Project Proposal

CS 577

Sep 28, 2023

DeepFake Detection

Authors

Ana Tomash
Shorouq Alasbal
Norah Alamri

Problem description

Deepfake technology has emerged as a groundbreaking project in artificial intelligence, enabling the synthesis of hyper-realistic, computer-generated audio and video content that can convincingly mimic real human expressions and voices. While deepfakes can be used for harmless purposes, such as entertainment or parody, they can also be used for malicious purposes, such as spreading misinformation or damaging someone's reputation. As deepfake technology continues to improve, it is becoming increasingly difficult to distinguish between real and fake videos.

Project Goal

This project aims to develop and apply deep-learning techniques to effectively detect deepfake videos. The model will be trained on a dataset of real and fake videos, and evaluated on its ability to distinguish between the two accurately.

Literature Review

Deepfake technology relies on sophisticated deep-learning algorithms, particularly generative adversarial networks (GANs) and deep neural networks (DNNs), to generate convincingly realistic synthetic content, such as images and videos. These algorithms are well-known for their capacity to mimic human features, expressions, and voices with astonishing accuracy.

Moreover, there are some of the key techniques employed in the "FakeDeep" project:
Deepfake Detection Algorithms: These algorithms are designed to identify the presence of deepfake content within multimedia, such as images, videos, and audio recordings. Common approaches include:

- Convolutional Neural Networks (CNNs): Utilized for image and video analysis, CNNs can learn patterns and anomalies in visual data that may indicate deepfake manipulation.[3]

- Recurrent Neural Networks (RNNs): Applied to analyze sequential data, RNNs are useful for detecting manipulated audio or video sequences.[4]

- Capsule Networks: These networks can capture hierarchical relationships in data, which can be beneficial for recognizing inconsistencies in deepfake content.[5]

Preliminary plan (milestones)

In this project, we aim to compare the performance of these deep-learning techniques and evaluate their performance objectively. To achieve this, we will follow this primary plan:

Milestone 1: (2 weeks)

- Review the literature on deepfake detection and identify the most promising approaches.
- Select a deep learning framework and dataset for the project.

Milestone 2: (2 weeks)

- Model Implementation: Implementing and fine-tuning the deep-learning models associated with deepfake detection.
- Data Preprocessing: involves resizing, cropping, normalizing, and augmenting the data to make it consistent and standardized for deep learning training and evaluation.

Milestone 3: (4 weeks)

- Training and Validation: Training each deep-learning model using the appropriate dataset, optimizing hyperparameters, and validating the models against relevant metrics.
- Accuracy Assessment: Comparing the performance of each model by assessing the accuracy of each model against the true values. .

Milestone 4: (2 weeks)

- Results Analysis: Analyzing the results obtained from the accuracy assessments and drawing conclusions about which technique exhibits the highest accuracy.
- Documentation and Reporting: Documenting the entire process, including data selection, model implementation, training, and evaluation, and preparing a comprehensive report.

In summary, the project plan aims to objectively compare the performance of different deep-learning techniques for deepfake generation by implementing and evaluating the models on a diverse dataset of real and synthetic content. The findings will be used to make recommendations for the most suitable deep-learning technique for specific applications.

References

1. Rössler, Andreas, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. "Faceforensics: A large-scale video dataset for forgery detection in human faces." arXiv preprint arXiv:1803.09179 (2018).
2. Hsu, Chih-Chung, Chia-Yen Lee, and Yi-Xiu Zhuang. "Learning to detect fake face images in the wild." In 2018 international symposium on computer, consumer and control (IS3C), pp. 388-391. IEEE, 2018.
3. Bonettini, Nicolo, Edoardo Daniele Cannas, Sara Mandelli, Luca Bondi, Paolo Bestagini, and Stefano Tubaro. "Video face manipulation detection through ensemble of cnns." In 2020 25th international conference on pattern recognition (ICPR), pp. 5012-5019. IEEE, 2021.
4. Güera, David, and Edward J. Delp. "Deepfake video detection using recurrent neural networks." In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), pp. 1-6. IEEE, 2018.
5. Nguyen, Huy H., Junichi Yamagishi, and Isao Echizen. "Capsule-forensics: Using capsule networks to detect forged images and videos." In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2307-2311. IEEE, 2019.
6. Rana, Md Shohel, Mohammad Nur Nobi, Beddhu Murali, and Andrew H. Sung. "Deepfake detection: A systematic literature review." IEEE access 10 (2022): 25494-25513.