Preface: IJCAI 2023 Workshop on Deepfake Audio Detection and Analysis (DADA 2023)

Introduction

Deepfake audio detection is an emerging topic in artificial intelligence (AI) fields. However, the current research focuses more on performing binary classification between real and fake audio. There is nonetheless also an interest in surpassing the constraints of binary real/fake classification, and actually analyzing the deepfake audio (e.g. locating the manipulated regions in a partially fake audio and recognizing the algorithm of the deepfake audio). To this end, the Second Audio Deepfake Detection Challenge (ADD 2023) is underway to foster further research in audio deepfake detection. In addition, we also launch the IJCAI 2023 Workshop on Deepfake Audio Detection and Analysis (DADA 2023). In this workshop, we aim to bring together AI researchers from the fields of audio deepfake detection and generation, deepfake audio analysis, audio fake game, adversarial attacks and defense to further discuss recent research and future directions.

This year's workshop was held in conjunction with the International Joint Conference on Artificial Intelligence (IJCAI 2023). The workshop received 41 submissions, all of which were peer-reviewed by members of our program committee. After the review phase, we selected 21 papers for presentation, and among them, we awared one as the best paper. The workshop discusses recent advances in deepfake generation, detection and analysis. The topics studied include, among other things, adversarial attacks and defense for audio AI systems, manipulation or fake region location, deepfake algorithm recognition.

Invited Speakers

Zhizheng Wu

Biography: Zhizheng Wu is an associate professor at the Chinese University of Hong Kong, Shenzhen. Prior to that, he led teams and performed research at Meta, JD.com, Apple, the University of Edinburgh, and Microsoft Research Asia. Zhizheng received his Ph.D. from Nanyang Technological University, Singapore. Zhizheng is the creator of Merlin, an open-source speech synthesis toolkit. He co-initiated and co-organized the first speaker verification spoofing and countermeasures challenge and the Voice Conversion Challenge 2016. He organized the Blizzard Challenge 2019. Zhizheng is an associate editor of IEEE/ACM Transactions on Audio Speech and Language Processing and an elected member of the IEEE Speech and Language Processing Technical Committee.
Talk Title: Recent Advances in Voice Spoofing Detection

Accepted Papers

The following full papers presenting original research works were accepted. Yi et al. report an overview of the the Second Audio Deepfake Detection Challenge (ADD 2023), which consists of four distinct subchallenges. Different from previous challenges (e.g. ADD 2022), ADD 2023 focuses on surpassing the constraints of binary real/fake classification, and actually localizing the manipulated intervals in a partially fake speech as well as pinpointing the source responsible for generating any fake audio.

There are a number of contributions focused on improving the performance of speech synthesis systems. Zhan et al. present a speech synthesis system, which has yielded promising results in the adversarial process with the fake audio detection model. Hua et al. report a multi-stage audio synthesis system that participated in Track 1.1 in ADD 2023 Challenge. Zhang et al. study effective adversarial attacks for black-box SRSs. By adopting the boundary information, Zhang et al. propose DAoB, a new adversarial example generation strategy that effectively improves the transferability of adversarial examples. Wu et al. review the defense methods against adversarial attacks and partially fake speech attacks that have recently emerged.

In addition, there are many recent advances in the field of deepfake audio detection. Martín-Doñas et al. present the system for the ADD 2023 challenge Track 2 based on a pre-trained wav2vec2 feature extractor and downstream neural networks for detection and clustering of partial deepfakes. Zhang et al. propose a method to address the problem of low detection accuracy of models facing fake audio generated by new emerging spoofing algorithms. Li et al. improve the performance of the manipulation region system by mitigating the biases introduced by AASIST at the utterance level and Wav2Vec2 at the frame level. Li et al. present their proposed system on ADD 2023 Challenge Track 2- Manipulation Region Location (RL). The system apply the Convolutional Recurrent Neural

© 2023 Author:Pleasefillinthe\copyrightclause macro
CEUR Workshop Proceedings (CEUR-WS.org)

CEUR-WS.org/Vol-3597/preface.pdf

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

Network where the CNN extracts high temporal resolution features and RNN models the context information. Wu et al. describe the system of USTC-NERCSLIP submitted to the Track 1.2 of ADD 2023 Challenge, which involves the wav2vec2-based feature extractor and the AASIST-based classifier. Liu et al. propose a novel TranssionADD system as a solution to the challenging problem of model robustness and audio segment outliers in the trace competition. Xie et al. propose a Shuffle Mix Aggregation and Separation Domain Generalization (SMASDG) method for audio deepfake detection, which enables single-domain generalization with excellent robustness and generality. Zhang et al. introduce the system developed for ADD 2023 Track 1.2. The system introduces the Energy-based Open-World Softmax (EOW-Softmax) to calibrate model confidence and achieves the second place in the challenge. propose the multi-perspective information fusion (MPIF) Res2Net with random Specmix for fake speech detection (FSD). The main purpose of this system is to improve the model's ability to learn precise forgery information for FSD task in low-quality scenarios. Wang et al. propose a novel low-rank adaptation method (LoRA) to improve the efficiency and performance of the wav2vec2 model for the fake audio detection task.

On the topic of deepfake algorithm recognition, Han et al. present the models employed by CAU_KU team participating in Track 1.2 and Track 3 of the ADD 2023 challenge. The team utilized various deepfake models, including the W2V2 pretrained model and a modified AASIST architecture. Wang et al. introduce the NPU-ASLP system for the Deepfake Algorithm Recognition task. Contributions include data augmentation, model architecture, fine-tuning strategy, and model ensemble. Tian et al. propose a manifold-based multi-model fusion approach for open-set recognition. This approach constructs a manifold space to fuse the deep embedding features extracted by different models and computes the geodesic distance between the manifold spaces of different deepfake algorithm. Lu et al. propose cosine similarity based kNN distance to separate unknown samples from known ones. Together with data augmentation methods and logits based model fusion, the system wins first place in ADD 2023 Track 3. Zeng et al. present the system for the ADD 2023 Challenge Deepfake Algorithm Recognition Track, which is based on pre-trained wav2vec2.0-base and ECAPA-TDNN models. Qin et al. regard the deepfake algorithm recognition as speaker verification and propose the center-based maximum similarity method to determine the test audio category.

The success of DADA 2023 would not have been possible without the enthusiasm and contributions of a group of people. We sincerely thank the IJCAI 2023 Workshop Chairs, Hadi Hosseini (Penn State University, US) and Viviana Mascardi (the University of Genova, IT) for their support. We sincerely hope that the participants enjoyed the DADA workshop and that this collection of papers will inspire and facilitate more future research in the deepfake audio community.

Jiangyan Yi

Macao, 2023

Organisation

General Chairs

- Jianhua Tao, Tsinghua University
- Haizhou Li, National University of Singapore, the Chinese University of Hong Kong
- Jiangyan Yi, Institute of Automation, Chinese Academy of Sciences

Program Committee

- Ha-Jin Yu, University of Seoul
- Kong Aik LEE, Singapore Institute of Technology
- Hung-yi Lee, National Taiwan University
- Jennifer Williams, University of Southampton
- Qing Guo, Agency for Science, Technology, and Research (A*STAR), Singapore
- Juan Manuel Martín Doñas, Vicomtech Foundation, San Sebastián-Donostia (Spain)

Publication Chair

- Cunhang Fan, Anhui University