Literature Review

Nicholas Tee

1 Introduction

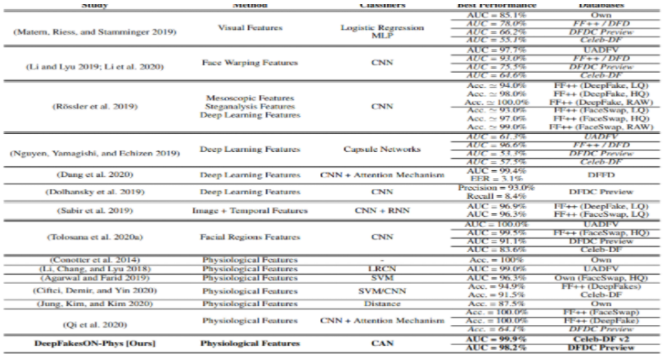
The main focus of this paper will be to discuss the current state of research on machine learning and artificial intelligence. It will focus on the applications researchers have used this technology for and how it may affect our quality of life. The paper will discuss three different documents that focus on advancements and applications in computer graphics and computer vision. The first paper will talk about the usage of computer vision technology to detect deepfakes. The second will talk about the effectiveness of our current facial recognition algorithms that are in use today and the flaws that they contain. The final paper will discuss a new method of enhancing resolution in videos.

2 Counteracitng Deepfakes

The first paper I want to talk about is "DeepFakesON-Phys: DeepFake Detection based on Heart Rate Estimation" by Javier Hernandez-Ortega, Ruben Tolosana, Julan Fierrez, and Aythami Morales, from the Biometrics and Data Pattern Analytics lab in Universidad Autonoma de Madrid. This paper tackles something that I think is overlooked. Most people are familiar with what a deep fake is; what people overlook is the growing technology steadily improving deep fakes. Although the technology comes with many benefits through its various applications, it can also bear malice. Especially in this time when the internet and media usage is at an all-time high, we are very susceptible to misinformation. The act of spreading fake news can be detrimental in many different ways. For instance, someone could easily use a deep fake to disguise as a government official and misinform the general public about very sensitive or essential topics. An example of this already exists, Jordan Peele used this technology to create a video of him disguised as former US president Barack Obama, the video was made three years ago.

Attempts to detect deep fakes have already been made. However, they are not that effective and are prone to detecting plenty of false-positive results. The paper challenges the use of physiological features to detect deep fakes. The paper does not go into too much technical detail on how the system works. What is mentioned is that the project uses existing results and tech made purely for the detection of heart rates. The DeepPhys deep learning model trained for heart rate estimation was implemented and modified to detect deep fakes. The final heart rate estimation is not used as a factor, but rather the information that is linked to the prediction of a person's heart rate is what is used. The system has two main steps when processing the videos. The system will first process the frames into a "Normalized Frame Difference" and a "Normalized Frame" for each frame. Once the frame has been processed, the "Normalized Fram Difference" will be put through the Motion model, designed to detect changes between consecutive frames. At the same time, the "Normalized Frame" is put through the Appearance Model, which analyzes the static information in the frame and determines which frames are the most likely to contain good information on whether or not a deep fake exists in the video. That information is then passed to the Motion model.

The paper also goes through the multiple tests that the group has done. They first took their system and compared its effectiveness and accuracy with previous studies. They looked at over ten studies and used the same databases that they did. This can be seen through the table that is presented in the table. In figure 1, you can see the different researchers, their methods, and their maximum accuracies with specific databases. You can see that all the methods that use physiological features seem to be far more effective than the rest. You can also see that between all the physiological studies, the method proposed in this paper was the most effective. It is difficult to critique their solution since I do not have an in-depth knowledge of how their system functions. However, the one critique that I can give is that I would like to see its effectiveness when using its average accuracy rather than its best performance. I would also like to see its effectiveness with random videos on the internet, rather than a set of databases that have already been used throughout the years.



Study	Method	Classifiers	Best Performance	Databases
(Matern, Riess, and Stamminger 2019)	Visual Features	Logistic Regression MLP	AUC = 85.1%	Own
			AUC = 78.0%	*FF++ / DFD*
			AUC = 66.2%	*DFDC Preview*
			AUC = 55.1%	*Celeb-DF*
(Li and Lyu 2019; Li et al. 2020)	Face Warping Features	CNN	AUC = 97.7%	UADFV
			AUC = 93.0%	*FF++ / DFD*
			AUC = 75.5%	*DFDC Preview*
			AUC = 64.6%	*Celeb-DF*
(Rössler et al. 2019)	Mesoscopic Features Steganalysis Features Deep Learning Features	CNN	Acc. ≃ 94.0% Acc. ≃ 98.0% Acc. ≃ 100.0%	FF++ (DeepFake, LQ) FF++ (DeepFake, HQ) FF++ (DeepFake, RAW)
			Acc. ≃ 93.0% Acc. ≃ 97.0% Acc. ≃ 99.0%	FF++ (FaceSwap, LQ) FF++ (FaceSwap, HQ) FF++ (FaceSwap, RAW)
(Nguyen, Yamagishi, and Echizen 2019)	Deep Learning Features	Capsule Networks	*AUC = 61.3%*	*UADFV*
			AUC = 96.6%	*FF++ / DFD*
			AUC = 53.3%	*DFDC Preview*
			AUC = 57.5%	*Celeb-DF*

(Dang et al. 2020)	Deep Learning Features	CNN + Attention Mechanism	AUC = 99.4%
			EER = 3.1%
				DFFD

(Dolhansky et al. 2019)	Deep Learning Features	CNN	Precision = 93.0%
			Recall = 8.4%
				DFDC Preview

(Sabir et al. 2019)	Image + Temporal Features	CNN + RNN	AUC = 96.9%
			AUC = 96.3%
				FF++ (DeepFake, LQ)
				FF++ (FaceSwap, LQ)

(Tolosana et al. 2020a)	Facial Regions Features	CNN	AUC = 100.0%	UADFV
			AUC = 99.5%	FF++ (FaceSwap, HQ)
			AUC = 91.1%	DFDC Preview
			AUC = 83.6%	Celeb-DF
(Conotter et al. 2014)	Physiological Features	-	Acc. = 100%	Own
(Li, Chang, and Lyu 2018)	Physiological Features	LRCN	AUC = 99.0%	UADFV
(Agarwal and Farid 2019)	Physiological Features	SVM	AUC = 96.3%	Own (FaceSwap, HQ)

(Ciftci, Demir, and Yin 2020)	Physiological Features	SVM/CNN	Acc. = 94.9%
			Acc. = 91.5%
				FF++ (DeepFakes)
				Celeb-DF

(Jung, Kim, and Kim 2020)	Physiological Features	Distance	Acc. = 87.5%	Own
(Qi et al. 2020)	Physiological Features	CNN + Attention Mechanism	Acc. = 100.0%	FF++ (FaceSwap)
			Acc. = 100.0%	FF++ (DeepFake)
			Acc. = 64.1%	*DFDC Preview*
DeepFakesON-Phys [Ours]	**Physiological Features**	**CAN**	**AUC = 99.9%**	**Celeb-DF v2**
			AUC = 98.2%	**DFDC Preview**

Fig. 1. Figure1: table from paper (Harnandex-Ortega et al)

3 Face Recognition

The following paper that I read is titled"On the Robustness of Face Recognition Algorithms Against Attacks and Bias" by Richa Singh, Akshay Agarwal, Maneet Singh, Shruti Nagpal, and Mayank Vatsa from IIIT-Delhi. The paper's primary focus is on face recognition technology and the recognition models that are used today. The paper challenges the robustness of said models. The paper talks about how these models, although reliable in some ways, are still susceptible to digital attacks and biases. These attacks refer to alterations made in the image that the model processes. Biases refer to how some models are more effective and accurate with specific population subsets than others. This is an important issue because, when face recognition becomes a more popular method of identification, we need to be sure that the systems in use are as robust as possible. If not, this could lead to identity theft or illegal access to sensitive information.

The paper goes over several different types of attacks and how they affect the accuracy of face recognition software. These different methods include presentation attacks, which refers to an attack directly at the system, disguise, makeup, and plastic surgery. Initially, I thought that what the paper was covering was somewhat redundant. I thought many of these "attacks" were particular cases that we would not normally encounter in the real world. However, I realized that the paper was trying to show how people may exploit the minor flaws within these recognition models to benefit themselves. Since this is still all relatively new technology, there is still much to learn and improve. Since face recognition is about finding a face despite all the slight variations present in the input image, I think that this paper can give others insight into the flaws of recognition software and use this research to improve our current tech.

4 Image Enhancements

The last paper that I want to summarize is titled "Low-Resolution Information Also Matters Learning Multi-Resolution Representations for Person Re-Identification" by Guoqing Zhang, Yuhao Chen, Weisi Lin, Arun Chandram, Xuan Jing. This project focuses mainly on a new method of sharpening videos. We know that despite all the advancements in camera technology, most security cameras still suffer from low-resolution images. Thus, much effort has been put into increasing and improving the image quality of the videos. The paper proposes that instead of focusing on improving the image quality soley, we should use the already available information that exists within the low-resolution version of the images.

A model that takes in both a low and high-resolution image as training data is then used to reconstruct the image. The network in question has different branches. Each branch focuses on a distinct feature of the image. The paper also shows the system's effectiveness relative to previous studies and their methods of increasing resolutions. It is shown that the technique that they proposed seems to be much more effective and accurate. I wanted to talk about this paper as I think that the problem that this solution is challenging is quite essential. If developed to a robust level, I also believe that this technology can be quite helpful in different applications. For instance, this technology can be used in conjunction with the other two papers mentioned in this report to create high-quality images for the training data. It would also mean that we would no longer be limited to high-resolution photos when developing projects such as these three papers.

References

[1] Hernandez-Ortega, Javier, et al. DeepFakesON-Phys: DeepFakes Detection Based on Heart Rate Estimation. Universidad Autonoma De Madrid.

[2] Singh, R., Agarwal, A., Singh, M., Nagpal, S., & Vatsa, M. (2020). On the Robustness of Face Recognition Algorithms Against Attacks and Bias. Proceedings of the AAAI Conference on Artificial Intelligence, 34(09), 13583-13589. https://doi.org/10.1609/aaai.v34i09.7085

[3] Zhang, Guoqing, et al. Low Resolution Information Also Matters ... - Ijcai.org. https://www.ijcai.org/proceedings/2021/0179.pdf.