Yue **Gao**

Researcher · Microsoft Research

No. 5 Dan Ling Street, Haidian District, Beijing, China

(+86) 188-0018-5532 | yuegao@microsoft.com | yuegao.me | hologerry | Citations: 250+

Summary

Yue is a researcher currently working at the Media Computing Group at Microsoft Research Asia. With a focus on computer vision, Yue is particularly interested in generative models, video processing, and self-supervised learning. As the first author or co-first author, Yue has published several papers on top conferences such as SIGGRAPH/SIGGRAPH Asia, NeurIPS, and CVPR. Yue has also contributed to Microsoft Teams projects for conference meeting scenarios. With 5 years of experience, Yue has a deep understanding of the latest techniques and technologies and is able to quickly and efficiently apply the knowledge to solve complex problems.

Work Experience

Microsoft Research — *Beijing, China*

Researcher — *July 2021 - Present*

- Projects: Head Pose and Expression Freely Controllable Talking Head Video Generation; Sign Language Understanding.

Microsoft Research — *Beijing, China*

Research Intern — *June 2020 - June 2021*

- Projects: Self-Supervised Learning for Object Detection; High Fidelity and Arbitrary Face Image Synthesis.

Apple — *Beijing, China*

Research Intern — *Dec. 2019 - May 2020*

- Project: Invoice Verification System.

Education

Peking University — *School of Electronics Engineering and Computer Science*

Master of Natural Science in Computer Applied Technology — *Sep. 2018 - July 2021*

- Graduate with honors: Outstanding Graduates, Merit Student, Outstanding Student of Wangxuan Institute of Computer Technology.
- Thesis: Attributes Controllable Image Synthesis; Supervisors: Prof. Zhouhui Lian.

Beijing Institute of Technology — *School of Computer Science*

Bachelor of Science in Internet of Things Engineering — *Sep. 2014 - June 2018*

- Graduate with honors: 5 Times of Study Excellence Scholarship.

Projects

Freely Controllable Talking Head Video Generation (CVPR 2023) — *Project Page*

Researcher at Microsoft Research — *Apr. 2022 - Present*

- Talking head generation has quality and controllability limitations despite generating videos based on source identity and target motion.
- Our PECHead method generates high-fidelity face reenactment and talking head videos, enabling free control over head pose and expression through head movement-based landmark estimation.
- PECHead uses learned and predefined face landmarks for global and local motion estimation, improving image synthesis with the a novel alignment module. And the video-based pipeline further enhances generated video smoothness and naturalness.
- This project is continuously ongoing. We are currently integrating the diffusion models, and trying to expand the scope to more scenarios, such as appearance manipulation based on identity.
- Responsible for almost all the research work, including method design, implementation, and experiments.

Sign Language Understanding

Researcher at Microsoft Research — *Dec. 2022 - Present*

- Keypoints have limitations in representing sign language sequences, making it difficult to represent action sequences and sensitive to noise. We aim to propose a better representation to aid in sign language understanding.
- We propose using SMPL-X parameters to represent sign language sequences, including pose and shape parameters of signers. This representation satisfies human prior constraints and better represents sign language.
- Based on this 3D representation, we propose some natural data augmentation techniques that further enhance the performance of sign language recognition tasks.
- Quantitative experiments show that our proposed sign language representation outperforms the existing keypoint-based representation in downstream sign language recognition tasks.

Self-Supervised Learning for Object Detection (NeurIPS 2021 Spotlight) — *Project Page*

Research Intern at Microsoft Research — *Dec. 2020 - May 2021*

- Image-level contrastive representation learning is effective for transfer learning. Our method aligns object-level representations with selective search bounding boxes as object proposals, uses the same network architecture as the detection pipeline, and has object detection properties such as invariance to translation and scale. I am responsible for implementing all the code and experiments and part of writing.

High Fidelity and Arbitrary Face Image Synthesis (CVPR 2021)

Project Page

Research Intern at Microsoft Research — *June 2020 - Nov. 2020*

- Based on GANs, we propose a novel wavelet-based face editing framework, HifaFace, for high-fidelity and arbitrary face editing, which can also be applied to face related tasks as a data augmentation method.
- This work utilized the large-scale unlabeled dataset FFHQ through semi-supervised learning, and proposed a novel loss to edit faces in a continuous space at any degree.

Invoice Verification System

Research Intern at Apple — *Dec. 2019 - May 2020*

- Our Invoice Verification System uses object detection and recognition to determine the legality of uploaded invoices. This system aims for faster inference speed by employing optimized model design and quantization techniques. Additionally, a website has been developed to showcase the system online.

Semantic Attributes based Font Generation (SIGGRAPH 2020)

Project Page

Master's student at Peking University — *Aug. 2019 - Mar. 2020*

- The system establishes a mapping relationship from semantic attributes to glyph images to inspire new font designs and enable ordinary users to create fonts.
- The generative model uses semantic attribute values to synthesize fonts, and a semi-supervised learning scheme and the attribute attention mechanism are designed to significantly improve the quality of the generated glyph images and the ability of font style transfer.
- The responsibilities include designing the framework and core module, coding, conducting experiments and data processing, and developing the online demonstration system.

Artistic Glyph Image Synthesis (SIGGRAPH Asia 2019)

Project Page

Master's student at Peking University — *Sep. 2018 - July 2019*

- The purpose of this work is to reduce the repetitive work of font designers. It is the first time to realize the generation of artistic fonts based on a small number of samples in any language.
- This method is based on GAN and style transfer, using the target character and artistic font style reference set as input to generate stylized target characters. The novel local texture refinement loss improves local texture details. New large-scale Chinese dataset is constructed to promote Chinese special effect font generation field.

Publications

1. High-Fidelity and Freely Controllable Talking Head Video Generation
   Yue Gao, Yuan Zhou, Jinglu Wang, Xiao Li, Xiang Ming, Yan Lu
   Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2023
2. Learning to Prompt for Open-Vocabulary Object Detection With Vision-Language Model
   Yu Du, Fangyun Wei, Zihe Zhang, Miaojing Shi, ***Yue Gao***, Guoqi Li
   Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2022
3. Aligning Pretraining for Detection via Object-Level Contrastive Learning
   Fangyun Wei*, ***Yue Gao****, Zhirong Wu, Han Hu, Stephen Lin (* equal contribution)
   Proceedings of the Advances in Neural Information Processing Systems (NeurIPS, spotlight), 2021
4. High-Fidelity and Arbitrary Face Editing
   Yue Gao, Fangyun Wei, Jianmin Bao, Shuyang Gu, Dong Chen, Fang Wen, Zhouhui Lian
   Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR), 2021
5. Attribute2Font: Creating Fonts You Want From Attributes
   Yizhi Wang*, ***Yue Gao****, Zhouhui Lian (* equal contribution)
   ACM Transactions on Graphics (TOG, SIGGRAPH Technical Paper), 2020
6. Artistic Glyph Image Synthesis via One-Stage Few-Shot Learning
   Yue Gao*, Yuan Guo*, Zhouhui Lian, Yingmin Tang, Jianguo Xiao (* equal contribution)
   ACM Transactions on Graphics (TOG, SIGGRAPH Asia Technical Paper), 2019
7. *Patent*: An Automatic Generation Method of Artistic Font Based on One-Stage Few-Shot Learning
   Zhouhui Lian, ***Yue Gao***, Yuan Guo, Yingmin Tang, Jianguo Xiao
   China Patent, CN.201910670478.8

Honors & Awards

June 2021 **Star of Tomorrow**, Microsoft Research Asia Internship Program

June 2021 **Outstanding Graduates**, Peking University

Dec. 2020 **Merit Student**, Peking University

Dec. 2019 **Outstanding Student**, Wangxuan Institute of Computer Technology of Peking University

2014 - 2018 **Study Excellence Scholarships (5 times)**, Beijing Institute of Technology

Skills

Programming Languages C/C++, Python, Bash, HTML, CSS, JavaScript, Java

Frameworks PyTorch, NumPy, OpenCV, TensorFlow, Keras

Tools Git, Docker, LaTeX, Visual Studio, VS Code, Xcode, Blender, Matlab