity Few-shot Image Synthesis. *International Conference on Learning Representations (ICLR)* 2021
6. **Yizhe Zhu**, Martin Renqiang Min, Asim Kadav, Hans Peter Graf. S3VAE: Self-Supervised Sequential VAE for Representation Disentanglement and Data Generation. *International Conference on Computer Vision and Pattern Recognition (CVPR)* 2020
7. Xingchao Peng, Zijun Huang, **Yizhe Zhu**, Kate Saenko. Federated Adversarial Domain Adaptation. *International Conference on Learning Representations (ICLR)* 2020
8. Bingchen Liu, **Yizhe Zhu**, Zuohui Fu, Gerard de Melo, Ahmed Elgammal. OOGAN: Disentangling GAN with One-Hot Sampling and Orthogonal Regularization. (**AAAI**) 2020
9. **Yizhe Zhu**, Jianwen Xie, Bingchen Liu, Ahmed Elgammal. Learning Feature-to-Feature Translator by Alternating Back-Propagation for Zero-Shot Learning. *International Conference on Computer Vision (ICCV)* 2019
10. **Yizhe Zhu**, Jianwen Xie, Zhiqiang Tang, Xi Peng, Ahmed Elgammal. Semantic-Guided Multi-Attention Localization for Zero-Shot Learning. *Neural Information Processing Systems (NeurIPS)* 2019
11. Zhiqiang Tang, Xi Peng, Tingfeng Li, **Yizhe Zhu**, Dimitris N Metaxas. AdaTransform: Adaptive Data Transformation. *International Conference on Computer Vision (ICCV Oral)* 2019
12. **Yizhe Zhu**, Mohamed Elhoseiny, Bingchen Liu, Ahmed Elgammal. A Generative Adversarial Approach for Zero-Shot Learning from Noisy Texts. *International Conference on Computer Vision and Pattern Recognition (CVPR)* 2018.
13. Zhiqiang Tang, Xi Peng, Shijie Geng, **Yizhe Zhu**, Dimitris Metaxas. CU-Net: Coupled U-Nets. *British Machine Vision Conference (BMVC Oral)* 2018.

14. **Yizhe Zhu**, Ahmed Elgammal. A Multilayer-Based Framework for Online Background Subtraction with Freely Moving Cameras. *International Conference on Computer Vision (ICCV)* 2017
15. Mohamed Elhoseiny\*, **Yizhe Zhu\***, Han Zhang, Ahmed Elgammal. Link the head to the beak: Zero Shot Learning from Noisy Text Description at Part Precision. *International Conference on Computer Vision and Pattern Recognition (CVPR)* 2017 (\* means Co-first authors)

## ACADEMIC SERVICES

---

Conference Reviewer: CVPR, ICCV, ECCV, ICLR, NeurIPS, AAAI	2018 ~ present
Journal Reviewer: TPAMI, IJCV	2018 ~ present

## AWARDS, GRANTS, & HONORS

---

Student Travel Grant, US National Science Foundation (NSF)	2017 ~ 2021
Research Assistant Scholarship	2018 ~ 2020
Teaching Assistant Scholarship	2016 ~ 2017
Excellent Student Award in SHU	2012
Recognition Award from Shanghai Innovation Experiment Program for University Students	2012
National Scholarship of University, Department of Education of China	2011 ~ 2012