JING HE

🌐 jingheya.github.io

📞 +86-185-7171-3207 | ✉ jingheyaa@gmail.com | ⑂ jingheya | 🐦 Jingheya

EDUCATION

Hong Kong University of Science and Technology 09/2023 ~ 06/2026 (expected)
Information Hub, Guangzhou Campus *Guangzhou, China*
- **Doctor of Philosophy in Artificial Intelligence**
- Supervisor: Prof. Ying-Cong Chen
- Research Topic: **Visual Generative Models** (**Customized Image Generation**, **Diffusion for Perception**)

Xiamen University 09/2020 ~ 06/2023
School of Informatics *Xiamen, China*
- **Master of Engineering in Artificial Intelligence**
- Supervisor: Prof. Rongrong Ji
- Research Topic: **Visual Generative Models** (**Image Generation**)

Wuhan Institute of Technology 09/2016 ~ 06/2020
School of Computer Science and Engineering *Wuhan, China*
- **Bachelor of Engineering in Digital Media Technology**
- Grades: 3.62/4.0, Rank: 2/87

RESEARCH (*EQUAL CONTRIBUTION, ORDER RANDOMIZED.)

LOTUS: Diffusion-based Visual Foundation Model for High-quality Dense Prediction (620+ Stars) 09/2024
ICLR 2025 | Project page: `lotus3d.github.io` *Guangzhou*
Jing He*, Haodong Li*, Wei Yin, Yixun Liang, Leheng Li, Kaiqiang Zhou, Hongbo Zhang, Bingbing Liu, Ying-Cong Chen
TL;DR: Based on Stable Diffusion, Lotus delivers SoTA performance on monocular depth & normal estimation with a simple yet effective fine-tuning protocol that better fits the pre-trained visual prior for dense prediction.

DisEnvisioner: Disentangled and Enriched Visual Prompt for Customized Image Generation (110+ Stars) 04/2024
ICLR 2025 | Project page: `disenvisioner.github.io` *Guangzhou*
Jing He*, Haodong Li*, Yongzhe Hu, Guibao Shen, Yingjie Cai, Weichao Qiu, and Ying-Cong Chen
TL;DR: DisEnvisioner effectively identifies and enhances the subject-essential features while filtering out other irrelevant ones, enabling exceptional image customization in a tuning-free manner with only a single image.

PixelFolder: An efficient progressive pixel synthesis network for image generation 03/2022
ECCV 2022 | Project page: `github.com/jingheya/PixelFolder` *Xiamen, Shanghai*
Jing He, Yiyi Zhou, Qi Zhang, Jun Peng, Yunhang Shen, Xiaoshuai Sun, Chao Chen, Rongrong Ji
TL;DR: PixelFolde is a progressive pixel synthesis network designed to improve image generation efficiency by utilizing a multi-stage approach and innovative pixel folding operations, while achieving state-of-the-art image quality.

OmniBooth: Learning Latent Control for Image Synthesis with Multi-modal Instruction 10/2024
Arxiv 2024 | Project page: `len-li.github.io/omnibooth-web/` *Guangzhou*
Leheng Li, Weichao Qiu, Xu Yan, Jing He, Kaiqiang Zhou, Yingjie Cai, Qing Lian, Bingbing Liu, Ying-Cong Chen
TL;DR: OmniBooth is a framework for flexible image synthesis using multi-modal instructions for spatial control and customization, leveraging a novel latent control signal for integration.

Towards Efficient Diffusion-Based Image Editing with Instant Attention Masks   08/2023

AAAI 2024 | Project page: `github.com/xiaotianqing/InstDiffEdit`   *Xiamen*

*Siyu Zou, Jiji Tang, Yiyi Zhou, **Jing He**, Chaoyi Zhao, Rongsheng Zhang, Zhipeng Hu, Xiaoshuai Sun*

TL;DR: InstDiffEdit is an efficient diffusion-based image editing method that utilizes cross-modal attention in diffusion models to generate instant masks for image editing.

Learning dynamic prior knowledge for text-to-face pixel synthesis   04/2022

ACM MM 2022 | Project page: `github.com/pengjunn/PixelFace`   *Xiamen*

*Jun Peng, Xiaoxiong Du, Yiyi Zhou, **Jing He**, Yunhang Shen, Xiaoshuai Sun, Rongrong Ji*

TL;DR: PixelFace is a dynamic pixel synthesis network for text-to-face (T2F) generation, which uses dynamic knowledge embeddings derived from text descriptions to guide pixel synthesis.

Towards open-ended text-to-face generation, combination and manipulation   04/2022

ACM MM 2022 | Project page: `github.com/pengjunn/OpenFace`   *Xiamen*

*Jun Peng, Han Pan, Yiyi Zhou, **Jing He**, Xiaoshuai Sun, Yan Wang, Yongjian Wu, Rongrong Ji*

TL;DR: OpenFaceGAN is an end-to-end network for open-ended text-to-face generation, combination, and manipulation, which directly maps text descriptions to faces using a multi-modal latent space.

INTERNSHIP

Tencent Youtu Lab   08/2021 ~ 01/2022

Research Intern   *Shanghai, China*

- Mentors: Chao Chen, Qi Zhang
- Research Topic: **Image Generation**

HONORS

- HKUST(GZ) Postgraduate Student Fellowship
- WIT First-class Scholarship
- WIT Outstanding graduate
- WIT Outstanding Student Leader