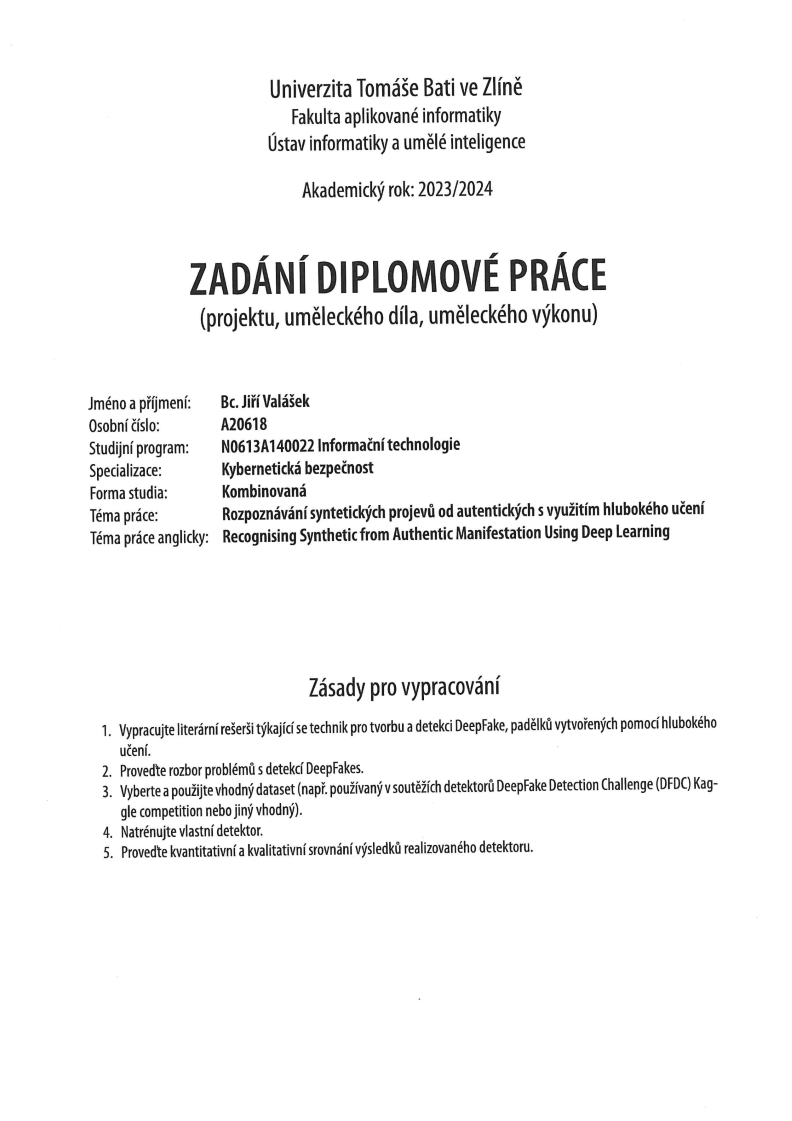
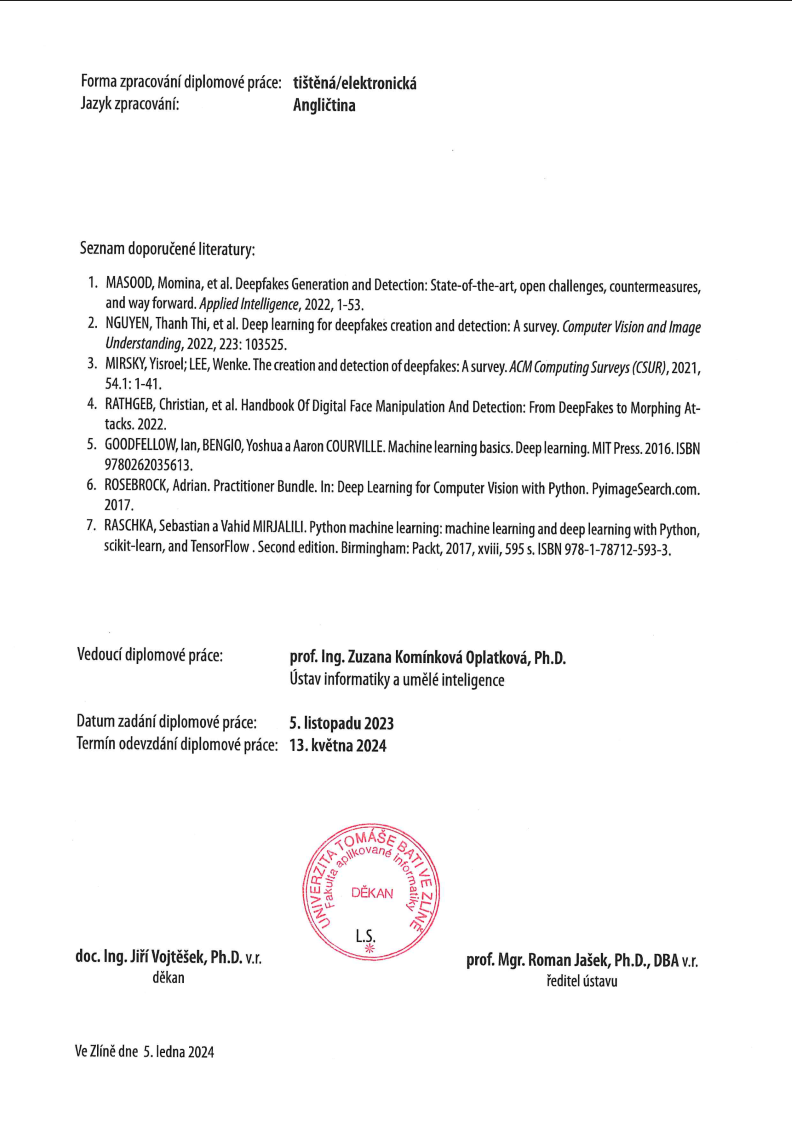
|  |  |
| --- | --- |
|  | |
| Recognising Synthetic from Authentic Manifestation Using Deep Learning | |
|  | |
| Bc. Jiří Valášek | |
|  | |
|  |  |
| 2024 | fai |
|  |  |
|  | |

\*\*\*Scanned submission page 1\*\*\*

\*\*\*Scanned submission page 2\*\*\*

**I hereby declare that:**

* I understand that by submitting my Diploma thesis, I agree to the publication of my work according to Law No. 111/1998, Coll., On Universities and on changes and amendments to other acts (e.g. the Universities Act), as amended by subsequent legislation, without regard to the results of the defence of the thesis.
* I understand that my Diploma Thesis will be stored electronically in the university information system and be made available for on-site inspection, and that a copy of the Diploma/Thesis will be stored in the Reference Library of the Faculty of Applied Informatics, Tomas Bata University in Zlin, and that a copy shall be deposited with my Supervisor.
* I am aware of the fact that my Diploma Thesis is fully covered by Act No. 121/2000 Coll. On Copyright, and Rights Related to Copyright, as amended by some other laws (e.g. the Copyright Act), as amended by subsequent legislation; and especially, by §35, Para. 3.
* I understand that, according to §60, Para. 1 of the Copyright Act, TBU in Zlin has the right to conclude licensing agreements relating to the use of scholastic work within the full extent of §12, Para. 4, of the Copyright Act.
* I understand that, according to §60, Para. 2, and Para. 3, of the Copyright Act, I may use my work - Diploma Thesis, or grant a license for its use, only if permitted by the licensing agreement concluded between myself and Tomas Bata University in Zlin with a view to the fact that Tomas Bata University in Zlín must be compensated for any reasonable contribution to covering such expenses/costs as invested by them in the creation of the thesis (up until the full actual amount) shall also be a subject of this licensing agreement.
* I understand that, should the elaboration of the Diploma Thesis include the use of software provided by Tomas Bata University in Zlin or other such entities strictly for study and research purposes (i.e. only for non-commercial use), the results of my Diploma Thesis cannot be used for commercial purposes.
* I understand that, if the output of my Diploma Thesis is any software product(s), this/these shall equally be considered as part of the thesis, as well as any source codes, or files from which the project is composed. Not submitting any part of this/these component(s) may be a reason for the non-defence of my thesis.

**I herewith declare that:**

* I have worked on my thesis alone and duly cited any literature I have used. In the case of the publication of the results of my thesis, I shall be listed as co-author.
* That the submitted version of the thesis and its electronic version uploaded to IS/STAG are both identical.

In Zlin; dated: .....................................

Student´s Signature

ABSTRAKT

DeepFake tvoří značnou hrozbu pro dnešní společnost. Informace v kyberprostoru mohou mít dosah milionů lidí v řádech hodin. První slovo vždy platí a náprava napáchaných škod je složitá, nákladná a nikdy plně účinná. Proto se tato práce zaměřuje na detekci synteticky upravených nebo generovaných videí.

Klíčová slova: DeepFake, AI,

ABSTRACT

DeepFakes pose a significant threat to current society. At present, information can reach millions in order of hours. The first information is always taken as truth and damage repair is difficult, costly and never fully effective. That is why this work focuses on detection of synthetically edited or generated videos.

Keywords: DeepFake, AI,

Acknowledgements

Acknowledgements, motto and a declaration of honour saying that the print version of the Bachelor's/Master's thesis and the electronic version of the thesis deposited in the IS/STAG system are identical, worded as follows:

I hereby declare that the print version of my Bachelor's/Master's thesis and the electronic version of my thesis deposited in the IS/STAG system are identical.

Contents

[introduction 8](#_Toc160311424)

1. [theory 9](#_Toc160311425)

[1 DeepFake 10](#_Toc160311426)

[1.1 History 10](#_Toc160311427)

[1.2 Categories 10](#_Toc160311428)

[1.2.1 Face-swap 10](#_Toc160311429)

[1.2.2 Lip-sync 10](#_Toc160311430)

[1.2.3 Facial Reenactment 10](#_Toc160311431)

[1.2.4 Face-synthesis 10](#_Toc160311432)

[1.2.5 Attribute-manipulation 10](#_Toc160311433)

[1.2.6 Video generation 10](#_Toc160311434)

[1.3 State of the Art 10](#_Toc160311435)

[1.4 Limitations 10](#_Toc160311436)

[2 DeepFake Detection 11](#_Toc160311437)

[2.1 Approaches 11](#_Toc160311438)

[2.1.1 Temporal Features across Video Frames 11](#_Toc160311439)

[2.1.2 Visual Artifacts vithin a Vide Frame 11](#_Toc160311440)

[2.2 State of the Art 11](#_Toc160311441)

[2.3 Limitations 11](#_Toc160311442)

1. [analysis 12](#_Toc160311443)

[3 Methodology 13](#_Toc160311444)

[3.1 Datasets 13](#_Toc160311445)

[3.1.1 Chosen Datasets 13](#_Toc160311446)

[3.2 Model design 13](#_Toc160311447)

[3.3 Training 13](#_Toc160311448)

[3.3.1 Technical Approach 13](#_Toc160311449)

[3.3.2 Optimization 13](#_Toc160311450)

[4 Results 15](#_Toc160311451)

[5 Discussion 16](#_Toc160311452)

[5.1 Subheading 16](#_Toc160311453)

[Conclusion 17](#_Toc160311454)

[bibliography 18](#_Toc160311455)

[List of abbreviations 19](#_Toc160311456)

[list of figures 20](#_Toc160311457)

[list of tables 21](#_Toc160311458)

[appendices 22](#_Toc160311459)

introduction

Deepfake term is used to describe pictures, video and audio recordings that have been created or edited using deep learning in order to imitate a target person or persons performing given activity. Present-day deepfake content can appear very realistic and easily deceive a large number of the target audience. The means of creating such content are also widely available due to many open-source projects, publicly accessible programs and applications. The usage of deepfakes can be beneficial, but also poses significant risks to security.

Deepfake technology offers many uses to the entertainment industry. It can be used to “bring back to life” historical figures or actors that have passed away. This is how late actress Carrie Fisher was able to posthumously appear in The Rise of Skywalker. Another usage of the technology is in gaming industry where it enables creators to make more realistic environment. However, all these benefits aren’t able to outweigh the dangers of this technology.

The potential to do damage is considerable when this technology can be used to imitate anyone who shares personal media content online. This creates a sizable vulnerability for all users of social networks with public profiles sharing their content with anyone online or on given social network as well as highly publicized persons like celebrities, politicians and state officials. Malicious actors in cyberspace can use publicly available content together with deepfake technology in various ways to cause damage.

https://www.wired.com/story/most-deepfakes-porn-multiplying-fast/#:~:text=Researchers%20worry%20that%20doctored%20videos%20may%20disrupt%20the,finds%20that%2096%20percent%20of%20deepfakes%20are%20pornographic.

* Premise - loads of „public“ images and videos online
  + Dangers of social media & general habits of sharing everything online
* Deep Fake dangers
  + Disinformation
    - Political
      * https://nypost.com/2023/03/22/chilling-deepfakes-claiming-to-show-trumps-arrest-spread-across-twitter/
    - Military
      * https://arxiv.org/pdf/2206.12043.pdf
  + Cybercriminality – extortion, scareware, phishing
    - Examples – phishing commercials with Petr Pavel
    - https://cc.cz/deepfake-v-byznysu-kolega-volal-s-mou-kopii-ctvrt-hodiny-rika-sef-miliardoveho-gymbeamu/
  + Revenge porn
    - Examples – celebrities (not images)
  + Hoaxes
    - Example - Trump arrested
* Easy to use apps
* Possible futures - SORA and similar services
* Approaches to the fight against DeepFakes
  + Verified sources
    - Snaha o boj proti deepfakes: Existuji standardizacni projekty ktere si kladou za cil vyresit problem s autenticitou medii (fotky, videa). Project Origin a CAI (Content Authenticity Initiative), ktere se zabyvaji technickym resenim a C2PA (The Coalition for Content Provenance and Authenticity) jehoz cilem je vytvorit otevreny standard, ktery zajisti fungovani celeho retezce (tvurci, vydavatele, konzumenti) s resenim pro content authenticity.
    - V podstate jde o to, ze samotne zarizeni (kamera, fotak) budou k vyvtorenemu kontentu pridavat data jako datum a cas, pozice, typ fotaku, autor - proste vsechno co uz stejne normalne delaji - a vsechno se to pak vcetne contentu samotneho podepise soukromym klicem vydanym vyrobcem. Dale bude mozne v celem retezci pridavat dalsi data provadet zmeny a dale je podepisovat, mela by se tam vytvorit overitelna historie toho co se s danym videem, nebo fotkou delo a jak.
    - Existuji prvni produkty jako napriklad Leica M11P ktera uz tohle umi a budou nasledovat dalsi. Ve vyvoji je i cip, ktery muze byt vyuzity v android telefonech.https://c2pa.org/about/about/
    - https://www.originproject.info/about
    - https://leicarumors.com/2023/10/26/leica-m11-p-camera-with-content-authentication-and-summicron-m-28mm-f-2-asph-lens-officially-announced.aspx/
  + Detection
* Why is detection important
  + News media
  + Social media
* Stating what is this thesis trying to accomplishext

|  |  |
| --- | --- |
|  | theory |

# DeepFake

* Include teoretical background into AI models as well

## History

* Alexnet, ImageChallenge CNN/DNN classifiers
* Fully Convolutional Networks - https://arxiv.org/abs/1411.4038
* U-NET - <https://arxiv.org/abs/1505.04597>, https://pyimagesearch.com/2022/02/21/u-net-image-segmentation-in-keras/
* Pix2Pix – conditional GAN - <https://arxiv.org/abs/1611.07004>, <https://www.tensorflow.org/tutorials/generative/pix2pix>
* Cycle GAN - https://arxiv.org/abs/1703.10593,https://colab.research.google.com/github/tensorflow/docs/blob/master/site/en/tutorials/generative/cyclegan.ipynb

## Categories

Text

### Face-swap

Face-swap deepfakes is the most prevalent method due to its relative simplicity and usefulness. In this category, generative adversarial networks are used t

### Lip-sync

Text

### Facial Reenactment

Text

### Face-synthesis

Text

### Attribute-manipulation

Text

### Video generation

Text

## State of the Art

Text

## Limitations

Text

# DeepFake Detection

Text

## Approaches

Text

### Temporal Features across Video Frames

Text

### Visual Artifacts within a Video Frame

Text

## State of the Art

Text

## Limitations

Text

|  |  |
| --- | --- |
|  | analysis |

# Methodology

Text

## Datasets

* [ForgeryNet Github](https://github.com/yinanhe/forgerynet)
* [Celeb-DF Github](https://github.com/yuezunli/celeb-deepfakeforensics)
* [DF-W Github](https://github.com/jmpu/webconf21-deepfakes-in-the-wild) – send request via email (Google Form does not work)
* [CDDB Github](https://github.com/Coral79/CDDB)
* [FF++ Github](https://github.com/ondyari/FaceForensics) – aquired compresion 23/40, not enough storage for raw
* [Facebook DFDC](https://ai.meta.com/datasets/dfdc/) – requires AWS ([Transferring data from Amazon S3 to Cloud Storage](https://cloud.google.com/architecture/transferring-data-from-amazon-s3-to-cloud-storage-using-vpc-service-controls-and-storage-transfer-service)
* [DeepFakeTimid](https://www.idiap.ch/en/scientific-research/data/deepfaketimit) - aquired

Text

### Chosen Datasets

* Reasons
* What DeepFake categories are included

## Model design

Description of tested models

## Training

Why cloud needed to be used

### Technical Approach

* [Artificial Intelligence and Machine Learning Capabilities and Application Programming Interfaces at Amazon, Google, and Microsoft](https://dspace.mit.edu/handle/1721.1/146689)
* Why Google Cloud Platform was selected
* How training was done on GCP

### Optimization

* Optimization algorithm
* How dataset was used for training, testing and evalution
* Be careful to avoid methodological errors of using the same datapoints

# Results

Text

# Discussion

Text

## Subheading

Text

Conclusion

Text

bibliography

[1] Text

List of abbreviations

|  |  |  |
| --- | --- | --- |
| ABC |  | First abbreviation – meaning |
| B |  | Second abbreviation - meaning |
| C |  | Third abbreviation - meaning |
|  |  |  |

list of figures

**Nenalezena položka seznamu obrázků.**

list of tables

**Nenalezena položka seznamu obrázků.**

appendices

appendix p i: appendix title