Using Domain Knowledge RCNN for Deepfake Detection

Yahya Sherif Solayman Mohamed
School of Computing
Asia Pacific University of Technology and innovation (APU)
Kuala Lumpur, Malaysia
G.o.yahya.sherif@gmail.com

Nowshath K. Batcha
School of Computing
Asia Pacific University of Technology and innovation (APU)
Kuala Lumpur, Malaysia
nowshath.kb@staffemail.apu.edu.my

***Abstract*— As deep fakes and other similar synthetic image and video generation and manipulation techniques become more prominent the world is entering a new era where media is no longer a determinant or factual evidence of trustworthiness. While media sabotage is not a new concept, the use of artificial intelligence techniques such as deepfakes makes the generation of those fake images and videos more accessible and far cheaper in terms of time and effort. This paper explores the most modern of those synthetic image and video generation technologies: deepfakes. It is vital to raise awareness about the existence of the technology, the level of accessibilities, and its impact. To limit the impact of disinformation those technologies can spread, we propose a novel Recurrent Convolutional Neural Network model that focuses on the facial and other relevant domain-specific knowledge of the video to detect deepfake videos. This novel model abstracts away from irrelevant details in the inputted video which greatly improves the consistency of the detection. Furthermore, by identifying the features deepfakes struggle in producing, the model will learn about the features and attributes to prioritize during evaluation.**

Keywords*— *Artificial intelligence (AI), Deepfakes, Deep neural networks, Convolutional neural network, recurrent neural network, Pairwise learning, synthetic image generation, forgery detection, GAN.

I. INTRODUCTION

A. *The rise of deep fakes*

In April of 2018, BuzzFeed released a video titled "You Won't Believe What Obama Says In This Video!" (BuzzFeedVideo, 2018). The video started with the former president of the united stated Barack Obama giving a formal address warning the audience that we are in an era where technology can be used to make it seem like "anyone is saying anything." The video took a quick turn as Obama added "they could have me say for instance…" and went on to speak out of character even saying, "President Trump is a total and complete [explitive]." It was then revealed to the audience that this video in itself is a deep fake designed to raise awareness about the existence and level of accuracy of this technology.

This video is not an isolated case! There are millions of astonishing deepfake videos on the internet that can leave one terrified. Another instance is the viral "Zuckerberg speaks frankly" video where the CEO of Facebook Mark Zuckerberg seemingly confesses the vile purpose of Facebook (bill_posters_uk, 2019). Additionally, there is a less popular video where Trump is made to play a character in the movie breaking bad (Ctrl Shift Face, 2019).

B. *What are deepfakes?*

After watching those videos, one might wonder: what is a deepfake and how do they work? The term AI-synthesized media (Commonly referred to as "deepfakes") describes a machine learning algorithm or a neural network that achieves one of three tasks:

- Puppet mastery: where a figure's face or body is animate (Agarwal et al., 2019);
- Voice puppetry: an audio-driven facial video synthesis approach where the targeted character's lips utter the inputted text (Thies, 2019);
- Faceswap: takes the face region of the source video and inserts it on a targeted video making it seem like the former is doing the actions of the latter (Rössler et al., 2019).

It is worth noting that while the term deepfake became synonymous with facial image manipulation techniques, it is also used to refer to a specific manipulation method through online forums (Rössler et al., 2019). For this paper, the term deepfake will denote deep neural network algorithms used for the listed tasks (hence the "deep" portion) (Sample, 2020).

C. *Deepfake algorithm*

To understand how deepfakes work, it is helpful to explore the method used by one of the earliest attempts at deepfake creation: FakeApp. The algorithm runs thousands of pictures and frames by what is known as an encoder that extracts latent, or hidden, features of the inputted images. Using those features a decoder learns to reconstruct the image or video back using the encoder. Thus, to create a deepfake the algorithm uses the decoder with another person's encoder reconstructing the person's features to look similar to the one whose the encoder was based on as illustrated in Table I (Sample, 2020).

D. *Generative adversarial networks (GANs)*

Jumping into a more recent algorithm and the most powerful thus far: Generative adversarial networks (GANs). GANs generate very accurate and detailed deepfakes using a novel algorithm with two models first introduced by Ian Goodfellow (2014). GAN algorithms learn irregularities or patterns from the inputted image or video such that they can generate new images or videos that look like they were in the dataset (Brownlee, 2019). Fig 2., for example, uses a GAN model to generate artificial faces that did not exist before using two inputted images: denoted as the source and destination:

TABLE I. PAIRWISE LEARNING PERFORMANCE

Method/Target	WGAN-GP		DCGAN		WGAN		LSGAN		PGGAN	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
The proposed	0.986	0.751	0.929	0.916	0.988	0.927	0.947	0.986	0.988	0.948

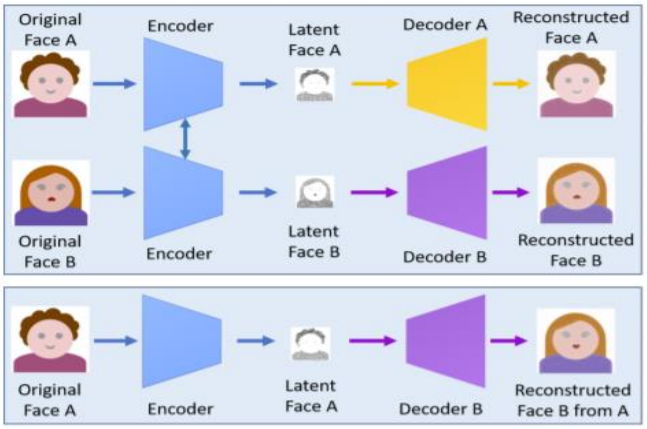

Fig. 1. double encoder-decoder deepfake creation (Nguyen, 2019)

Fig. 2. Generating Realistic Artificial Faces (Horev, 2019)

GANs work by training two deep neural networks: a generative model and a discriminative model. The former is trained to generate new examples using the inputted data, and the latter is trained to discriminate and determine if a given data is from the data set or not. The two models get trained simultaneously until the discriminative mistakenly believes that at least half of the data generated by the generative model is real (Brownlee, 2019).

II. LITERATURE REVIEW

It is a common trend in the field of artificial intelligence that when an AI is used to accomplish an objective, parallel research is conducted to identify weaknesses in the said AI algorithm, and deepfakes are no exception. For example, the United States Defense Advanced Research Projects Agency (DARPA) started a research scheme project called medifor (short for Media Forensics) to increase its efforts in detecting fraudulent visual images (Nguyen, 2019). Cooperations have also taken a role in detecting deepfakes. In 2019, Facebook teamed up with Microsoft to provide a new dataset to help the AI community with deepfake detection in what they called the deepfake challenge (ai.facebook.com, 2019). The literature review explores two of the more recent and perhaps very impactful deepfake detection papers.

A. *Deep Fake Image Detection Based on Pairwise Learning*

Hsu et al. (2020) proposed a pairwise learning algorithm to detect deep fake images. This paper used the CelebA data set created and introduced in "Deep Learning Face Attributes in the Wild. In Proceedings of the International Conference on Computer Vision, Santiago" (Liu et al., 2015) to generate real to fake data pairs using five state-of-the-art GAN models:

- DCGAN (Deep convolutional GAN) (Radford, 2015)
- WGAP (Wasserstein GAN) (Arjovsky, 2017)
- WGAN-GP (WGAN with Gradient Penalty) (Gulrajani, 2017)
- LSGAN (Least Squares GAN) (Mao, 2017)
- PGGAN (Karras, 2017)

This dataset will be useful when integrating the proposed system.

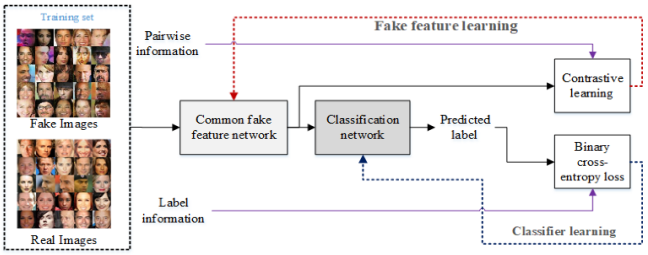

Fig. 3. The flowchart of the proposed fake face detector (Hsu et al., 2020)

The algorithm starts with feature extraction using a Common fake feature network (CFFN) which employees the Siamese network architecture (Chopra, 2005). To improve the representativeness of the fake images, the CFFN consists of three or five dense units depending on validation data being faced or general images. Each dense unit has a varying number of dense blocks (Huang et al., 2018). The CFFN extracts discriminative features between the data pairs, that is the pairwise information, and feeds it into a convolutional neural network, concatenated to the last layer of the CFFN for classification. The results were outstanding as shown in table 1, the algorithm got way over 90 percent precision on all except the WGAB-GP recall.

B. *FaceForensics++*

"FaceForensics++: Learning to Detect Manipulated Facial Images" (Rössler et al., 2019) is an outstanding piece of work that introduces a new dataset for image forgery detection, insights on image and video synthesis, human detection

results, a new benchmark for deep fake detection, and reported results of the model's performance against other detectors in 3 different video qualities after implementing a proposed pre-processing method for accuracy improvement. Each component in this study, shown in Fig 4., will contribute to the significance and will help with training the proposed model. Most of those components will be further explored in later sections of the paper.

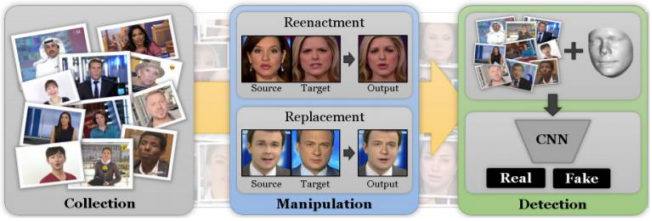

Fig. 4. FaceForensics++ is a dataset of facial forgeries that enables researchers to train deep-learning-based approaches (Rössler et al., 2019)

The paper first illustrates the video collection gathered from YouTube and then using both face replacement and reenactment methods a new data set of synthesized videos is created. After evaluating human candidates' ability to distinguish between real and fake videos, the paper proposes a new method for detecting fake videos.

Fig. 5. domain-specific forgery detection pipeline (Rössler et al., 2019)

Rather than training a model to classify the video, the proposed system first uses a model face tracker (Thies, 2017) to find and extra the face, as illustrated in Fig 5., and then feed it to the model for classification. As for the classification task, the system uses a couple of previously introduced detectors upgraded with the new pipeline: Steg. Features + SVM (Fridrich, 2012), Cozzolino et al.(2017), Bayar and Stamm (2016), Rahmouni et al.(2017), MesoNet (Afchaŕ et al., 2018), and XceptionNet (Chollet, 2017).

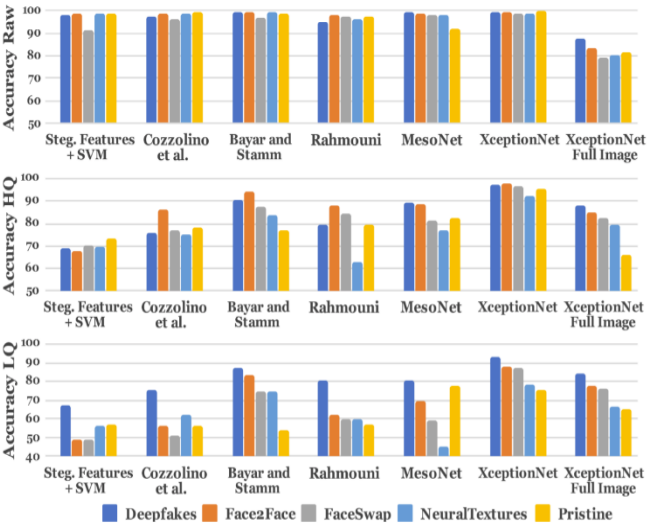

Fig. 6. Binary precision values of the baselines when trained on all four manipulation methods simultaneously. Aside from the Full Image XceptionNet, the proposed pre-extraction of the face region as input is used. (Rössler et al., 2019)

The results of each model's performance are detailedly illustrated in Fig 6., where the XceptionNet outperformed. This is unlike the Full Image XceptionNet which is the same model but without implementing the proposed pipeline. Lastly, the proposed algorithm also reports which video synthesis it detected with great accuracy.

Fig. 7. Classified video synthesis method (Niessner, 2019)

III. PROBLEM STATEMENT

While the results of the Deep Fake Image Detection Based on Pairwise Learning were astounding the proposed model is only limited to fake image detection, which while valuable does not contribute to detecting fake videos like the ones illustrated in the introduction. This is problematic, as fake videos are more dangerous and impactful. Think about it: while a fake image can raise distress and conflict, a fake video, like the one released by BuzzFeed (BuzzFeedVideo, 2018), is more believable and realistic which deepens the uncertainty. Furthermore, the spread of fake videos without a proper detector will only serve to increase public distrust in media.

A study into the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News was conducted by Cristian Vaccari, Andrew Chadwick (2020) in which they showed participants the fake Obama video (BuzzFeedVideo, 2018), but removed the clip where it is revealed that the video is a deep fake and assessed how participants were willing to trust other sources. The researchers recorded their results, as shown below, and found that the knowledge that the video was fake does not impact the subject's susceptibility to false statements. It did however increase the percipient's level of uncertainty illustrating the impact of fake videos on the community. Thus, the pairwise learning paper needs to expand to video detection to make a noticeable impact.

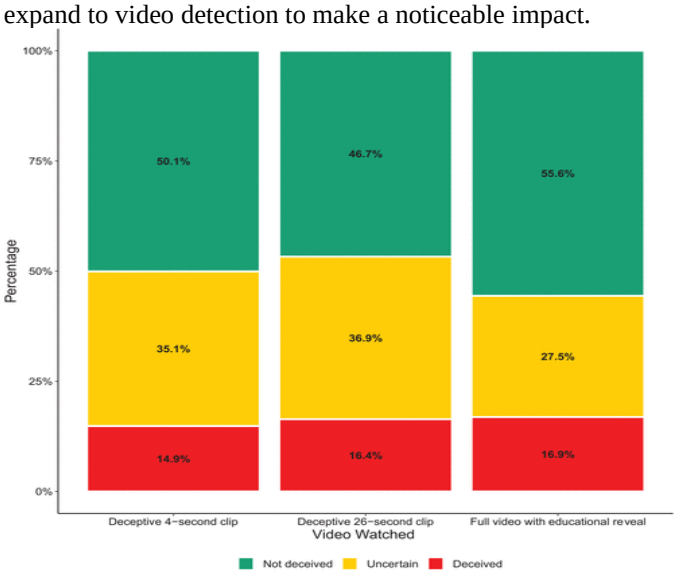

Fig. 8. Assessment of the truthfulness of the video, by treatment (Vaccari and Chadwick, 2020)

Looking at the work of the FaceForensics++ paper, it did not introduce a new method to detecting fake videos as much as it illustrated an improvement in addressing the problem. This is consistent with the paper's abstract to "show that the use of additional domain-specific knowledge improves forgery detection to unprecedented accuracy" (Rössler et al., 2019). Thus, the aim of the paper is not to solve the problem, but rather to lay the foundation for a novel deep fake detection method to be built on. This theme is consistent with other contributions to the field of Deep fake detection and computer science in general like the ones discussed at the beginning of the literature review section.

IV. Aim and Objectives of the Research

With a better understanding of deep fakes, their algorithm, and the huge impact they stand to deal with, this paper aims to propose a novel algorithm for efficient deep fake video detection. For this aim to be met the following objectives must be fulfilled:

- Understanding the weakness of deep fakes.
- Developing a reliable feature recognition model.
- Configure a neural network for video processing and classification.
- Leveraging domain-specific knowledge.

V. Research Questions

For a reliable integration of a novel solution to the stated problem, some questions must be addressed. While this paper has explored the concept of deep fake and their underlying algorithms, found in the introduction section, and the latest and most accurate algorithms used to detect deep fakes, found in the literature review section, there are a couple more questions to address:

- How much data will a model need for the training to return reliable results?
- Can new deep fake algorithms exploit the detector easily?
- What is the future research direction to expand on this piece of work?
- How to account for mathematical limitations.

VI. Significance of the Research

The significance of a deep fake detector is immeasurable as it can prevent so many lies and misinformation from spreading. Thus far only examples of harmless experiments with deep fakes have been illustrated, but this technology is already used to spread fear as well as distrust. In April 2018, a video went viral on social media in India. The video was a CCTV camera footage of some kids playing cricket in the street when suddenly two men storm in and kidnap one of the kids on a motorbike. This video caused a lot of mob violence and the death of around 9 people (BBC News, 2018). This video was a deep fake used for malicious intent. Those actions were a consequence of the little awareness in communities about this technology and the lack of an efficient method to detect it.

Furthermore, even if people were aware of the existence of deep fakes, that nevertheless does not solve the problem. The survey conducted in FaceForensics++ showed that deep fakes are very good at deceiving the naked eye. The results of the survey conducted by FaceForensics++, shown below, showed that participants detected less than 80% of the image or video synthesizing techniques. Furthermore, in some cases, the results are worse than a coin flip.

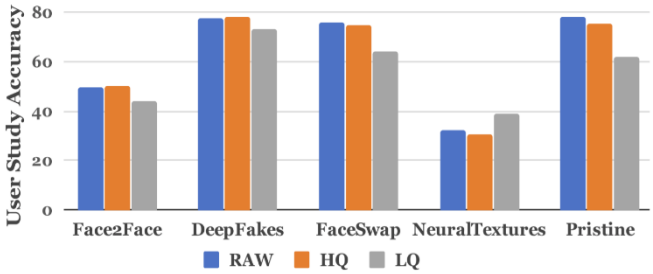

Fig. 9. Forgery detection results of our user study with 204 participants (Rössler et al., 2019)

Additionally, Deep fakes are becoming more prominent, affordable, and easy to integrate. A paper by the title of "First Order Motion Model for Image Animation" introduced a new algorithm for creating deep fakes by animating an inputted image using a driving video without the need for any labels, annotation, or knowledge about the image (Siarohin et al., 2019). Those technologies make developing deep fake videos trivial even for a college student, such as myself. Even if one lacks mastery in the field tutorials exist to help new users navigate through the code (Rubik's Code, 2020).

That is not the only technique made accessible. Another recent paper under the title of "Neural Voice Puppetry: Audio-driven Facial Reenactment" Developed a technique in which the inputted image is transformed into a video using an audio input as illustrated below (Thies, 2019).

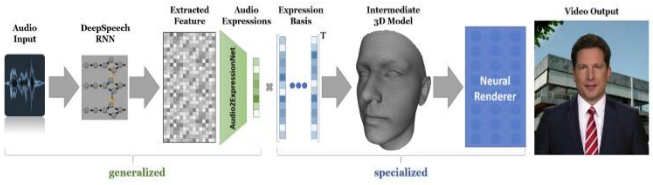

Fig. 10. Pipeline of Neural Voice Puppetry (Thies, 2019)

Those technologies and algorithms require less and less knowledge in the respective field as companies start to offer them. For example, in China, a new app by the name of Zao employs deep fake technology allowing users to swap their faces with characters in videos (Shanghai, 2019). Another instance is Deepfake web a website that at the right price generates deepfakes for clients (Deepfakes web β, n.d.).

Lastly, this research is significant because of fast those deep fake videos spread (Rashmi, 2019). Part of the reason the mob violence occurred in the last incident mentioned is how quickly the video spread. This should not be a surprise since something with such outrageous content will quickly spread among people.

VII. Methodology

A. *Data collection*

The most important aspect of any neural network is the data used to train it. One can build the best neural network for

a given task, but without sufficient data, the model will be inconsistent. As previously stated, the proposed system will be trained using the novel new data sets provided to the AI community such as the new dataset provided by Facebook and Microsoft (ai.facebook.com, 2019), the pairwise deepfake detection paper (Hsu et al, 2020), and FaceForensics++ (Rössler et al., 2019).

Additionally, new deepfakes data can be generated by running quantitive research asking participants from across the globe to make certain features or expressions and create new deepfakes using the data collected. Such data collection will contribute to the development of the deep fake detector and the AI community. The results of a survey conducted to assess people's willingness to contribute to the proposed database show that while uncertain people are willing to contribute, those in the field of computer science are more likely to contribute.

B. Limitations and legal concerns

Another important aspect to consider is the limitations of the results. As proven in a study by Sakshi Agarwal and Lav R. Varshney (2019), deepfake detectors have limitations in their numerical algorithm that is guaranteed to generate errors. It is worth noting that those limitations can nonetheless be exceeded by accounting for specific features and limiting bias supporting the claim in the FaceForensics++ study that specific domain knowledge can boost the accuracy of a deepfake detection model.

As for legal concerns, the development process has no legal limitations as it is conducted based on the content of any participants involved, and in general, there are no laws to regulate or control deepfakes let alone the integration of a method to detect them, unless either is used for a malicious purpose (Sample, 2020).

VIII. Overview of the Proposed System

It has been proven thus far the strong impact that domain knowledge has on the accuracy of deep fake detection. Thus we propose a detailed image processing algorithm that extracts all of the relevant features and attributes to the model rather than letting the model deal with a set of frames without any context. One powerful option for this detection task is the one proposed in Precise Detailed Detection of Faces and Facial Features (Ding and Martinez, 2008).

Fig. 11. Shown here is an example of accurate and detailed face detection in each of the frames over a video sequence of an ASL sentence. The top row shows the automatic detection obtained with the algorithm. The bottom row provides a manual detection for comparison. (Ding and Martinez, 2008)

This approach divides segments of the detected facial features into sub-classes of constructors of the feature. The mouth being open is one such example. The reason this method is so powerful is that it can detect the eyes to a very precise degree. Not only can the center of the eye window be detected, but also the eye shape as illustrated below:

Fig. 12. Eye corners are represented with an asterisk. The iris is shown as a circle, of which, the red part is the visible component. The upper and lower contours of the shape of the eye are shown in blue. (Ding and Martinez, 2008)

Once the eye is detected the rest of the facial features can easily be inferred as their locations are respective of the eye.

Fig. 13. Shown here are five examples of the automatic detection of faces and facial features as given by the proposed approach (Ding and Martinez, 2008)

After applying this detailed pre-processing technique to the training video, we construct a recurrent convolutional neural network for deepfake detection. The extracted features will be used as the pairwise information that the model will use, similar to the work in Deep Fake Image Detection Based on Pairwise Learning reviewed in the literature review section. The model's structure and the layer will match that of the mentioned work but will employ a recurrent neural network. On the first iteration, the model takes in the first frame along with the extracted features and constructs a probability distribution mapping the likelihood that the video is a deepfake and which algorithm was used to construct it. On the second iteration and onwards the model feeds the probability distribution back into itself along with the next frame and updates the probability distribution using the probability distribution mapped thus far along with the current frame and the extracted features. The integration of a Recurrent neural network allows our model to parse video data.

IX. Conclusion

The rise of deepfake technology is increasing in speed as the technology is becoming more predominant and accessible. Without an efficient way to reliable detect fake videos more tragedies stand to rise as the mob violence in India, illustrated in further detail in the significance section, as the public's trust in media and news decreases.

This paper proposed a novel recurrent neural network that leverages domain-specific knowledge extracted during pre-processing to better detect deepfake videos. Indeed, as has been illustrated throughout this paper the use of domain-specific knowledge is quite powerful in detecting fake media.

X. Future Research Directions

To address the issue illustrated in the Limits of deepfake detection of the "A robust estimation viewpoint" paper (Agarwal, 2019), discussed in the significance section, the

model proposed in the Deep Learning Face Attributes in the Wild paper (Liu et al., 2015) can be integrated to detect attributes of the detected facial region.

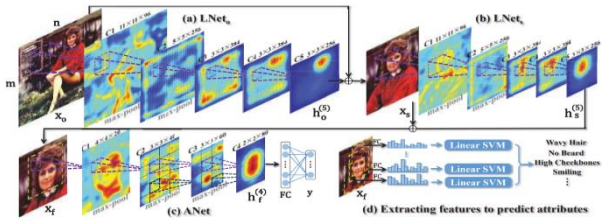

Fig. 14. Deep Learning Face Attributes in the Wild proposed pipeline of attribute prediction (Liu et al., 2015)

The paper proposes two cascaded convolutional neural networks: ANet and LNet. Both CNNs are fine-tuned jointly but are pre-trained differently.

- LNet is pre-trained by massive general object categories for face localization.
- ANet is pre-trained by massive face identities for attribute prediction.

This novel approach is unique and similar in concept to GANs as it divides the tasks into two models rather than expecting one to do all the work. Thus, this approach outperformed the state-of-art models with a sizable margin as illustrated in the Fig 15.

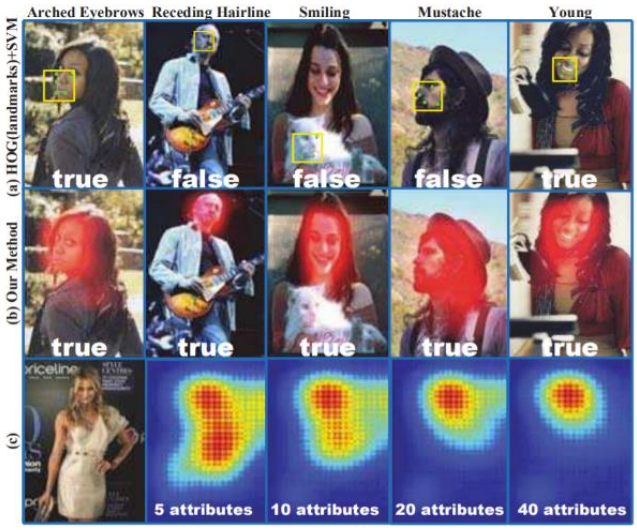

Fig. 15. (a) Inaccurate localization and alignment lead to prediction errors on attributes by existing methods (b) LNet localizes face regions by averaging the response maps of attribute filters. ANet predicts attributes without alignment (c) Face localization with the averaged response map when LNet is trained with different numbers of attributes (Liu et al., 2015)

The proposed system could even be trained to recognize uncertainty raised by this model to detect fraudulent features that a deep fake failed to generate.

REFERENCES

Afchar, D., Nozick, V., Yamagishi, J. and Echizen, I., 2018, December. Mesonet: a compact facial video forgery detection network. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS) (pp. 1-7). IEEE.

Agarwal, S., and Varshney, L. R. (2019). Limits of deepfake detection: A robust estimation viewpoint. arXiv preprint arXiv:1905.03493.

Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K. and Li, H., 2019, June. Protecting World Leaders Against Deep Fakes. In CVPR Workshops (pp. 38-45).

ai.facebook.com. (2019). Creating a dataset and a challenge for deepfakes. [online] Available at: https://ai.facebook.com/blog/deepfake-detection-challenge [Accessed 7 Apr. 2021].

Arjovsky, M., Chintala, S. and Bottou, L., 2017, July. Wasserstein generative adversarial networks. In International conference on machine learning (pp. 214-223). PMLR.

Bayar, B. and Stamm, M.C., 2016, June. A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security (pp. 5-10).

BBC news: India WhatsApp "child kidnap" rumours claim two more victims. (2018). BBC News. [online] 11 Jun. Available at: https://www.bbc.com/news/world-asia-india-44435127 [Accessed 19 Apr. 2021].

bill_posters_uk. (2019). Login • Instagram. [online] Available at: https://www.instagram.com/p/BypkGIvFfGZ/?utm_source=ig_web_copy_link [Accessed 6 Apr. 2021].

Brownlee J. (2019). A Gentle Introduction to Generative Adversarial Networks (GANs). [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/.

BuzzFeedVideo (2018). You Won't Believe What Obama Says In This Video! YouTube. Available at: https://www.youtube.com/watch?v=cQ54GDm1eL0.

Chollet, F., 2017. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1251-1258).

Chopra, S. (2005). Learning a similarity metric discriminatively, with application to face verification. In IEEE Conference on Compter Vision and Pattern Recognition (pp. 539-546).

Cozzolino, D., Poggi, G. and Verdoliva, L., 2017, June. Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection. In Proceedings of the 5th ACM Workshop on Information Hiding and Multimedia Security (pp. 159-164).

Ctrl Shift Face. (2019). Better Call Trump: Money Laundering 101 [DeepFake]. [online] Available at: https://www.youtube.com/watch?v=Ho9h0ouemWQ&t=51s [Accessed 6 Apr. 2021].

Deepfakes web β. (n.d.). Deepfakes web β | The best online faceswap app. [online] Available at: https://deepfakesweb.com/.

Ding L, Martinez AM. Precise detailed detection of faces and facial features. In2008 IEEE Conference on Computer Vision and Pattern Recognition 2008 Jun 23 (pp. 1-7). IEEE.

Fridrich, J. and Kodovsky, J., 2012. Rich models for steganalysis of digital images. IEEE Transactions on Information Forensics and Security, 7(3), pp.868-882.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y. (2014). Generative Adversarial Nets. Available at: https://arxiv.org/pdf/1406.2661.pdf.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V. and Courville, A., 2017. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028.

Horev R. (2019). Style-based GANs – Generating and Tuning Realistic Artificial Faces | Lyrn.AI. [online] Available at: https://www.lyrn.ai/2018/12/26/a-style-based-generator-architecture-for-generative-adversarial-networks/.

Hsu, C. C., Zhuang, Y. X., and Lee, C. Y. (2020). Deep fake image detection based on pairwise learning. Applied Sciences, 10(1), 370.

Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K. (2018). Densely Connected Convolutional Networks. [online] . Available at: https://arxiv.org/pdf/1608.06993.pdf.

Karras, T., Aila, T., Laine, S. and Lehtinen, J., 2017. Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196.

Liu, Z., Luo, P., Wang, X. and Tang, X., 2015. Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision (pp. 3730-3738).

Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z. and Paul Smolley, S., 2017. Least squares generative adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2794-2802).

Nguyen, T.T., Nguyen, C.M., Nguyen, D.T., Nguyen, D.T. and Nahavandi, S., 2019. Deep learning for deepfakes creation and detection. arXiv preprint arXiv:1909.11573, 1.

Niessner M. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images (ICCV 2019). [online] Available at: https://www.youtube.com/watch?v=x2g48Q2I2ZQ&t=150s [Accessed 11 Apr. 2021].

Radford, A., Metz, L. and Chintala, S., 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

Rahmouni, N., Nozick, V., Yamagishi, J. and Echizen, I., 2017, December. Distinguishing computer graphics from natural images using convolution neural networks. In 2017 IEEE Workshop on Information Forensics and Security (WIFS) (pp. 1-6). IEEE.

Rashmi, H. (2019). "6 Deep Fakes." National Academies of Sciences, Engineering, and Medicine. 2019. Implications of Artificial Intelligence for Cybersecurity: Proceedings of a Workshop. Washington, DC: The National Academies Press. DOI: 10.17226/25488.

Rössler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J. and Nießner, M. (2019). FaceForensics++: Learning to Detect Manipulated Facial Images. Available at: https://arxiv.org/pdf/1901.08971.pdf.

Rubik's Code. (2020). Create Deepfakes in 5 Minutes with First Order Model Method. [online] Available at: https://rubikscode.net/2020/05/25/create-deepfakes-in-5-minutes-with-first-order-model-method/ [Accessed 4 Apr. 2021].

Sample, I. (2020). What are deepfakes – and how can you spot them? [online] the Guardian. Available at: https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them.

Shanghai (2019). Chinese deepfake app Zao sparks privacy row after going viral. [online] the Guardian. Available at: https://www.theguardian.com/technology/2019/sep/02/chinese-face-swap-app-zao-triggers-privacy-fears-viral.

Siarohin, A., Lathuilière, S., Tulyakov, S., Ricci, E. and Sebe, N., 2019. First-order motion model for image animation. Advances in Neural Information Processing Systems, 32, pp.7137-7147.

Thies, J. (2019). Face2Face: Real-time facial reenactment. it - Information Technology, 61(2-3), pp.143–146.

Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C. and Nießner, M. 2017. Face2Face: Real-time Face Capture and Reenactment of RGB Videos.

Vaccari, C. and Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. Social Media + Society, 6(1), p.205630512090340