Exposing DeepFake Videos By Detecting Face Warping Artifacts

# Yuezun Li; Siwei Lyu

## 2018

We describe a new deep learning based method that can effectively distinguish AI-generated fake videos from real videos

# Scholarcy Highlights

* The number of fake videos and their degrees of realism have been limited by the lack of sophisticated editing tools, the high demand on domain expertise, and the complex and time-consuming process involved
* We evaluate our four models on each frame of all videos based on Area Under Curve (AUC) metric, where the performance of VGG16, ResNet50, ResNet101 and ResNet152 models on lower quality (LQ) and higher quality (HQ) video sets are 84.6%, 99.9%, 97.6%, 99.4% and 57.4%, 93.2%, 86.9%, 91.2% respectively, see Figure 6 and Figure 7
* We describe a new deep learning based method that can effectively distinguish AI-generated fake videos (DeepFake Videos) from real videos
* Our method is based on the observations that current DeepFake algorithm can only generate images of limited resolutions, which are needed to be further transformed to match the faces to be replaced in the source video
* Such transforms leave certain distinctive artifacts in the resulting DeepFake Videos, which can be effectively captured by a dedicated deep neural network model
* We evaluate our method on several different sets of available DeepFake Videos which demonstrate its effectiveness in practice

# Scholarcy Summary

## Introduction

The increasing sophistication of mobile camera technology and the ever-growing reach of social media and media sharing portals have made the creation and propagation of digital videos more convenient than ever before.

The number of fake videos and their degrees of realism have been limited by the lack of sophisticated editing tools, the high demand on domain expertise, and the complex and time-consuming process involved.

The time of fabrication and manipulation of videos has decreased significantly in recent years, thanks to the accessibility to large-volume training data and high-throughput computing power, but more to the growth of machine learning and computer vision techniques that eliminate the need for manual editing steps.

A new vein of AI-based fake video generation methods known as DeepFake has attracted a lot of attention recently.

It takes as input a video of a specific individual (’target’), and outputs another video with the target’s faces replaced with those of another individual (’source’).

With proper post-processing, the resulting videos can achieve a high level of realism

## Objectives

As our aim is to expose the artifacts between fake face area and surrounding area, the RoIs are chosen as the rectangle areas that contains both the face and surrounding areas

## Methods

We detect synthesized videos by exploiting the face warping artifacts resulted from the DeepFake production pipeline.

The current DeepFake algorithms create synthesized face images of fixed sizes.

These faces are undergone an affine transform to match the poses of the target faces that they will replace (see Figure 1 (g) – (h)).

The facial region and surrounding regions in the original image/video frame will present artifacts, the resolution inconsistency due to such transforms after the subsequent compression step to generate the final image or video frames.

We propose to use a Convolutional Neural Network (CNN) model to detect the presence of such Face detection.

Shape refinement (c) Transform matrix + (d) DeepFake (e) (g) (f).

## Findings

ResNet networks have about 10% better performance compared to VGG16, due to the residual connections, which make the learning process more effective.

VGG16, ResNet, ResNet101 and ResNet152 can achieve AUC performance 84.5%, 98.7%, 99.1%, 97.8% respectively.

In this video based evaluation metric, ResNet network still performs ∼ 15% better than VGG16

## Conclusion

We describe a new deep learning based method that can effectively distinguish AI-generated fake videos (DeepFake Videos) from real videos.

Our method is based on the observations that current DeepFake algorithm can only generate images of limited resolutions, which are needed to be further transformed to match the faces to be replaced in the source video.

Such transforms leave certain distinctive artifacts in the resulting DeepFake Videos, which can be effectively captured by a dedicated deep neural network model.

We evaluate our method on several different sets of available DeepFake Videos which demonstrate its effectiveness in practice.

We would like to evaluate and improve the robustness of our detection method with regards to multiple video compression.

We currently using predesigned network structure for this task, but for more efficient detection, we would like to explore dedicated network structure for the detection of DeepFake videos