

# 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 AI**

*Applied Scientist II*

May 2018 - Present

*Seattle, WA*

- Leading the video/Multimedia understanding research, including action recognition, action detection, and multimedia understanding. Multiple publications in ECCV, CVPR, ACM MM, ACL, and INTER-SPEECH.
- Designed and developed the efficient action recognition/detection training framework which has been open-sourced as part **GluonCV-Torch** and used as training frameworks in production pipelines.
- Actively developing and maintaining the SOTA action recognition/detection model zoo, which is also available as part of **GluonCV-Torch model zoo**.
- Leading the content moderation video pipeline and multimedia pipeline. The **content moderation** is a service that detects content that is inappropriate, unwanted, or offensive.
- Leading the media segmentation and understanding pipeline. The **media segmentation** is a service that breaking up videos into clips at shot-level and scene-level.

**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*

- Computer vision-based Multi-label action recognition and surgical phase detection (action temporal localization); publications in CVPR, ACM MM.
- Visual-acoustic human emotion recognition and sentiment analysis; publications in ACL, ACM MM, COLING.
- Sensor network based concurrent action recognition; publications in Sensys, UbiComp.
- Developed and deployed the data collection system in a trauma room in Children's National Medical Center, the system collects RGB videos, depth videos, directional audio recordings, and passive RFID signal sequences for action recognition research and medical training purposes.

**Image Processing Lab, UESTC**

*Under-graduate Research Assistant*

June 2012 - June 2013

*Chengdu, China*

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

## SELECTED PUBLICATION

---

Full list of publication can be find at Google Scholar

\* denotes equally contributed.

1. Xinyu Li\*, Yanyi Zhang\*, Chunhui Liu, Bing Shuai, Yi Zhu, Biagio Brattoli, Hao Chen, Ivan Marsic, and Joseph Tighe. "VidTr: Video Transformer Without Convolutions." arXiv preprint arXiv:2104.11746 (2021).
2. Mingze Xu, Yuanjun Xiong, Hao Chen, Xinyu Li, Wei Xia, Zhuowen Tu and Stefano Soatto. "Long Short-Term Transformer for Online Action Detection." arXiv preprint arXiv:2107.03377 (2021).
3. 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).
4. Chunhui Liu\*, Xinyu Li\*, Hao Chen, and Joseph Tighe "Selective Feature Compression for Efficient Activity Recognition Inference." arXiv preprint arXiv:2104.00179 (2021).
5. 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.
6. 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.
7. Li, Xinyu, Chunhui Liu, Bing Shuai, Yi Zhu, Hao Chen, and Joseph Tighe. "NUTA: Non-uniform Temporal Aggregation for Action Recognition." arXiv preprint arXiv:2012.08041 (2020).
8. 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).
9. 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.
10. 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.
11. 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.
12. Li, Xinyu, Venkata Chebiyyam, and Katrin Kirchhoff. "Speech Audio Super-Resolution for Speech Recognition." In INTERSPEECH, pp. 3416-3420. INTERSPEECH 2019.
13. 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.
14. Gu, Yue, Xinyu Li, Kaixiang Huang, Shiyu Fu, Kangning Yang, Shuhong Chen, Moliang Zhou, and Ivan Marsic. "Human conversation analysis using attentive multimodal networks with hierarchical encoder-decoder." In Proceedings of the 26th ACM international conference on Multimedia, pp. 537-545. ACM MM 2018.

15. Gu, Yue, Kangning Yang, Shiyu Fu, Shuhong Chen, Xinyu Li, and Ivan Marsic. “Hybrid Attention based Multimodal Network for Spoken Language Classification.” In Proceedings of the 27th International Conference on Computational Linguistics, pp. 2379-2390. ACL 2018.
16. Li, Xinyu, Yanyi Zhang, Jianyu Zhang, Yueyang Chen, Huangcan Li, Ivan Marsic, and Randall S. Burd. “Region-based activity recognition using conditional GAN.” In Proceedings of the 25th ACM international conference on Multimedia, pp. 1059-1067. ACM MM 2017.
17. Li, Xinyu, Yanyi Zhang, Jianyu Zhang, Moliang Zhou, Shuhong Chen, Yue Gu, Yueyang Chen, Ivan Marsic, Richard A. Farneth, and Randall S. Burd. “Progress estimation and phase detection for sequential processes.” Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies 1, no. 3 (2017): 1-20. Ubicomp 2017.
18. Li, Xinyu, Yanyi Zhang, Ivan Marsic, Aleksandra Sarcevic, and Randall S. Burd. “Deep learning for rfid-based activity recognition.” In Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM, pp. 164-175. SenSys 2016.

## OPEN-SOURCE PACKAGES

---

### GluonCV-Torch

Project Link

- GluonCV provides implementations of state-of-the-art (SOTA) deep learning algorithms in computer vision. I am leading the GluonCV development in PyTorch.
- I wrote the efficient distributed video training pipeline and reproduced the multigrid training.
- I wrote the initial action recognition GluonCV-torch model zoo and now leading GluonCV-torch video model zoo maintenance (including action recognition, action detection, and self-supervised action classification).

### GluonMM

- 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.).
- I am leading the video transformer pipeline including video classification, spatio-temporal action detection, and movie scene segmentation.
- I am actively contributing to the image pipeline by reproducing a list of image classification models including ViT, DEiT, and SwinViT, etc.
- I am actively contributing to decord, the fast video and audio decoder open-sourced as part of Gluon.

## PROFESSIONAL SERVICES

---

Conference Reviewer: CVPR, ICCV, ACM Multimedia, Ubicomp, CHI

Journal Reviewer: Pattern Recognition Letters, IMWUT, Transaction of Mobile Computing.