Warsaw University of Technology

FACULTY OF
ELECTRONICS AND INFORMATION TECHNOLOGY

Institute of Institute of Electronics and Information Technology

Bachelor’s diploma thesis

in the field of study Computer science
and specialisation Computer Systems and Networks

Audio Deepfake Detection: An Iterative Approach with
Feature Matching Self-Supervised Learning (FMSL)

Ansh Choudhary
student record book number 317174

thesis supervisor
dr hab. inż. Włodzimierz Kasprzak

WARSAW 2025

Audio Deepfake Detection: An Iterative Approach with Feature Matching Self-Supervised Learning (FMSL)

Abstract: Enhancing TTS Deepfake Audio Detection through Iterative Modeling and a Novel Pipeline Approach with Feature Matching Self-Supervised Learning (FMSL)

This thesis presents a systematic approach to enhancing Text-to-Speech (TTS) deepfake audio detection through iterative modeling and a novel pipeline solution combining multiple layers of feature extraction and classification methods. The work consists of two main stages: systematic exploration of 8 baseline models (Maze 1-8) to identify common performance limitations, followed by development and validation of a pipeline approach that integrates Feature Matching Self-Supervised Learning (FMSL) with various components such as Wav2Vec2, Transformer, and other methods. The key contribution is the first systematic identification of core limitations in baseline TTS deepfake detection models through iterative architectural exploration, followed by novel pipeline application achieving 83% performance improvement.

Experiments were conducted on the ASVspoof 2019 dataset, utilizing diverse pipeline architectures from simple RawNet2 to advanced models incorporating multiple layers such as Wav2Vec2, Transformer, and FMSL components. The identification of geometric bottlenecks in feature representations led to the development of a pipeline solution that combines self-supervised learning with feature matching to shape feature manifolds. Detailed analysis of Maze6 with FMSL pipeline demonstrates dramatic precision improvement from 27.66% to 50.83%.

Keywords: TTS deepfake detection, pipeline approach, iterative modeling, FMSL, forgery detection, self-supervised learning

Audio Deepfake Detection: An Iterative Approach with Feature Matching Self-Supervised Learning (FMSL)

Streszczenie. Ulepszanie wykrywania TTS deepfake'ów audio poprzez iteracyjne modelowanie i nowe podejście pipeline z Feature Matching Self-Supervised Learning (FMSL)

Niniejsza praca przedstawia systematyczne podejście do ulepszania wykrywania Text-to-Speech (TTS) deepfake'ów audio poprzez iteracyjne modelowanie i nowe rozwiązanie pipeline łączące wiele warstw metod ekstrakcji cech i klasyfikacji. Praca składa się z dwóch głównych etapów: systematycznej eksploracji 8 modeli bazowych (Maze 1-8) w celu identyfikacji wspólnych ograniczeń wydajności, a następnie opracowania i walidacji podejścia pipeline, które integruje Feature Matching Self-Supervised Learning (FMSL) z różnymi komponentami takimi jak Wav2Vec2, Transformer i inne metody. Głównym wkładem jest pierwsza systematyczna identyfikacja podstawowych ograniczeń w modelach wykrywania TTS deepfake'ów poprzez iteracyjną eksplorację architektoniczną, po której następuje nowe zastosowanie pipeline osiągające 83% poprawę wydajności.

Eksperymenty przeprowadzono na zbiorze danych ASVspoof 2019, wykorzystując różnorodne architektury pipeline od prostego RawNet2 po zaawansowane modele łączące wiele warstw takich jak Wav2Vec2, Transformer i komponenty FMSL. Identyfikacja ograniczeń geometrycznych w reprezentacjach cech doprowadziła do opracowania rozwiązania pipeline, które łączy uczenie samokontrolowane z dopasowywaniem cech w celu kształtowania rozmaitości cech. Szczegółowa analiza Maze6 z pipeline FMSL wykazuje dramatyczną poprawę precyzji z 27,66% do 50,83%.

Słowa kluczowe: wykrywanie TTS deepfake, podejście pipeline, iteracyjne modelowanie, FMSL, wykrywanie fałszerstw, uczenie samokontrolowane

Politechnika Warszawska
Warsaw University of Technology

………......................
miejscowość i data
place and date

……………………………..
imię i nazwisko studenta
name and surname of the student

……………………………..
numer albumu
student record book number

……………………….……….
kierunek studiów
field of study

OŚWIADCZENIE

DECLARATION

Świadomy/-a odpowiedzialności karnej za składanie fałszywych zeznań oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie, pod opieką kierującego pracą dyplomową.
Under the penalty of perjury, I hereby certify that I wrote my diploma thesis on my own, under the guidance of the thesis supervisor.

Jednocześnie oświadczam, że:
I also declare that:

- niniejsza praca dyplomowa nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 roku o prawie autorskim i prawach pokrewnych (Dz.U. z 2006 r. Nr 90, poz. 631 z późn. zm.) oraz dóbr osobistych chronionych prawem cywilnym,
- *this diploma thesis does not constitute infringement of copyright following the act of 4 February 1994 on copyright and related rights (Journal of Acts of 2006 no. 90, item 631 with further amendments) or personal rights protected under the civil law,*
- niniejsza praca dyplomowa nie zawiera danych i informacji, które uzyskałem/-am w sposób niedozwolony,
- *the diploma thesis does not contain data or information acquired in an illegal way,*
- niniejsza praca dyplomowa nie była wcześniej podstawą żadnej innej urzędowej procedury związanej z nadawaniem dyplomów lub tytułów zawodowych,
- *the diploma thesis has never been the basis of any other official proceedings leading to the award of diplomas or professional degrees,*
- wszystkie informacje umieszczone w niniejszej pracy, uzyskane ze źródeł pisanych i elektronicznych, zostały udokumentowane w wykazie literatury odpowiednimi odnośnikami,
- *all information included in the diploma thesis, derived from printed and electronic sources, has been documented with relevant references in the literature section,*
- znam regulacje prawne Politechniki Warszawskiej w sprawie zarządzania prawami autorskimi i prawami pokrewnymi, prawami własności przemysłowej oraz zasadami komercjalizacji.
- *I am aware of the regulations at Warsaw University of Technology on management of copyright and related rights, industrial property rights and commercialisation.*

Oświadczam, że treść pracy dyplomowej w wersji drukowanej, treść pracy dyplomowej zawartej na nośniku elektronicznym (płycie kompaktowej) oraz treść pracy dyplomowej w module APD systemu USOS są identyczne.
I certify that the content of the printed version of the diploma thesis, the content of the electronic version of the diploma thesis (on a CD) and the content of the diploma thesis in the Archive of Diploma Theses (APD module) of the USOS system are identical.

....
czytelny podpis studenta
legible signature of the student

Contents

Introduction

1. Problem Statement

1.1. Rising Threat of Sophisticated Deepfake Audio

1.2. Need for Robust, Generalizable Detection Models

2. Research Goal

2.1. Two-Stage Methodology Overview

2.2. Core Research Questions

3. Research Contributions

3.1. Novel Iterative Approach

3.2. Core Limitation Identification

3.3. FMSL Innovation

3.4. Performance Breakthrough

3.5. Standardization Framework

4. Thesis Organization

Dataset and Preprocessing

5. ASVspoof2019 Dataset Description

5.1. Dataset Characteristics

5.2. Evaluation Metrics

6. Data Processing Pipeline

6.1. Feature Extraction Process

6.2. Data Augmentation and Normalization

7. Train/Validation/Test Splits

7.1. Data Partitioning Strategy

Baseline Model Development - An Iterative Approach (Mazes 1-8)

8. Iterative Model Development Philosophy

8.1. Systematic Exploration Strategy

8.2. Justification for Each Modification

9. Detailed Model Development (Maze 1-8)

9.1. MAZE1: RawNet2 Baseline

9.2. MAZE2: Trainable SincConv RawNet2

9.3. MAZE3: SE + Transformer RawNetSinc

9.4. MAZE4: SpecAugment RawNetSinc

9.5. MAZE5: SpecAugment + FocalLoss RawNetSinc

9.6. MAZE6: RawNet + Wav2Vec2 + Transformer

9.7. MAZE7: Wav2Vec2 + SpecAugment RawNet

9.8. MAZE8: Advanced Multi-Modal Architecture

10. Architectural Analysis Framework

10.1. Feature Extraction Capabilities

10.2. Sequence Modeling Approaches

10.3. Pooling and Classification

Identification of Core Limitation in Baseline Models

11. Aggregate Performance Analysis

11.1. Comprehensive Baseline Results

11.2. Performance Visualization

12. Deep Analysis of Common Failure Patterns

12.1. Error Analysis Across Models

12.2. Common Thread Identification

13. Formulating the Core Problem Hypothesis

13.1. Identified Core Limitation

13.2. Supporting Evidence from Results

14. Problem Characterization

14.1. Technical Description of the Limitation

14.2. Impact on Detection Performance

Proposed Solution - Feature Matching Self-Supervised Learning (FMSL)

15. Introduction to FMSL

15.1. FMSL Concept and Theory

15.2. FMSL Architecture

16. Justification for FMSL as Solution

16.1. Direct Problem-Solution Mapping

16.2. Theoretical Advantages

17. FMSL Implementation Details

17.1. Standardized FMSL Framework

17.2. Integration with Baseline Models

Experimental Validation and Results Analysis

18. FMSL vs Baseline Comparison

18.1. Head-to-Head Performance Analysis

18.2. Universal Improvement Validation

19. Deep Dive on MAZE6: The Optimal Architecture

19.1. Why MAZE6 Baseline Performed Best

19.2. Why MAZE6 + FMSL Saw Most Improvement (83%)

20. Analysis of MAZE7 & MAZE8 Performance Plateau

20.1. Limited Improvement Observation

20.2. Hypothesis for Plateau

21. Statistical Validation

21.1. McNemar's Test Results

21.2. Comprehensive Results Summary

22. Deep Dive Analysis: The Efficiency of FMSL on the Optimal Maze6 Architecture

22.1. Quantitative Performance and Statistical Validation

22.2. Visual Validation of FMSL's Impact

22.3. Theoretical Foundation and Related Work

22.4. Statistical Significance and Robustness

22.5. Synthesis and Conclusion

Justification of Standardized Experimental Parameters

23. Architectural Standard: ’filters’: [128, [128, 128], [128, 256]]

23.1. Reasoning Behind Standard Config

23.2. Evidence for Standardization

24. Choice of Loss Function: Categorical Cross-Entropy (CCE)

24.1. Justification for CCE Usage

24.2. Alternative Loss Functions Considered

25. Standardized Training Parameters

25.1. Consistent Training Framework

25.2. Hyperparameter Optimization

26. Justification Summary

Conclusion and Future Work

27. Holistic Summary of Research Journey

27.1. Complete Research Narrative

27.2. Key Achievements

28. Driving the Point Home: Comprehensive Results

28.1. Summary Visualizations

28.2. Key Performance Metrics

29. Future Research Directions

29.1. Potential Extensions

29.2. Open Questions

List of Symbols and Abbreviations

FMSL – Feature Matching Self-Supervised Learning
ASVspoof – Automatic Speaker Verification Spoofing

List of Figures

List of Tables

List of Appendices