

# Yuren Cong

COMPUTER VISION · MACHINE LEARNING

Balatonstr. B2.70, 10319, Berlin, Germany

☎ (+49) 15257600297 | ✉ congyuren@hotmail.com | 🏠 yrcong.github.io | 📺 yrcong | 🎓 Google Scholar

## Summary

I am now working as a senior machine learning scientist at Picsart, and I am a fourth-year PhD student at the Institute for Information Processing, Leibniz University Hanover, focusing on scene understanding, generative models, and Embodied AI. I am co-advised by Prof. Michael Ying Yang and Prof. Bodo Rosenhahn. Previously, I obtained my Bachelor's degree from Hefei University of Technology and my Master's degree from Leibniz University Hanover.

## Main Publications

- [1] **Yuren Cong**, Hanno Ackermann, Wentong Liao, Michael Ying Yang and Bodo Rosenhahn. **Nodis: Neural Ordinary Differential Scene Understanding**, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.
- [2] **Yuren Cong**, Wentong Liao, Hanno Ackermann, Bodo Rosenhahn and Michael Ying Yang. **Spatial-Temporal Transformer for Dynamic Scene Graph Generation**, *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021
- [3] **Yuren Cong**, Michael Ying Yang and Bodo Rosenhahn. **RelTR: Relation Transformer for Scene Graph Generation**, *IEEE Trans on Pattern Analysis and Machine Intelligence (PAMI)*, 2023
- [4] **Yuren Cong**, Jinhui Yi, Bodo Rosenhahn and Michael Ying Yang. **SSGVS: Semantic Scene Graph-to-Video Synthesis**, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2023
- [5] **Yuren Cong**, Martin Renqiang Min, Li Erran Li, Bodo Rosenhahn and Michael Ying Yang. **Attribute-Centric Compositional Text-to-Image Generation**, *arXiv preprint*, 2023
- [6] **Yuren Cong**, Wentong Liao, Jiawei Ren, Bodo Rosenhahn, Michael Ying Yang. **Learning Similarity between Scene Graphs and Images with Transformers**, *arXiv preprint*, 2023
- [7] **Yuren Cong**, Mengmeng Xu, Christian Simon, Shoufa Chen, Jiawei Ren, Yanping Xie, Juan-Manuel Perez-Rua, Bodo Rosenhahn, Tao Xiang and Sen He. **FLATTEN: Optical Flow-guided Attention for Consistent Text-to-Video Editing**, *In International Conference on Learning Representations (ICLR)*, 2024
- [8] Shoufa Chen, Mengmeng Xu, Jiawei Ren, **Yuren Cong**, Sen He, Yanping Xie, Animesh Sinha, Ping Luo, Tao Xiang and Juan-Manuel Perez-Rua. **GenTron: Delving Deep into Diffusion Transformers for Image and Video Generation**, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024

## Experience

### Picsart

SENIOR MACHINE LEARNING SCIENTIST

Berlin, Germany

Mar. 2024 - PRESENT

#### • Video generation models

- Develop a human-centric video generation framework aimed at improving content personalization.

### Meta AI

RESEARCH SCIENTIST INTERN

London, United Kingdom

Jun. 2023 - Oct. 2023

#### • Image and video generation models

- Propose a training-free text-to-video editing framework that enforces the patches on the same optical flow path across different frames to attend to each other in the attention module, thus improving visual consistency. (cf. Publication [7])
- Propose a diffusion Transformer model for text-to-image and text-to-video generation. (cf. Publication [8])

### Cruisewatch

MACHINE LEARNING CONSULTANT

Hanover, Germany

Feb. 2020 - Jun. 2023

- Cruise travel recommendation system
- NLP-based system for customer emails

### Robert Bosch GmbH

RESEARCH INTERN

Reutlingen, Germany

July. 2018 - Dec. 2018

- Time-series classification algorithms for MEMS

### Leibniz University Hanover

RESEARCH ASSISTANT

Hanover, Germany

Nov. 2017 - Feb. 2018

- Multiple object tracking algorithm optimization

## Education

---

### Leibniz University Hanover

*Hanover, Germany*

PHD CANDIDATE

*Feb. 2020 - May 2024 (expected)*

- **Scene understanding:**
  - Propose a scene graph generation method to infer visual relationships by solving neural variants of ODE. (cf. Publication [1])
  - Extend image-based scene graph generation to video-based scene graph generation. (cf. Publication [2])
  - View scene graph generation as set prediction and address the limitations of existing models. (cf. Publication [3])
- **Generative models:**
  - Propose a multi-modal learning framework connecting video scene graphs and videos, and a generative model for scene graph-to-video synthesis. (cf. Publication [4])
  - Improve the compositional generalization in text-to-image generation. (cf. Publication [5])
- **Multi-modal learning:**
  - Propose a contrastive learning framework that can measure the similarity of scene graphs and images. (cf. Publication [6])
- **Teaching assistant:**
  - Lecture: Imaging systems for medical technology
  - Lab: Sampling and quantization

### Leibniz University Hannover

*Hannover, Germany*

M.S. IN ELECTRICAL ENGINEERING AND INFORMATION TECHNOLOGY

*Oct. 2016 - Sept. 2019*

### Hefei University of Technology

*Hefei, China*

B.S. IN ELECTRICAL ENGINEERING

*Sept. 2011 - July. 2015*

## Program Committees

---

**Reviewer**, Computer Vision and Pattern Recognition (CVPR)

**Reviewer**, European Conference of Computer Vision (ECCV)

**Reviewer**, International Conference of Computer Vision (ICCV)

**Reviewer**, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)

**Reviewer**, International Society for Photogrammetry and Remote Sensing (ISPRS)

## Skills

---

**Professional experience** Computer Vision, Machine Learning

**Programming** Python, MATLAB, C/C++

**Library** Pytorch, NumPy, Pandas, Tensorflow, OpenCV

**Operating System** Linux, Mac OS, Windows

**Language** English (fluent), Chinese (native), German (fluent)

**Other** Git, LaTeX, SLURM, Microsoft Office