BRNO UNIVERSITY OF TECHNOLOGY
FACULTY OF INFORMATION TECHNOLOGY

Deepfake Detection Framework

Bc. Jan Bernard, xberna18@stud.fit.vutbr.cz
Supervisor: Mgr. Kamil Malinka, Ph.D.

a.k.a (DF)^2

Motivation

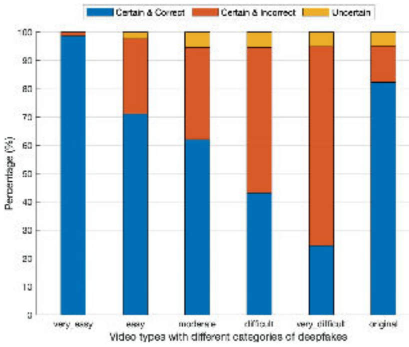
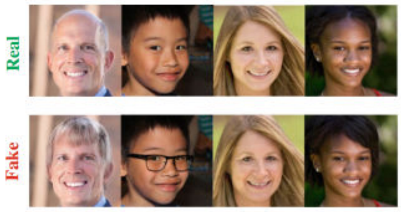


Figure 1: Subjective answers from ANOVA test for different deepfake categories. Retrieved from [1]

Deefake is the buzzword that has no agreed-upon technical definition. It consists of two words, deep and fake. Deep is referring to deep machine learning, which is used for creating fake voices, images, or even videos.

It is a fast growing technical field of study and could be a major threat to society because the human ability to recognize fake media from the originals is in contradiction to their quality.

Diffret types of deepfakes



- Audio
- Image
- Video

Figure 2.1: Examples of real and fake attribute manipulation category. Retrieved from [2].

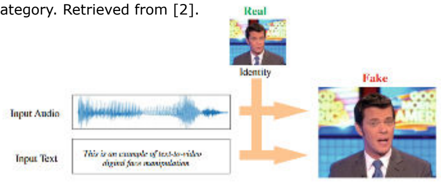

Real
Identity
Input Audio
Input Text
This is an example of text-to-video digital face manipulation
Fake

Figure 2.2: Examples of real and fake audio/text to video fake category. Retrieved from [2]

Detection framework

Architecture

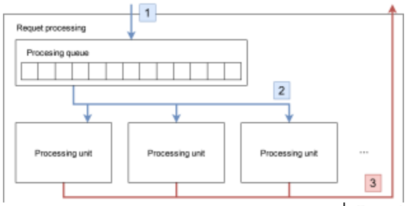


Figure 3.2: Request processing detail

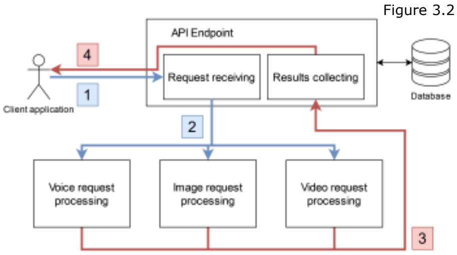


Figure 3.1: High-level design of whole framework

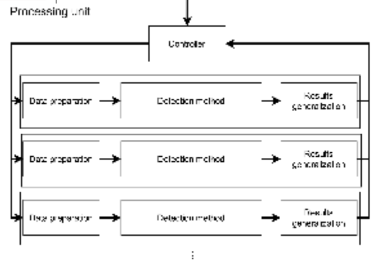


Figure 3.3: Processing unit pipeline

Implementation

The framework was implemented on the previously defined architecture. Processing units can be scaled based on number of waiting messages in processing queue. This allows to process more messages at a time and overall improve performence.

The framework can found on publicly available repository.

Framework is divided into two parts:
- **processing** - API Endpoint, message broker, processing units
- **monitoring** - metric collector, observability platform, etc.

Client application

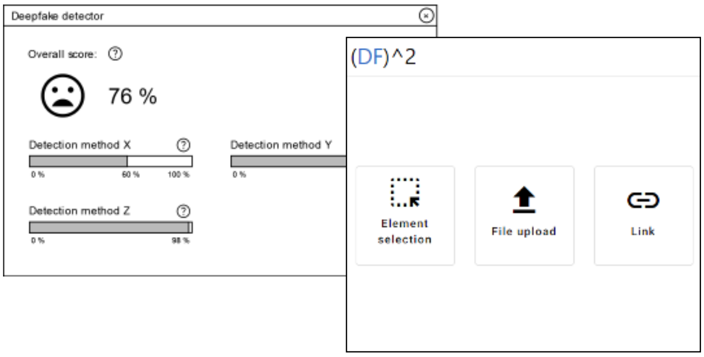


Figure 5.1: Client application wireframes and implementation screens

Results

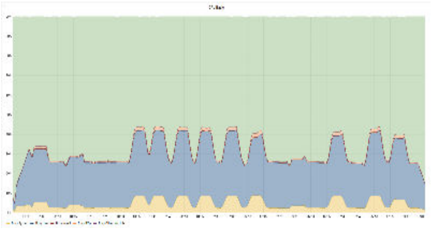


Figure 6.1: CPU load during small bursts test

Three different test scenarios were created to test the reliability of the framework. In graph in figure 6.1 we can see CPU usage during one of test scenario. Overall results were success. The framework is able to handle a large number of files in a relatively short time.

References

[1] Korshunov, P. and Marcel, S. The Threat of Deepfakes to Computer and Human Visions. In: Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks. Springer International Publishing, 2022, p. 97–115. DOI: 10.1007/978-3-030-87664-7_5. ISBN 978-3-030-87664-7. Available at: https://doi.org/10.1007/978-3-030-87664-7_5.

[2] Ibsen, M., Rathgeb, C., Fischer, D., Drozdowski, P. and Busch, C. An Introduction to Digital Face Manipulation. In: Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks. Springer International Publishing, 2022, p. 3–26. DOI: 10.1007/978-3-030-87664-7_5. ISBN 978-3-030-87664-7. Available at: https://doi.org/10.1007/978-3-030-87664-7_5.