Ziyue Xiang (Alan Xiang)

Applied Research Intern @ Tencent Americas | Ph.D. Candidate @ Purdue University, West Lafayette

Expected Graduation: Jan. 2024

Sunnyvale, California, USA
ziyue.alan.xiang@gmail.com

xiang71@purdue.edu
ziyue-alan-xiang
alanshawn.com
github.com/xziyue
UKO6iCgAAAAJ
0000-0001-6054-5801

Language: Mandarin, English, Cantonese
Pronoun: He/His

Last Update: September 6, 2023

Contents

Summary

I am a Ph.D. candidate with experience in audio/video/image compression, signal processing, computer vision, and deep learning. I am currently working on developing next generation international video coding standards. My Ph.D. thesis is focused on digital media forensics, with topics including media metadata analysis, bitstream analysis, deepfake detection, and media manipulation detection.

Education

Ph.D. in Electrical and Computer Engineering *Aug. 2020–Jan. 2024*
Purdue University *West Lafayette, IN, USA*
- Advisor: Prof. Edward J. Delp

MSc. in Computer Science *Aug. 2018–May 2020*
Syracuse University *Syracuse, NY, USA*

BSc. in Information and Computing Science *Sept. 2014–June 2018*
Sun Yat-sen University *Guangzhou, Guangdong, China*
- Best undergraduate thesis award

Experience

Applied Research Intern *June 2023–Dec. 2023*
Tencent Americas *Palo Alto, CA, USA*
- Developing next generation video coding standard beyond Versatile Video Coding (VVC)

Research Assistant *Aug. 2020–Dec. 2023*
Purdue University *West Lafayette, IN, USA*
- Using computer vision, signal processing, and machine learning techniques to conduct media forensics research

Research Assistant *Aug. 2018–July 2020*
Syracuse University *Syracuse, NY, USA*
- Developing algorithms to automatically analyze the integrity of scientific figures at Prof. Daniel E. Acuna's SOS+CD lab

Teaching Assistant *Jan. 2020–May 2020*
Syracuse University *Syracuse, NY, USA*
- Teaching assistant for *CIS-375: Discrete Mathematics* offered by Prof. Andrew C. Lee

Selected Projects

Video Coding Standard Beyond Versatile Video Coding (VVC)
- Develop new intra prediction methods for the latest MPEG video coding standard known as the **Enhanced Compression Model (ECM)**
- Evolve new techniques to debug, improve, and verify complex image/video processing software
- Convert developed intellectual properties to patent applications

Analysis of MP4 Container Metadata Using Machine Learning
- In project [5], a custom MP4 container metadata parser was developed with the ability to export the tree structure and interpret vendor-specific data formats

Contents

- Employed feature engineering techniques to convert the MP4 metadata tree into feature vectors and used random forest classifiers to analyze the MP4 containers
- In a followup project [1], **Self-supervised Learning** strategies were used to train a **Graph Neural Network** for MP4 metadata analysis

Analysis of H.264 Bitstream Using Deep Learning

- In project [4], the **openh264** H.264 decoder was modified to extract encoding parameters such as macroblock type, motion vectors, quantization parameters from the bitstream
- Analyzed the H.264 encoding parameters with a **Vision Transformer**-based **Siamese Neural Network** to determine the capturing device of a given video bitstream

Analysis of MP3 Bitstream Using Deep Learning

- In project [3], the **minimp3** MP3 decoder was modified to extract MP3 encoding parameters such as MDCT coefficients, scalefactors, Huffman table indices from the bitstream
- The MP3 encoding parameters were analyzed by a **Convolution Neural Network+Transformer Neural Network** hybrid network to determine which parts of the MP3 audio are compressed more than once; the compression inconsistency can be used for audio manipulation detection

Synthetic (Deepfake) Audio Detection

- Modified MP3 decoder (**minimp3**) and AAC decoder (**fdk-aac**) to extract encoding parameters from audio bitstreams
- In project [9], the AAC scalefactors were used with **Transformer Neural Networks** for synthetic speech detection
- In project [2], the MP3 encoding parameters were used to compute the spectrogram efficiently, which was analyzed by popular neural network architectures such as **ResNet**, **EfficientNetV2**, **MobileNetV3** for synthetic speech detection
- In project [8], a **Variational Autoencoder** was used to extract **Disentangled Representation** for synthetic speech signals, which allows interpretable synthetic speech detection

Analyzing Scientific Figures

- In project [6], hand-crafted features based on a series of One-class **Support Vector Machines** were designed for scientific figure manipulation detection, with a neural network classifier as the backend
- In project [13], the **Oriented FAST and Rotated BRIEF (ORB)** features of ~7 million scientific figures from the PubMed Open Access Subset were computed using **Apache Spark**, which was used to fit a probabilistic model to estimate the confidence of figure reuse instances found by research integrity investigators
- In project [14], the **Faster-RCNN** network was fine-tuned to localize figure components in scientific figures such as figure body, color bar, and legend; the extracted results were used with optical character recognition, color matching, connect component labeling, and thresholding techniques to extract data points from scientific figures

Skills

- Programming languages: Python, C++
- Machine learning frameworks: TensorFlow, PyTorch, scikit-learn
- Image/audio/video processing
- The ability to understand and modify video/audio/image codec software such as *openh264*, *fdk-aac*, *ECM*, and *libav*
- Experience in computer graphics (OpenGL, ray tracing, simulation)
- Experience in compiler design (front-end, LLVM IR, optimization)
- Experience in computer vision (object detection, semantic segmentation, homography

Contents

estimation, camera calibration, binocular vision, etc.)

Publications

[1] MTN: Forensic Analysis of MP4 Container Metadata Using Graph Neural Networks
Ziyue Xiang, Amit Kumar Singh Yadav, Paolo Bestagini, Stefano Tubaro, Edward J. Delp
IEEE/CVF Computer Vision and Pattern Recognition Workshops (CVPRW), 2023

[2] Extracting Efficient Spectrograms From MP3 Compressed Speech Signals for Synthetic Speech Detection
Ziyue Xiang, Amit Kumar Singh Yadav, Paolo Bestagini, Stefano Tubaro, Edward J. Delp
ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec), 2023

[3] Forensic Analysis and Localization of Multiply Compressed MP3 Audio Using Transformers
Ziyue Xiang, Paolo Bestagini, Stefano Tubaro, Edward J. Delp
International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2022

[4] H4VDM: H.264 Video Device Matching
Ziyue Xiang, Paolo Bestagini, Stefano Tubaro, Edward J. Delp
International Conference on Pattern Recognition Workshops (ICPRW), 2022

[5] Forensic Analysis of Video Files Using Metadata
Ziyue Xiang, János Horváth, Sriram Baireddy, Paolo Bestagini, Stefano Tubaro, Edward J. Delp
IEEE/CVF Computer Vision and Pattern Recognition Workshops (CVPRW), 2021

[6] Scientific Image Tampering Detection Based on Noise Inconsistencies: A Method and Datasets
Ziyue Xiang, Daniel E. Acuna
arXiv preprint arXiv:2001.07799, 2020

[7] PS3DT: Compression Robust Synthetic Speech Detection Transformer
Amit Kumar Singh Yadav, **Ziyue Xiang**, Kratika Bhagtani, Paolo Bestagini, Stefano Tubaro, Edward J. Delp
IEEE International Conference on Machine Learning and Applications (ICMLA), 2023

[8] DSVAE: Interpretable Disentangled Representation for Synthetic Speech Detection
Amit Kumar Singh Yadav, Kratika Bhagtani, **Ziyue Xiang**, Paolo Bestagini, Stefano Tubaro, Edward J. Delp
IEEE International Conference on Machine Learning and Applications (ICMLA), 2023

[9] ASSD: Synthetic Speech Detection in the AAC Compressed Domain
Amit Kumar Singh Yadav, **Ziyue Xiang**, Emily R. Bartusiak, Paolo Bestagini, Stefano Tubaro, Edward J. Delp
International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2023

[10] An Overview of Recent Work in Multimedia Forensics
Kratika Bhagtani, Amit Kumar Singh Yadav, Emily R. Bartusiak, **Ziyue Xiang**, Ruiting Shao, Sriram Baireddy, Edward J. Delp
International Conference on Multimedia Information Processing and Retrieval (MIPR), 2022

[11] Deepfake Detection Using Multiple Data Modalities
Hanxiang Hao, Emily R. Bartusiak, David Güera, Daniel Mas Montserrat, Sriram Baireddy, **Ziyue Xiang**, Sri Kalyan Yarlagadda, Ruiting Shao, János Horváth, Justin Yang
Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks, 2022

[12] Sat U-net: a Fusion Based Method for Forensic Splicing Localization in Satellite Images
János Horváth, **Ziyue Xiang**, Edoardo Daniele Cannas, Paolo Bestagini, Stefano Tubaro, Edward J Delp
Multimodal Image Exploitation and Learning, 2022

[13] Estimating a Null Model of Scientific Image Reuse to Support Research Integrity Investigations
Daniel E. Acuna, **Ziyue Xiang**
arXiv preprint arXiv:2003.00878, 2020

[14] SciEye: A System for Finding the Underlying Datasets for Scientific Figures
Ziyue Xiang, Edward J. Delp
GitHub Project (https://github.com/xziyue/SciEye), 2022

Professional Service

- Reviewer for *Journal of Visual Communication and Image Representation*
- Reviewer for *Conference on Vision and Intelligent Systems*
- Reviewer for *Journal of Testing and Evaluation*
- Reviewer for *Forensic Science International: Digital Investigation*
- I am an active member of the LaTeX developer community; my LaTeX3 language tutorial has over 10,000 views; I also contributed CTAN packages such as *luaprogtable* and *smart-eqn*