BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2023
http://excel.fit.vutbr.cz

Testing the robustness of a voice biometrics system against deepfakes

Jakub Reš

Abstract

Topic of this paper is a methodology of testing the robustness of a biometric system against deepfakes. The main problem currently lies in insufficient coverage of testing against the presentation attack using deepfakes in ISO/IEC standards. The aim of this paper is to cover the hole, resulting from emergence of deepfake technology, by proposing an extended methodology, based on the existing one, that focuses on fixing the issue. The solution of proposed problem started by studying the state of the art for deepfakes and standard practices of biometric system testing. Second step is forming my own method of testing a existing voice biometric system – Phonexia and conduction said test. And at last, generalization of the procedure to be suitable for repeatable testing of wide range of biometric systems, thus forming a methodology. In the paper, I proposed and documented a method of testing the voice biometric system. The test was designed as a scenario, where the Phonexia voice biometric system is used as a remote verification tool for the voice-as-a-password use-case. For the purpose of demonstration, the online publicly available dataset was used. On top of test design, I set a non-standard metric for the test evaluation to show possibilities of focus on different kinds of deepfakes. After carrying out tests, I formulated the procedure into a generic repeatable methodology, containing practices and recommendations. The contribution of this work lies in incorporating deepfakes into the existing standard methodologies of testing a biometric systems, hence forming and demonstrating a repeatable methodology.

*xresja00@stud.fit.vutbr.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

[Motivation] Deepfakes are new emerging problem [1]. Besides all the possibilities of manipulation and blackmailing, deepfakes are often used also as a tool for spoofing attacks on the biometric access systems. For this reason, the developers are required to act fast and adapt their existing system to face new kinds of attacks. The development cycle consists not only from implementation of new detection methods, but it also contains the testing phase. Testing is usually done using standard methodologies. However the current standards are outdated and does not consider deepfakes as an alternative for testing. This work aims to fix this issue and propose methodology that fixes the problem.

[Problem definition] Today standards for biometric system testing contain a proven generic practices for companies to follow when evaluating their products. But the standards only consider conventional methods of forgeries generation. As all these methods, deepfakes need to be acknowledged and treated carefully. This work focuses on demonstration of the test method design and execution for a provided biometric system, followed by proposing a repeatable way of testing the biometric system's robustness against the spoofing attack using deepfakes, based on the existing standards.

[Existing solutions] As mentioned before, existing solutions, mainly in form of standards, contain generic procedures for testing a biometric system. The standards this work considers are primarily ISO/IEC 19795-1, ISO/IEC 19795-2 and ISO/IEC 30107-3 [2][3][4].

In terms of deepfake-specific testing, there are numerous papers about testing the detection methods, but very little to none that tests the implemented methods as a part of any system. Despite that they offer valuable information and practices worth taking inspiration from, such as [5] or [6].

[Our solution] Solution this work proposes lies in combining the existing methods of testing, as outlined in standards, and the approaches of testing the deepfake detection methods. Adopting the long-established practices and extending them by useful additions relating to deepfakes and their evaluation appears to be the right approach.

[Contributions] My work offers a solution to the problem of testing biometric systems against spoofing attack using deepfakes. It presents a general repeatable methodology, not only the procedure, but also recommendations on what to watch out for and not to neglect during the performance evaluation.

2. Testing method

For the purpose of the methodology formulation and demonstrating the procedure I have proposed my own testing method of a supplied, commercially used voice biometric system Phonexia. The method is based on the standard-recommended practices with the addition of the focus on using publicly available deepfake data sets and proposing non-standard metrics as an example of possible monitoring of the various types of deepfakes according to their creation methods.

This procedure is shown in the Fig. B of the poster. The figure depicts the bases of method and the resulting parts:

- Goal – the goal of the test (measure the robustness of biometric system against different groups of deepfakes with the main goal of determining the resistance to different methods of creating forgeries)
- Use-case – the use-case of system that will be tested (verification, voice-as-a-password)
- Attacker model – the motivation, opportunity and resources of the potential attacker
- Scenario – the summary of previous parts into testing scenario
- Dataset – the publicly available dataset used (ASVSpoof19)
- Metrics – metrics of biometric system to be evaluated (standard/custom)

3. Test execution report

According to the proposed method, the testing of the system, as a tool used for remote voice-as-a-password verification, has been conducted. The work contains detailed procedure performed, including description of environment, the tested system properties and description of communication between the tester and tested subjects. Next the work describes the experiments, how the scenario was simulated and fulfilled, how the dataset was modified according to system requirements and what results were measured. At last I described evaluation of measured values and how proposed metrics could help identifying possible vulnerabilities.

The Fig. C shows measured performance as a proposed metric AUC (Area under curve) vs. deepfake types and how this metric helps understanding weaknesses towards different methods of deepfake generation.

4. Methodology

Fig. D shows the gist of proposed methodology. The five main parts are:

- Planning phase – the first phase of every testing. This phase is focused on gathering the important information about the system itself, the potential attackers and defining the main goal of the test.
- Data acquisition – this phase is all about recommendations regarding acquiring the proper dataset. Whether by collecting your own samples or using prepared available databases.
- Text execution – this phase is usually the same as the existing standards propose.
- Evaluation – the phase about metrics and reasoning behind using custom ones.
- Interpretation – the last short phase about discussing the results relevance according to statistic rules established by standards.

Attribution

All icon graphic in the poster was acquired from website https://thenounproject.com under the Common Creations licence. Authors of icons are listed bellow.

Robot – Adiyogi, Deepfake – Becris, Document – Rank Sol, Test – UNKNOWN, Target – Lemon Liu, Clapper – Atif Arshad, Hacker – Andy Horvath, Scenario – WEBTECHOPS LLP, Dataset – Srinivas Agra, Metrics – Srinivas Agra, Mindmap – Noura Mbarki

References

[1] Martin Havlík. Deepfake jako pokročilá manipulativní technika k šíření propagandy. online, 2023. www.vojenskerozhledy.cz.

[2] International Organization for Standardization. Iso/iec 19795-1:2006 – biometric performance testing and reporting — part 1: Principles and

framework. Standard, International Organization for Standardization, April 2006.

[3] International Organization for Standardization. Iso/iec 19795-2:2007 – biometric performance testing and reporting – part 2: Testing methodologies for technology and scenario evaluation. Standard, International Organization for Standardization, February 2007.

[4] International Organization for Standardization. Iso/iec 30107-3:2017 – biometric presentation attack detection — part 3: Testing and reporting. Standard, International Organization for Standardization, September 2017.

[5] Yuezun Li and Siwei Lyu. Exposing deepfake videos by detecting face warping artifacts. *CoRR*, abs/1811.00656, 2018.

[6] Xin Yang, Yuezun Li, and Siwei Lyu. Exposing deep fakes using inconsistent head poses. *CoRR*, abs/1811.00661, 2018.