A Project Report

on

Detection of Deepfake in video

Submitted in partial fulfillment of the

requirement for the award of the degree of

Bachelor of technology

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Under The Supervision of

Mr Surender Singh Chauhan

Submitted By

Shubham Dubey
19SCSE1010423

Rahul Kumar
19SCSE1010286

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING / DEPARTMENT OF COMPUTERAPPLICATION

GALGOTIAS UNIVERSITY, GREATER NOIDA

INDIA

MONTH, YEAR

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING
GALGOTIAS UNIVERSITY, GREATER NOIDA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the thesis/project/dissertation, entitled **"CAPS…."** in partial fulfillment of the requirements for the award of the B.tech submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of 4 month, Year to Month and Year, under the supervision of Mr Surendar Singh Chauhan, Department of Computer Science and Engineering/Computer Application and Information and Science, of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the thesis/project/dissertation has not been submitted by me/us for the award of any other degree of this or any other places.

Shubham Dubey(19SCSE1010423)

Rahul Kumar (19SCSE1010286)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Surendra Singh Chauhan

CERTIFICATE

The Final Project Viva-Voce examination of Shubham Dubey(19SCSE1010423) and Rahul Kumar(19SCSE1010286) has been held on 23 Dec 2021 and his/her work is recommended for the award of B.tech.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator

Signature of Dean

Date: December 22, 2021

Place: Greater Noida

Table of Contents

Acknowledgement

Place:Galgotias University

Date:22 Dec 2021

I would like to express my sincere thanks to Mr. Surendra Singh Chauhan , for his/her valuable guidance and support in completing my project.

I would also like to express my gratitude towards our Dean Mr. S.P.S Chauhan for giving me this great opportunity to do a project on Deepfake Detection in Video. Without their support and suggestions, this project would not have been completed.

Shubham Dubey

Rahul kumar

Signature.

ABSTRACT

Deep learning is an effective and useful technique that has been widely applied in a variety of fields, including computer vision, machine vision, and natural language processing. Deepfakes uses deep learning technology to manipulate images and videos of a person that humans cannot differentiate them from the real one.

In recent years, many studies have been conducted to understand how deepfakes work and many approaches based on deep learning have been introduced to detect deepfakes videos or images. In this paper, we conduct a comprehensive review of deepfakes creation and detection technologies using deep learning approaches. In addition, we give a thorough analysis of various technologies and their application in deepfakes detection. Our study will be beneficial for researchers in this field as it will cover the recent state-of-art methods that discover deepfakes videos or images in social contents. In addition, it will help comparison with the existing works because of the detailed description of the latest methods and dataset used in this domain.

Image and video forgery are posing a threat to the society in today's world. People can artificially create any audio or video clip. Artificial intelligence, mainly machine learning, manipulates images and videos in such a way that they are often visually indistinguishable from real ones . There are some prevalent techniques which are widely used to manipulate images/videos. Some are computer graphic based (e.g. Photoshop, GIMP, and Canva) and the rest are content changing. Deepfake, a deep learning-based method, is a serious contender among the content changing video falsification techniques. The term "deepfake" originates from the words "deep learning" and "fake". Use of deep learning networks (DNN) has made the process of creating convincing fake images and videos increasingly easier and faster. It is a technique in which a video or image of a person is manipulated by the image of another person using deep learning methods.

Introduction

Recently, a huge step towards automated video editing has been made: technical knowledge is now not required when preparing fake content that presents something that did not even occur. What caught the public attention in 2017 was the reveal of many fake videos on Reddit (1). The huge amount of misleading content made people realize how serious the problem was.

Fake video is usually addressed to as deepfake (a term coined by a Reddit user). The definition of deepfake is very simple: it's an AImanipulated video. It presents realistic, but not real content – usually people doing and saying fabricated things.

This technique is especially known from superimposing somebody's face over the background reality this person has never experienced. One of the most prominent example is the video of President Barack Obama saying things he never said (2). It should be noted that not only famous people are exposed to the deepfake technology (3). In the word full of fake news, the potential of harmful use of deepfake techniques is horrific (4).

Image frauds before the deepfake era were not so widespread. Back then, people committing these type of frauds needed domain expertise and some tool mastery – they had to know CGI or Photoshop. In the case of deepfakes, apps for creating forged videos are easily available, e.g. Deepfakes web β (5), FaceApp (6) and the Chinese app Zao (7). They allow anyone with a minimum knowledge of the topic to create deepfakes and disseminate them on the internet. In response to the increase of more and more sophisticated and realistic fake videos, the solution that helps everyone better assess the misinformation of the online content is of paramount importance.

The spread of fake news created by deepfakes (8) has inspired the group of global technology leaders like AWS, Facebook and Microsoft to come together and solve this problem. With the Partnership on AI's Media Integrity Steering Committee and several universities including Oxford,

Berkeley, and MIT, they have built the biggest contest on detecting AI-forged videos. Called Deepfake Detection Challenge, it's hosted on the Kaggle platform (9). The goal of the challenge is to prepare the technology that helps everyone better assess the legitimacy of content they see. An important feature of this technology, as emphasized by the PAI Steering Committee, is its availability – the algorithms should be as accessible and useful as possible for people working with news pieces on a daily basis.
The remainder of this article is organized as follows. Section II reports literature overview as well as a description of Deepfake Detection Challenge. In section III, details about the proposed approach are introduced. The inference procedure is described in section IV. Results in section V and conclusions from section VI conclude this paper

State of the art

2.1 Deepfake video generation

The origin of deepfake technique can be found in GANs, short for Generative Adversarial Networks, that were introduced in a paper by Ian Goodfellow and other researchers at the University of Montreal in 2014 (10). They have proposed a new framework (Fig. 1) for estimating generative models via an adversarial process. In it, two models are being simultaneously trained: a generative model

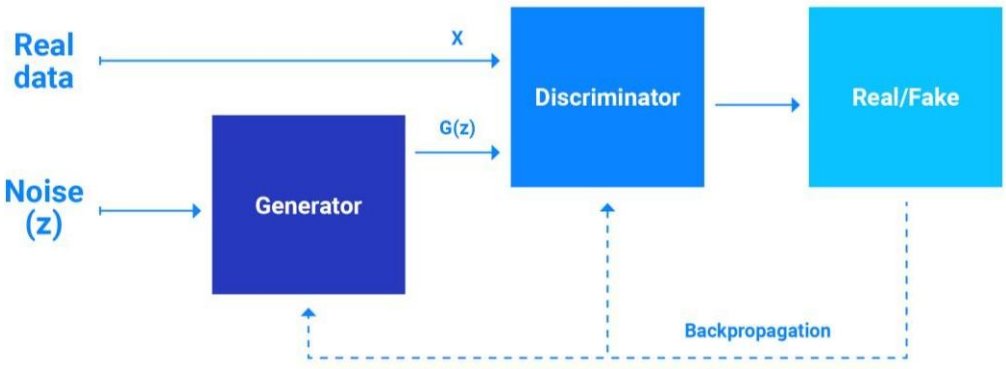


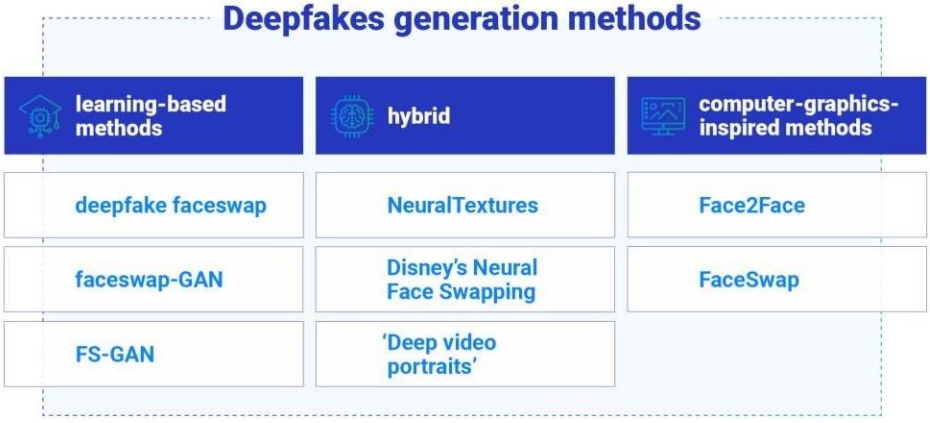
Figure 1: An idea of GANs learning

Considering mathematical foundations, in the space of arbitrary functions.

The GANs introduced by Goodfellow were the first of their kind and of course had their flaws. Just after the introduction of GANs, a huge interest was paid to this type of AI algorithms, which resulted in the

proliferation of GAN-based methods: ProGANs (11), StyleGAN (12),CycleGAN (13), StarGAN (14), IcGAN (15), AttGAN (16), STGAN (17), and MaskGAN (18), among others. The code of most GAN-based methods is publicly available and the number of developed methods delights and terrifies at the same time (19). GANs can be used for a number of helpful applications, in the fashion industry (20), medical imaging (21), or even drug discovery (22), but what has caught the public attention is the use of GANs to create deepfakes.

Unfortunately, the misuse of this technology overshadows its advantages. The problem of falsifying media content, while not new, was quite scarce – up until the AI developers have introduced the deepfake technology. Early face swaps were produced in 2016, when this article (23) was published showing how to swap faces using open source OpenCV library. But this technique only applied to images, not videos. The term deepfake referring to videos was coined in 2017 when a group from the Reddit community shared their first fake videos (1). Deepfake techniques can be divided into 2 main groups (Fig. 2): learning-based methods and computer-graphics-inspired approaches. Examples of learning-based methods are not limited to: Deepfakes Faceswap (best known for Nicolas Cage face-swaps, e.g. (25)), Faceswap-GAN (26) or FS-GAN. Face2Face and another FaceSwap (29) are examples of computer-graphics-based approaches. Hybrid approaches also exists. Astonishing results are obtained in animating still images.



Just after the deepfake technique introduction in 2017, plenty of easily available apps that do not require any technical knowledge appeared on the internet: FakeApp (6), ZAO (7), Deepfakes web β (5) or FaceApp (35). The majority of fake content generation code is publicly available on GitHub ((24), (26), (36), (37), (38)). The majority of AI-generated fake content is some kind of facial manipulation. According to (39) facial

manipulations could be divided into 4 main groups (Fig. 3): entire face synthesis (creating high-quality non-existent facial images), identity swap (replacing somebody's face with that of another person), attribute manipulation (editing selected parts of the face, e.g. adding glasses) and expression swap (aka face re-enactment, replacing somebody's expression with the expression of another person). Identity swap and face re-enactment are carried out at the video level.

It should be noted that AI-generated deceptive content is not just limited to face modification. Apps for modifying image content, e. g. landscapes ((40), (41)) or removing content (42) are also easily available.

2.2 Deepfake video detection methods

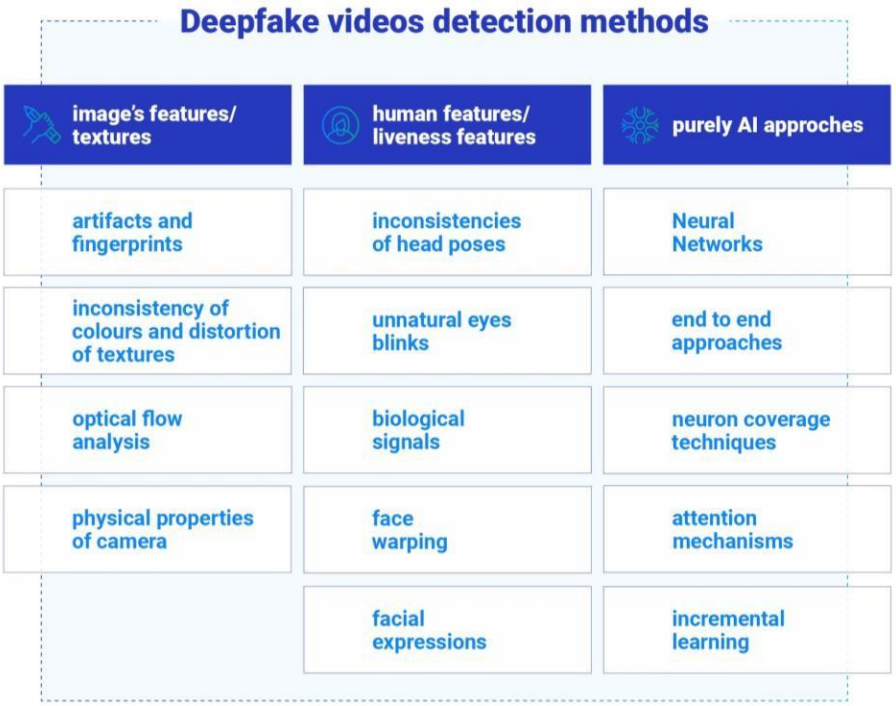
The idea of detecting deepfakes arised as soon as the deepfakes took the form that's undetectable by the human eye. Image-based detection of deepfake videos can be divided into 3 main groups (Fig. 4), based on the nature of detection method: 1. image features/textures, 2. human features / liveness features, 3. purely AI approaches.

Video is a sequence of images enriched with audio. It must be noted, that even though a video is a sequence of images, not all imagebased detection techniques could be applied to investigate the video forgery. Some imagebased detection techniques rely on high-quality photos that are not seen in videos due to compression.

The authors of publications devoted to the first group of methods – image features and textures – learned from the experience of researchers who studied image forensics before the deepfake era. That is why the basis of those methods lies in the pixel level. Authors developing those methods paid special attention to artefacts ((43), (44)), fingerprints ((45), (46)), color inconsistencies ((47), (48)), texture distortions (49) and optical flow analysis (50) or even physical properties of the camera (51).

The second group includes biologically inspired methods. The idea behind those methods is to pay attention to physical/physiological aspects of the image analysing human features or liveness features. The authors found that in deepfake videos there are inconsistencies of head poses (52), unnatural eye blinks ((53), (54)), biological signals are not preserved (55) and some face warping (56) can be noticed. There is also a method using the authentication of facial expressions (57), but the application of this method is only limited to famous people about whom many videos have been made (58). The authors of (59) showed a method where the combination of facial and behavioural biometrics is considered.

Purely AI approaches from the third group of detection methods don't take into account any hand-crafted or observed features. Especially suited neural networks are the root of the majority of purely AI approaches: ensembling of Convolutional Neural Networks (60), combination of Convolutional Neural Networks and Recurrent Neural Networks ((61), (62), (63)), Spatiotemporal Convolutional Networks (64), 3D Convolutional Neural Networks (65) or MesoNet (66). Other AI approaches also can be found: Capsule networks ((67), (68)), end-toend approaches (69), neuron coverage techniques (70), attention mechanisms (71), incremental learning (72) or multi-task learning (73).



A comparison of all of the mentioned detection techniques is fairly difficult because the authors of those methods used different evaluation metrics and also performance was measured on different databases. Nevertheless, a comparison attempt has been made in (39) and the conclusion is that AI-based methods perform the best. The methods for audio-based fake detection are also investigated by researchers.

2.3 Deepfake Detection Challenge

The Challenge (9) was founded by the collaborative initiative of global technology leaders such as AWS, Facebook, Microsoft, the Partnership on AI's Media Integrity Steering Committee, and academic partners. It was hosted on the Kaggle platform.

The first announcement of the challenge took place at the end of 2019. Preview dataset was released in December 2019. Full training and validation dataset was published in January 2020, and the deadline for the model submission was March 31, 2020. After the deadline, all the solutions were tested against the hidden test dataset, and the final leaderboard was published on June 12, 2020.

The challenge was open to anyone – not only researchers but also AI developers and enthusiasts.

The participants of the Deepfake Detection Challenge were equipped with training and validation datasets to build their models on. Then, the participants were asked to submit their code to a black box environment for testing. A public leaderboard was calculated throughout the competition. After the deadline, all the models were examined the second time against another, hidden test set. Based on those results, the final leaderboard was published.

Competitors could take two distinct paths: open or closed. The open path was for those who wished to share their code publicly, provided that their solutions abode by the open source licensing terms. There were prizes for 5 best solutions. Institutions that wanted to retain the ownership rights to their solutions, chose the closed path with no prizes involved.

Over 2 thousands AI enthusiasts from all around the word took part in the competition. More than 35,000 models were prepared.

After the revision of the existing deepfake detection methods, it was decided to choose purely AI-based approach. Owing to the fact that the labels for each video were provided in the dataset, the supervised learning approach was chosen. Promising results of Xception net utilization ((75), (76)) resulted in choosing this particular solution. For the sake of competitiveness, the decision was made to prepare an algorithm in accordance with Kaggle's Deepfake Detection Challenge terms.

3 Proposed method

Video is considered a combination of sequence of images and audio, so two types of content were extracted: snapshot images and audio.

At the beginning of the works, early-stage experiments were performed. Because of the limitations due to data labelling and not promising early-stage results, it was decided for the audio consideration to be abandoned, so the emphasis could be placed on fake image investigations.

Deepfake detection is understood here as a frame-level binary classification problem, where each video is classified as real or a fake. Only a classification of the video being real or fake should be delivered; the bounding boxes of the fake content don't have to be found.

3.1 Dataset

A unique dataset with a variety of high-quality videos was prepared for the competition (77). Remarkable academic institutions were involved in the preparation of the contest dataset. They have created the biggest dataset so far with both real (20%) and fake (80%) videos. The dataset was comprised of 124k videos (over 470 GB of .mp4 videos). Before the DFDC dataset there were only medium-sized sets of manipulated videos (e.g. much smaller Google's dataset (78), UADFV (52), DeepfakeTIMIT (79), FaceForensics++ (76), Celeb-DF (80) among others) not big enough to evaluate and improve computer vision algorithms.

The dataset was prepared just for the purpose of this challenge. A group of paid actors was involved in this database production. The actors signed a consent that their image would be manipulated. Fake videos available in the dataset were prepared using 8 facial modifications algorithms. The

authors of the database did their best to ensure diversity in terms of age, gender, skin tone and other characteristics. A Facebook AI blog post (81) reads:

"We altered the videos using a variety of different deepfake generation models,
refinement techniques such as image enhancement, and additional augmentations and distractors,
such as blur, frame-rate modification, and overlays.
Our goal was to make the data set representative of the variety of qualities and adversarial methods that could occur in real-world videos shared online"

Figure 5: Sample snapshots of faked videos

Dataset split

According to (82), within the whole database, we can specify 4 sets of videos:

Training set: public, data for training, videos and appropriate labels, training corpus of 115,000 videos.

Validation set: public: 400 videos, used for validation during training.

"Public" test set: withheld, used to compute scores for public leaderboard.

“Private” test set: privately held outside of the Kaggle platform, distinct dataset used to compute the private leaderboard, not shared with entrants, comprising of real world examples (with types of fake videos not included in the training dataset).

The purpose of providing two versions of the test dataset was to ensure that algorithms were robust enough. One of the major problem in training AI models is they fit training and validation data too much. Because of that, the models are not able to generalize from known examples to assess unfamiliar instances. The Facebook AI blog post talks about the nature of the dataset:

“One of the central unsolved challenges of detecting deepfakes is that it is hard to generalize from known examples to unfamiliar instances. We designed the DFDC with this in mind. To determine the winners, participants in the challenge submitted their code to a black box environment. This separate data set was not available to entrants, so they had to design models that could be effective even under unforeseen circumstances. The black box data set consisted of 10,000 videos that were not available to participants in the competition. It contains both organic content (both deepfakes and benign clips) found on the internet and new videos created for this project. We verified that the distribution of fake and real videos was identical to that of the public test set.”

From the perspective of the competitors, one can state the following pros and cons of the provided dataset

Dataset advantages:

The biggest dataset with deepfake video.

In case of fake videos, labels provided information on which video is original.

Dataset limitations:

The dataset is highly imbalanced.

Only 20% of all videos is labelled as real.

Only real/fake labels: one does not know if the video is altered on image or on audio.

No information about which algorithm was used to generate each video
Only limited types of deepfake – classifier overfitting.

3.2 Data pre-processing

Frame extraction

Frames were extracted from all the provided videos. Due to dataset imbalance (the dataset consists of 80% fake videos and 20% real videos), the number of frames extracted from the video depended on whether it was a fake or real video. To have a more balanced dataset, 12 frames were taken from each real video and 3 frames from fake videos. Not all the frames were extracted successfully, therefore the real / fake frames ratio was close to 2:3.

Face extraction

Deepfakes in the DFDC database were only limited to fakes on faces. It was decided then that the algorithm will be fed with face images extracted from the frames of original videos.

The pre-trained face extraction models were revised: OpenCV built-in face detector (83), FaceNet (84) model and MTCNN (85) face detector. OpenCV built-in face detector was chosen, because it is an open-source solution (competition requirements) having a good speed and accuracy performance.

Three main problems arised when extracting faces and building the dataset:

There were multiple false positives – objects that were wrongly detected as a face of an actor (e.g. depicting hands, portrait painting). An important part of the dataset creation stage was going through retrieved images and deleting the ones that were false positives.

If the deepfake quality was poor, the face was so distorted, that was not detected.

Some parts of deepfake videos are uneven: we can see original faces in some parts and fake ones in other parts. As the dataset consisted of way too many videos to be verified by a human, some of the faces that were labelled in the dataset as fake could actually represent a real face

Train, valid and test dataset determination

Due to the fact that the same actor could be found in multiple videos, and number of fake and real videos were not balanced, selecting appropriate data for the training, validation and test sets was an important part of the project.

Careful frame and face extraction described above was crucial in obtaining properly distributed datasets; in particular, the training dataset was not burdened by the class imbalance.

Due to the fact that training was performed on extracted frames, train, valid and internal test sets were built using video frames from the Kaggle training dataset.

3.3 AI approach

Xception net

Xception net (86) is an architecture selected to train binary classifier predicting the probability of the detected face being a deepfake. Xception is a deep convolutional neural network developed by Google. Its architecture consists of 3 parts: entry, middle and exit flow.

The characteristic parts of this network are depthwise separable convolutions and the “shortcut” connections between them. Xception turned out to outperform other popular architectures such as VGG (87), ResNet (88) and Inception (89) on ImageNet dataset.

Model ensamble

The model ensemble is a method of improving predictive performance that could be achieved by one model using the results of multiple models to get the final prediction. It is based on the idea that different models can have strengths in different areas, and combining them can give the best results. In this competition 3 Xception models were combined by taking the average of their predictions. These models were different by the number of epochs they were trained on, different dropout rates used, as well as slightly different selection of the training data.

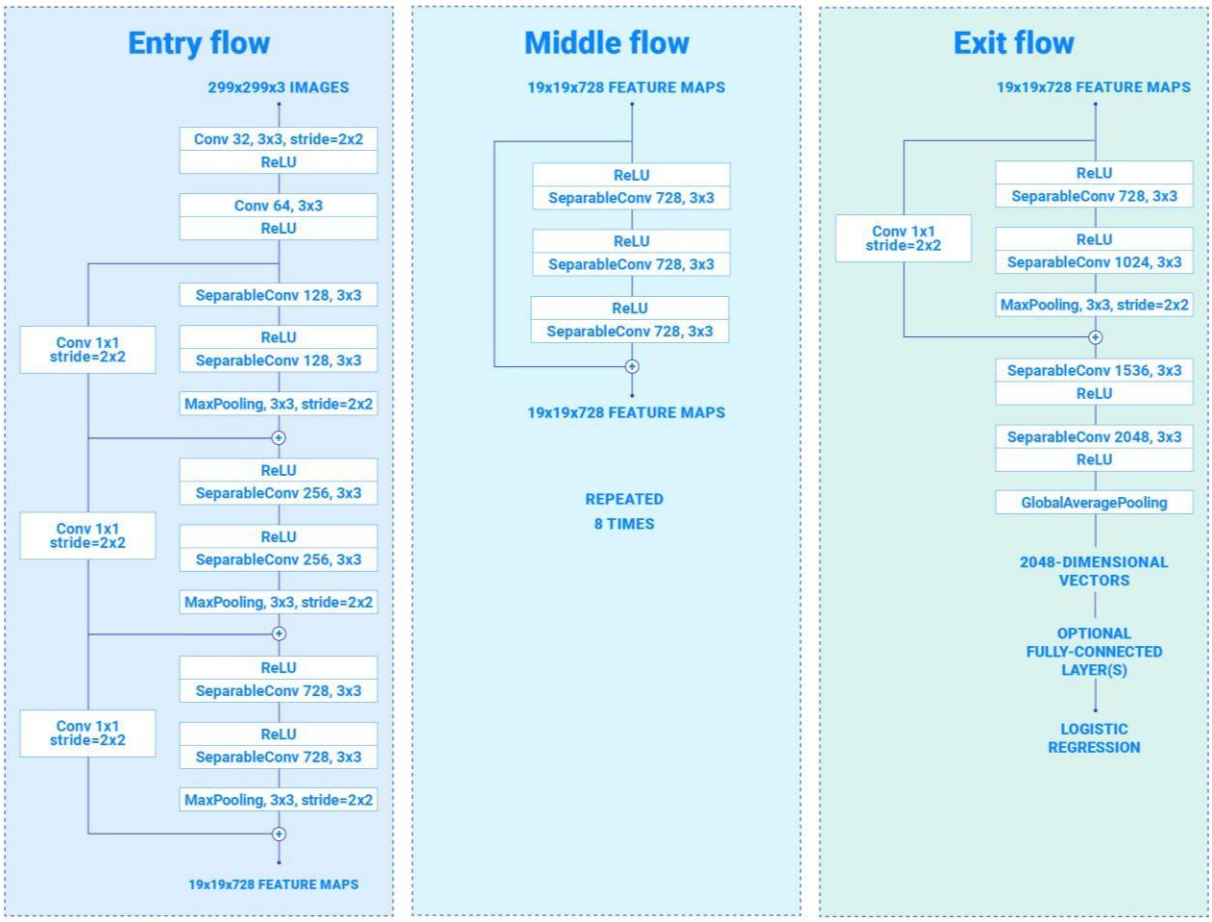


Figure 6: Xception net structure

4 Experiments

Infrastructure

Company's own computing resources were used in the training procedure. The training was performed on 4 Nvidia cards (GeForce GTX 1080 Ti). Single training on one card typically lasted 20 hours. With the available infrastructure, more than one model training was performed at a time.

Inference pipeline

The inference pipeline simplified block diagram is shown in the Fig. 7.

A detailed description of each step is provided.

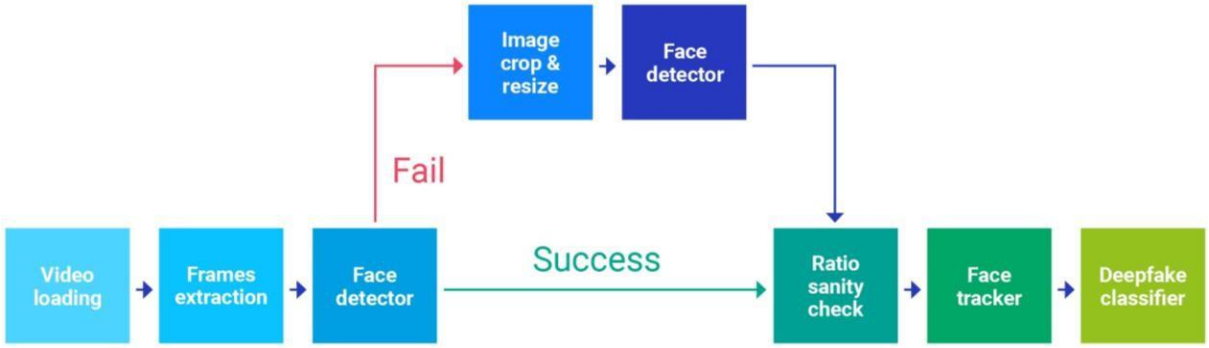


Figure 7: Inference pipeline

Frame extraction: the created framework extracts 30 equally distributed frames from the video. This number was selected not to exceed the limit of 9 hours for the inference of the testing set.

Face detector: as previously mentioned, face detection was performed using pre-trained deep neural network detector available in the OpenCV package.

Image crop & resize: this block of the framework is used in case of face detector failing to find a face in the picture – all the objects found in the picture have less than 50% of probability to be a face. One of the reasons could be the size of the face in the picture (e.g. an actor was standing too far from the camera). To solve this issue, it was decided to crop sides on the picture, where the face is least likely to be present, and then enlarge the picture. After this step there were two pictures: one being cropped at the bottom and on the sides, and the other one cropped at the top and on the sides.

Ratio sanity check: another framework block used to mitigate the problem with face detector wrongly recognizing certain objects as faces. This part ignores all the objects whose height to width proportions are between 1 and 3. All the detected objects with the ratio bigger than 2 and the probability of being a face lower than 60% are also ignored.

Face tracker: face tracking is performed by IoU (Intersection over Union) matching between the bounding boxes from two consecutive frames. IoU is a ratio of overlap area to object union area, as shown in Fig. 8 below:

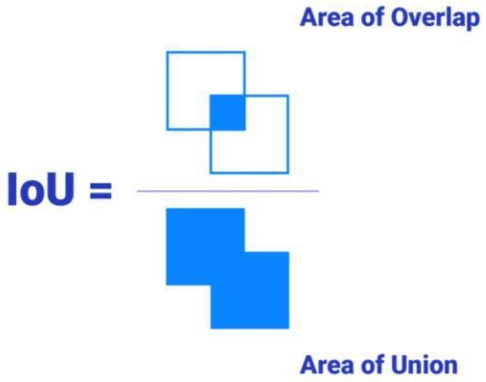


Figure 8: Intersection over Union (IoU)

In this pipeline, after tests, the IoU was set to 0.1. Tracking sequences with fewer than 2 faces were ignored and considered a false detection, and therefore these detected objects were ignored.

Deepfake classifier: All the extracted face images that successfully went through all the steps were fed to 3 Xception models. The probability of face detection was calculated as the average of probabilities predicted by these models. If there were more than 2 actors in one frame, we calculated the frame probability, which was the average of face probabilities in the frame. To finally decide whether a video is a deepfake or not, previously calculated frame probabilities from this video were averaged, which gave the final prediction of the model for the movie.

5 Results

The performance metrics defined for this competition was LogLoss. LogLoss is defined as:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Binary Cross-Entropy / Log Loss

where:

n is the number of videos being predicted,

Y^i is the predicted probability of the video being fake,

Yi is 1 if the video is fake, 0 if real,

Log() is the natural (base e) lagorithm.

The model developed for this challenge is based only on the images, it does not contain any audio information. An attempt to detect audio alternation was done at the beginning of the research, but did not produce satisfactory .The Public Score of the pipeline was 0.41802, whereas the Private Score (this one matters) was 0.52651 **placing the described AI approach in the top 7% of the proposed solutions.**

The final assessment of all the models' performance was done on a dataset consisting of data that could significantly differ from the training data. It results in a significant performance drop of the majority of the algorithms, but not the one presented in this paper.

Limitations

The model developed for this challenge is based only on the images, it does not contain any audio information. An attempt to detect audio alternation was done at the beginning of the research, but did not produce satisfactory results. This was also the effect of the dataset limitation – after the competition had ended, it turned out that „voice fakes" never appeared in the Public Dataset – therefore the lack of fake voice recognition did not hurt the overall performance of the model in the competition.

One of the requirements of the competition was that the inference jupyter notebook predicting 4000 videos can't run longer than 9 hours – without this restriction, the pipeline could have been easily tweaked to get better final results. The model developed for this challenge is only suited for the existing types of deepfakes. It will not cope with upcoming new deepfake technologies. This limitation is due to the selection of supervised learning

– the fact that “AI is only as good as the data that goes into it” (90) had to be accepted. If new type of deepfake appears, this model must be retrained to manage them.

6 Conclusions

Using a purely AI-based approach in deepfake detection has brought the expected results. The algorithm correctly classifies videos, is lightweight and can be implemented into more sophisticated forgery tools.

After the end of the competition it turned out that none of the top-performing solutions used standard or biologically-inspired approaches. This suggest that AI-generated content may be battled only by AI solutions, which was the idea behind choosing the XceptionNet.

Future works may be devoted to the incorporation of audio content to forgery detection (using data outside the DFDC dataset) and the pipeline adjustment (without any restrictions from the contest).

Notable achievement

The approach of altered video detection described in this paper was placed among the best 7% of Kaggle Deepfake Detection Challenge competitors, which means it was awarded a bronze medal. It was the first time deepfake problem had been brought up in such a big challenge. The method presented in this paper became one of the state-of-the-art methods and is now a baseline for future works.

The model developed for this challenge generalizes well to unknown samples. This fact is worth to be highlighted because even some of top-performing algorithms from the public leaderboard were not so good when making evaluations against a hidden dataset.

7 References

Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., and Li,H. (2019, June). Protecting world leaders against deep fakes. In *Computer Vision and Pattern Recognition Workshops* (pp. 38-45).

Tewari, A., Zollhoefer, M., Bernard, F., Garrido, P., Kim, H., Perez, P., and Theobalt, C. (2020). High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2), 357-370.

Lin, J., Li, Y., & Yang, G. (2021). FPGAN: Face de- identification method with generative adversarial networks for social robots. *Neural Networks*, 133, 132-147.

Liu, M. Y., Huang, X., Yu, J., Wang, T. C., & Mallya, A. (2021). Generative adversarial networks for image and video synthesis: Algorithms and applications. *Proceedings of the I EEE*, DOI: 10.1109/JPROC.2021.3049196.

Lyu, S. (2018, August 29). Detecting 'deepfake' videos in the blink of an eye. Available at http://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072

Bloomberg (2018, September 11). How faking videos became easy and why that's so scary. Available at https://fortune.com/2018/09/11/deep-fakes-obama-video/

Chesney, R., and Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Affairs, 98, 147.

Hwang, T. (2020). Deepfakes: A Grounded Threat Assessment. Centre for Security and Emerging Technologies, Georgetown University.

Zhou, X., and Zafarani, R. (2020). A survey of fake news: fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, DOI: https://doi.org/10.1145/3395046.

Kaliyar, R. K., Goswami, A., and Narang, P. (2020). Deepfake: improving fake news detection using tensor decomposition based deep neural network. *Journal of Supercomputing*, DOI: https://doi.org/10.1007/s11227-020-03294-y.

Guo, B., Ding, Y., Yao, L., Liang, Y., and Yu, Z. (2020). The future of false information detection on social media: new perspectives and trends. *ACM Computing Surveys (CSUR)*, 53(4), 1-36.

Tucker, P. (2019, March 31). The newest AI-enabled weapon: 'Deep-Faking' photos of the earth. Available at https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/

Fish, T. (2019, April 4). Deep fakes: AI-manipulated media will be 'weaponised' to trick military. Avail- able at https://www.express.co.uk/news/science/1109783/deep-fakes-ai-artificial-intelligence-photos-video-weaponised-china

Marr, B. (2019, July 22). The best (and scariest) examples of AI-enabled deepfakes. Available at https://www.forbes.com/sites/bernardmarr/2019/07/22/the-best-and-scariest-examples-of-ai-enabled-deepfakes/

Zakharov, E., Shysheya, A., Burkov, E., and Lempitsky, V. (2019). Few-shot adversarial learning of realistic neural talking head models. *arXiv preprint* arXiv:1905.08233.

Damiani, J. (2019, September 3). A voice deepfake was used to scam a CEO out of $243,000. Available at https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

Samuel, S. (2019, June 27). A guy made a deepfake app to turn photos of women into nudes. It didn't go well. Available at https://www.vox.com/2019/6/27/18761639/ai-deepfake-deepnude-app-nude-women-porn