Dr. **YU, LIJUN**

+1 412 758 0332 ⋄ lijun@lj-y.com ⋄ https://me.lj-y.com

EXPERIENCE

Full-Time

Google May 2024 - Present
Senior Research Scientist, Google Deepmind

During Education

Google

Research Intern / Student Researcher, Perception, Google Research May 2022 - Dec 2023
Student Researcher, Cloud AI Research, Google Cloud Jul 2021 - May 2022

GamesMind (Startup) Sept 2018 - Jul 2019
Co-Founder / Researcher

Microsoft Nov 2017 - May 2018
Research Intern, Social Computing Group, Microsoft Research Asia

EDUCATION

Carnegie Mellon University Aug 2019 - May 2024
Ph.D. in Language and Information Technologies
M.S. in Language Technologies

- Thesis: **Towards Multi-Task Multi-Modal Models: A Video Generative Perspective**.
- Committee: Dr. Alexander G. Hauptmann (advisor), Dr. Yonatan Bisk, Dr. Lu Jiang, Dr. Ming-Hsuan Yang.
- LTI Research Fellowship 2019-2024, Siebel Scholar 2021, Baidu Scholarship 2023.
- Graded A at CMU "hardest" course 15-410/605 Operating Systems Design and Implementation.

Peking University Sept 2015 - Jul 2019
B.S. in Computer Science and Technology
B. of Economics

- Summa Cum Laude, Top-Notch Class of Computer Science. Ranking: top 1%.
- Leo KoGuan Scholarship 2016, National Scholarship 2017, Arawana Scholarship 2018.
- Visiting Scholar at Carnegie Mellon University, Jun 2018 - Sept 2018.

PUBLICATIONS

Preprints

[P1] **A Versatile Diffusion Transformer with Mixture of Noise Levels for Audiovisual Generation**
Gwanghyun Kim, Alonso Martinez, Yu-Chuan Su, Brendan Jou, José Lezama, Agrim Gupta, Lijun Yu, Lu Jiang, Aren Jansen, Jacob Walker, Krishna Somandepalli.
arXiv 2024.

[P2] **Improving and Unifying Discrete&Continuous-time Discrete Denoising Diffusion**
Lingxiao Zhao, Xueying Ding, Lijun Yu, Leman Akoglu.
arXiv 2024.

[P3] **Photorealistic Video Generation with Diffusion Models**
Agrim Gupta, Lijun Yu, Kihyuk Sohn, Xiuye Gu, Meera Hahn, Li Fei-Fei, Irfan Essa, Lu Jiang, José Lezama.
arXiv 2023.

Refereed Conference Papers

[C1] **VideoPoet: A Large Language Model for Zero-Shot Video Generation**
Dan Kondratyuk, Lijun Yu*, Xiuye Gu*, José Lezama*, Jonathan Huang*, Grant Schindler, Rachel Hornung, Vighnesh Birodkar, Jimmy Yan, Ming-Chang Chiu, Krishna Somandepalli, Hassan Akbari, Yair Alon, Yong Cheng, Josh Dillon, Agrim Gupta, Meera Hahn, Anja Hauth, David Hendon, Alonso Martinez, David Minnen, Mikhail Sirotenko, Kihyuk Sohn, Xuan Yang, Hartwig Adam, Ming-Hsuan Yang, Irfan Essa, Huisheng Wang, David A. Ross, Bryan Seybold*, Lu Jiang* (* indicates core contributors with equal contribution.).*
In ICML 2024, International Conference on Machine Learning.

[C2] **Language Model Beats Diffusion: Tokenizer is Key to Visual Generation**
Lijun Yu, José Lezama, Nitesh B. Gundavarapu, Luca Versari, Kihyuk Sohn, David Minnen, Yong Cheng, Agrim Gupta, Xiuye Gu, Alexander G. Hauptmann, Boqing Gong, Ming-Hsuan Yang, Irfan Essa, David A. Ross, Lu Jiang.
In ICLR 2024, Twelfth International Conference on Learning Representations.

[C3] **Audio-Journey: Open Domain Latent Diffusion Based Text-to-Audio Generation**
Jackson Michaels, Juncheng Li, Laura Yao, Lijun Yu, Zach Wood-Doughty, Florian Metze.
In ICASSP 2024, IEEE International Conference on Acoustics, Speech and Signal Processing.

[C4] **SPAE: Semantic Pyramid AutoEncoder for Multi-Modal Generation with Frozen LLMs**
Lijun Yu, Yong Cheng, Zhiruo Wang, Vivek Kumar, Wolfgang Macherey, Yanping Huang, David A. Ross, Irfan Essa, Yonatan Bisk, Ming-Hsuan Yang, Kevin Murphy, Alexander G. Hauptmann, Lu Jiang.
In **NeurIPS (Spotlight)** 2023, Thirty-seventh Conference on Neural Information Processing Systems.

[C5] **MAGVIT: Masked Generative Video Transformer**
Lijun Yu, Yong Cheng, Kihyuk Sohn, José Lezama, Han Zhang, Huiwen Chang, Alexander G. Hauptmann, Ming-Hsuan Yang, Yuan Hao, Irfan Essa, Lu Jiang.
In **CVPR (Highlight)** 2023, IEEE/CVF Conference on Computer Vision and Pattern Recognition.

[C6] **DocumentNet: Bridging the Data Gap in Document Pre-Training**
Lijun Yu, Jin Miao, Xiaoyu Sun, Jiayi Chen, Alexander G. Hauptmann, Hanjun Dai, Wei Wei.
In EMNLP 2023, Conference on Empirical Methods in Natural Language Processing.

[C7] **Score-based Continuous-Time Discrete Diffusion Models**
Haoran Sun, Lijun Yu, Bo Dai, Dale Schuurmans, Hanjun Dai.
In ICLR 2023, Eleventh International Conference on Learning Representations.

[C8] **Rethinking Zero-Shot Action Recognition: Learning from Latent Atomic Actions**
Yijun Qian, Lijun Yu, Wenhe Liu, Alexander G. Hauptmann.
In ECCV 2022, European Conference on Computer Vision.

[C9] **Training-Free Monocular 3D Event Detection System for Traffic Surveillance**
Lijun Yu, Peng Chen, Wenhe Liu, Guoliang Kang, Alexander G. Hauptmann.
In BigData 2019, IEEE International Conference on Big Data.

[C10] **Traffic Danger Recognition with Surveillance Cameras without Training Data**
Lijun Yu, Dawei Zhang, Xiangqun Chen, Alexander G. Hauptmann.
In AVSS 2018, 15th IEEE International Conference on Advanced Video and Signals-based Surveillance.

Refereed Conference Workshop Papers

[W1] **Audio-Journey: Visual+LLM-aided Audio Encodec Diffusion**
Juncheng B Li, Jackson Michaels, Laura Jia Yao, Lijun Yu, Zach Wood-Doughty, Florian Metze.
In ICML ES-FoMO 2023, International Conference on Machine Learning: Efficient Systems for Foundation Models.

[W2] **Argus++: Robust Real-Time Activity Detection for Unconstrained Video Streams with Overlapping Cube Proposals**
Lijun Yu, Yijun Qian, Wenhe Liu, Alexander G. Hauptmann.
In WACV HADCV 2022, IEEE/CVF Winter Conference on Applications of Computer Vision: Human Activity Detection in Multi-Camera, Continuous, Long-Duration Video.

[W3] **TRM: Temporal Relocation Module for Video Recognition**
Yijun Qian, Guoliang Kang, Lijun Yu, Wenhe Liu, Alexander G. Hauptmann.

In WACV HADCV 2022.

[W4] **CMU Informedia at TRECVID 2021: Activity Detection with Argus++**
Lijun Yu, Yijun Qian, Wenhe Liu, Alexander G. Hauptmann.
In TRECVID 2021, Text Retrieval Conference Video Retrieval Evaluation.

[W5] **Learning Unbiased Transformer for Long-Tail Sports Action Classification**
Yijun Qian, Lijun Yu, Wenhe Liu, Alexander G. Hauptmann.
In MediaEval 2021, Multimedia Evaluation Benchmark.

[W6] **CMU Informedia at TRECVID 2020: Towards Real-Time Activity Recognition with Dense Spatio-Temporal Proposals**
Lijun Yu, Yijun Qian*, Wenhe Liu, Alexander G. Hauptmann (* indicates equal contribution.).*
In TRECVID 2020, Text Retrieval Conference Video Retrieval Evaluation.

[W7] **Zero-VIRUS: Zero-Shot Vehicle Route Understanding System for Intelligent Transportation**
Lijun Yu, Qianyu Feng, Yijun Qian, Wenhe Liu, Alexander G. Hauptmann.
In CVPR AI-City 2020, IEEE/CVF Conference on Computer Vision and Pattern Recognition: AI City Challenge.

[W8] **ELECTRICITY: An Efficient Multi-Camera Vehicle Tracking System for Intelligent City**
Yijun Qian, Lijun Yu*, Wenhe Liu, Alexander G. Hauptmann (* indicates equal contribution.).*
In CVPR AI-City 2020.

[W9] **Adaptive Feature Aggregation for Video Object Detection**
Yijun Qian, Lijun Yu, Wenhe Liu, Guoliang Kang, Alexander G. Hauptmann.
In WACV HADCV 2020, IEEE/CVF Winter Conference on Applications of Computer Vision: Human Activity Detection in Multi-Camera, Continuous, Long-Duration Video.

[W10] **Argus: Efficient Activity Detection System for Extended Video Analysis**
Wenhe Liu, Guoliang Kang, Po-Yao Huang, Xiaojun Chang, Lijun Yu, Yijun Qian, Junwei Liang, Liangke Gui, Jing Wen, Peng Chen, Alexander G. Hauptmann.
In WACV HADCV 2020.

[W11] **MMVG-INF-Etrol@ TRECVID 2019: Activities in Extended Video**
Xiaojun Chang, Wenhe Liu, Po-Yao Huang, Changlin Li, Fengda Zhu, Mingfei Han, Mingjie Li, Mengyuan Ma, Siyi Hu, Guoliang Kang, Junwei Liang, Liangke Gui, Lijun Yu, Yijun Qian, Jing Wen, Alexander G. Hauptmann.
In TRECVID 2019, Text Retrieval Conference Video Retrieval Evaluation.

[W12] **MOBA-Slice: A Time Slice based Evaluation Framework of Relative Advantage between Teams in MOBA Games**
Lijun Yu, Dawei Zhang, Xiangqun Chen, Xing Xie.
In IJCAI CGW 2018, International Joint Conference on Artificial Intelligence: Computer Games Workshop.

Technical Reports

[R1] **Robust Automatic Detection of Traffic Activity**
Alexander G. Hauptmann, Lijun Yu, Wenhe Liu, Yijun Qian, Zhiqi Cheng, Liangke Gui.
In CMU Mobility21 2023, University Transportation Center.

SELECTED TALKS

- **Language Model Beats Diffusion: Tokenizer is Key to Visual Generation**.
  - Invited talks at *Hong Kong Shanghai AI Forum, New York University, California Institute of Technology, Hong Kong University of Science and Technology, Chinese Academy of Sciences Institute of Computing Technology, Adobe, ByteDance, Baidu, Kunlun Tech, AISphere, and Peking University Alumni Association of Northern California*, 2024.
  - Job talks at *OpenAI, xAI, Nvidia*, 2024.
- **ArgusRoad: Road Activity Detection with Connectionist Spatio-Temporal Proposals**.
  - Invited talk at *ICCV ROAD Challenge*, 2021.
- **Argus++: Real-Time Activity Detection in Unknown Facilities with Dense Spatio-Temporal Proposals**.
  - Invited talks at *CVPR ActivityNet Challenge and WACV HADCV Workshop*, 2021.
- **Traffic Danger Recognition with Surveillance Cameras without Training Data**.

- Invited talk at *TRECVID*, 2019.
- Demo at *ICCV*, 2019.

PARTICIPATION AND SERVICES

Winning

- CVPR ActivityNet Challenge 2022: ActEV.
- ICCV ROAD Challenge 2021: Action Detection.
- CVPR ActivityNet Challenge 2021: ActEV and Kinetics-700.
- WACV ActEV Challenge 2021: SDL Unknown Facility.
- MediaEval 2021: Sports Video Stroke Classification.
- CVPR AI City Challenge 2020: City-Scale Multi-Camera Vehicle Tracking.
- TRECVID 2019 & 2020: Activities in Extended Videos.

Releasing

- MAGVIT: Masked Generative Video Transformer.
- Rethinking Zero-shot Action Recognition: Learning from Latent Atomic Actions.
- Pyturbo Package: A Pipeline System for Efficient Concurrent Execution.
- Zero-VIRUS: Zero-shot Vehicle Route Understanding System for Intelligent Transportation.
- ELECTRICITY: An Efficient Multi-camera Vehicle Tracking System for Intelligent City.
- AVI-R Package: A Robust Reader for AVI Video Files.

Reviewing

- CVPR 2023-2024, IEEE/CVF Conference on Computer Vision and Pattern Recognition.
- IJCV 2024, International Journal of Computer Vision.
- ICCV 2023, IEEE/CVF International Conference on Computer Vision.
- ECCV 2024, European Conference on Computer Vision.
- AAAI 2023-2024, AAAI Conference on Artificial Intelligence.
- WACV 2022-2023, IEEE/CVF Winter Conference on Applications of Computer Vision.
- ACMMM 2022, ACM International Conference on Multimedia.

Teaching

- Guest lecturer and teaching assistant, **Large-Scale Multimedia Analysis** (CMU 11775), Carnegie Mellon University, Jan 2022 - May 2022 & Sept 2022 - Dec 2022.
- Seminar instructor, **Introduction to Computer Systems** (Chinese version of CMU 15213), Peking University, Sept 2018 - Dec 2018.
- Chinese textbook translator for **Modern Operating Systems** (4th Edition) by Andrew S. T. and Herbert B., Peking University, Nov 2017.

Mentoring

- 5 M.S. students, Carnegie Mellon University, Sept 2021 - May 2024.
- 1 Ph.D. student, Carnegie Mellon University, Sept 2023 - May 2024.

Leading

- Secretary, Dean's Masters Advisory Committee, School of Computer Science, Carnegie Mellon University, Jan 2020 - Jun 2021.
- Vice President, Students' Union, School of Electronic Engineering and Computer Science, Peking University, Oct 2016 - May 2017.