

XINYU LI

7324066689 ◊ lixinyu.arthur@outlook.com ◊ www.arthurlxy.com

EDUCATION

Rutgers University

Ph.D. in Computer Engineering

Sep 2013 - Feb 2018

University of Electronic Science and Technology of China

B.S. in Communication Engineering

Sep 2009 - June 2013

EXPERIENCE

Amazon Prime Video

Senior Applied Scientist

Sep 2022 - present

Seattle, WA

- Leading the long-form video/Multimedia understanding product and research, content tagging, ranking and understanding.
- Leading the large-scale cross-modality learning and various downstream tasks.

ByteDance AML

Senior Research Scientist

Dec 2021 - Sep 2022

Seattle, WA

- Leading the video/Multimedia understanding research, including multimodal backbone design and large-scale modeling pre-training. Publications in CVPR, ICASSP, and WACV.
- Leading the multi-modality based downstream tasks, e.g. content moderation, video tagging, audio event detection.

Amazon AI

Senior Applied Scientist

May 2018 - Dec 2021

Seattle, WA

- Leading the video/Multimedia understanding research, including action recognition, action detection, and multimedia understanding. Publications in ECCV, CVPR, ICCV, NeurIPS, WACV, ACM MM, ACL, and INTERSPEECH.
- Designed and developed the efficient action recognition/detection training framework which has been open-sourced as part of **GluonCV-Torch** and **GluonMM**.
- Leading the video based products including the **content moderation** and **media segmentation**.

Amazon

Research Scientist Intern

June 2017 - Aug 2017

Seattle, WA

- Multi-stream fraud detection for amazon TRMS.
- Reinforcement learning based self-adaptive fraud detection system.

Multimedia Lab, Rutgers

Graduate Research Assistant

Sep 2013 - Feb 2018

New-brunswick, NJ

- Multi-modality based multi-label action recognition and action detection. The prototype system is deployed at an actual trauma room in Children's National Medical Center. Publications in CVPR, ACM MM, Sensys and Ubicomp, etc..
- Visual-acoustic human emotion recognition and sentiment analysis; publications in ACL, ACM MM, COLING.

- Single image dehazing based on dark-channel prior and wavelet transformation (Outstanding Capstone).
- Airport runway foreign object detection with adaboost.

SELECTED PUBLICATION

Selected publications in past 5 years, full list of publication can be find at Google Scholar

* denotes equally contributed.

1. Chen, Yuxiao, Jianbo Yuan, Yu Tian, Shijie Geng, Xinyu Li, Ding Zhou, Dimitris N. Metaxas, and Hongxia Yang. "Revisiting multimodal representation in contrastive learning: from patch and token embeddings to finite discrete tokens." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 15095-15104. 2023.
2. Xiaoyu Liu, Hanlin Lu, Jianbo Yuan, and Xinyu Li. "CAT: Causal Audio Transformer for Audio Classification." In ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1-5. IEEE, 2023.
3. Fan, David, Deyu Yang, Xinyu Li, Vimal Bhat, and Rohith MV. "Nearest-Neighbor Inter-Intra Contrastive Learning from Unlabeled Videos." ICLR workshop (2023).
4. Li, Xinyu, Yanyi Zhang, Jianbo Yuan, Hanlin Lu, and Yibo Zhu. "Discrete Cosin TransFormer: Image Modeling From Frequency Domain." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 5468-5478. 2023.
5. Jiaojiao Zhao*, Yanyi Zhang*, Xinyu Li*, Hao Chen, Shuai Bing, Mingze Xu, Chunhui Liu, Kaustav Kundu, Yuanjun Xiong, Davide Modolo, Ivan Marsic, Cees G.M. Snoek, Joseph Tighe. "TubeR: Tubelet Transformer for Video Action Detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR 2022 Oral.
6. A S M Iftekhar, Hao Chen, Kaustav Kundu, Xinyu Li, Joseph Tighe, Davide Modolo. "What to look at and where: Semantic and Spatial Refined Transformer for detecting human-object interactions." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR 2022 Oral.
7. Feng Cheng, Mingze Xu, Yuanjun Xiong, Hao Chen, Xinyu Li, Wei Li, Wei Xia. "Stochastic Backpropagation: A Memory Efficient Strategy for Training Video Models." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR 2022 Oral.
8. Bing Shuai, Xinyu Li, Kaustav Kundu, Joseph Tighe. "Id-Free Person Similarity Learning." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR 2022.
9. Li, Xinyu, Chunhui Liu, Bing Shuai, Yi Zhu, Hao Chen, and Joseph Tighe. "NUTA: Non-uniform Temporal Aggregation for Action Recognition." WACV 2022.
10. Zhe Wang, Hao Chen, Xinyu Li, Chunhui Liu, Yuanjun Xiong, Joseph Tighe, Charless Fowlkes. "SSCAP: Self-supervised Co-occurrence Action Parsing for Unsupervised Temporal Action Segmentation." WACV 2022.
11. Xinyu Li*, Yanyi Zhang*, Chunhui Liu, Bing Shuai, Yi Zhu, Biagio Brattoli, Hao Chen, Ivan Marsic, and Joseph Tighe. "VidTr: Video Transformer Without Convolutions." ICCV 2021.
12. Chunhui Liu*, Xinyu Li*, Hao Chen, and Joseph Tighe "Selective Feature Compression for Efficient Activity Recognition Inference." ICCV 2021.

13. Mingze Xu, Yuanjun Xiong, Hao Chen, Xinyu Li, Wei Xia, Zhuowen Tu and Stefano Soatto. "Long Short-Term Transformer for Online Action Detection." NeurIPS 2021 (Spotlight).
14. Zhang, Yanyi, Xinyu Li, and Ivan Marsic. "Multi-Label Activity Recognition using Activity-specific Features." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR 2021.
15. Shuai, Bing, Andrew G. Berneshawi, Xinyu Li, Davide Modolo, and Joseph Tighe. "Multi-object tracking with Siamese track-RCNN." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR 2021.
16. Jiaojiao Zhao*, Xinyu Li*, Chunhui Liu, Bing Shuai, Hao Chen, Cees Snoek and Joseph Tighe "TubeR: Tube-Transformer for Action Detection." arXiv preprint arXiv:2104.00969 (2021).
17. Zhu, Yi, Xinyu Li, Chunhui Liu, Mohammadreza Zolfaghari, Yuanjun Xiong, Chongruo Wu, Zhi Zhang, Joseph Tighe, R. Manmatha, and Mu Li. "A Comprehensive Study of Deep Video Action Recognition." arXiv preprint arXiv:2012.06567 (2020).
18. Shuai, Bing, Andrew Berneshawi, Manchen Wang, Chunhui Liu, Davide Modolo, Xinyu Li, and Joseph Tighe. "Application of Multi-Object Tracking with Siamese Track-RCNN to the Human in Events Dataset." In Proceedings of the 28th ACM International Conference on Multimedia, pp. 4625-4629. ACM MM 2020.
19. Li, Xinyu, Bing Shuai, and Joseph Tighe. "Directional temporal modeling for action recognition." In European Conference on Computer Vision, pp. 275-291. Springer, Cham, ECCV 2020.
20. Gu, Yue, Xinyu Lyu, Weijia Sun, Weitian Li, Shuhong Chen, Xinyu Li, and Ivan Marsic. "Mutual correlation attentive factors in dyadic fusion networks for speech emotion recognition." In Proceedings of the 27th ACM International Conference on Multimedia, pp. 157-166. ACM MM 2019.
21. Li, Xinyu, Venkata Chebiyyam, and Katrin Kirchhoff. "Speech Audio Super-Resolution for Speech Recognition." In INTERSPEECH, pp. 3416-3420. INTERSPEECH 2019.
22. Li, Xinyu, Venkata Chebiyyam, and Katrin Kirchhoff. "Multi-Stream Network with Temporal Attention for Environmental Sound Classification." Proc. Interspeech 2019 pp 3604-3608. INTERSPEECH 2019.

SELECTED OPEN-SOURCE PACKAGES

GluonCV-Torch

[Project Link](#)

- GluonCV provides implementations of state-of-the-art (SOTA) deep learning algorithms in computer vision. Available in mxnet and pytorch.

GluonMM

[Project Link](#)

- GluonCV-Transformer is an open-sourced library based on PyTorch, providing a list of SOTA transformer-based research implementations on various image tasks (image classification, object detection, semantic segmentation), video tasks (video classification, spatio-temporal action detection, long-video reasoning), and multimedia tasks (sound classification, video to text, retrieval, etc.).

TubeR: Tubelet Transformer

[Project Link](#)

- This repo contains the supported code to reproduce spatio-temporal action detection results of TubeR: Tubelet Transformer for Video Action Detection.

PROFESSIONAL SERVICES

Conference Reviewer: CVPR, ICCV, ECCV, WACV, ICLR, NeurIPS, ICASSP, InterSpeech.
Journal Reviewer: Pattern Recognition Letters, IMWUT, Transaction of Mobile Computing.