Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 7 Issue 10, October - 2020

A Model Recommendation For The Recognition Of Deepfake Videos Using LSTM Sequencing

Ayse Kok Arslan

***Abstract*—This research describes a new deep learning based method that can effectively distinguish AI-generated fake videos (DeepFake Videos) from real videos. The method is based on the observations that current DeepFake algorithm can only generate images of limited resolutions, which are then 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. The method is evaluated on several different sets of available DeepFake Videos which demonstrate its effectiveness in practice.**

Introduction

This research describes a new deep learning based method that can effectively distinguish AI (artificial intelligence) -generated fake videos (referred to as DeepFake videos hereafter) from real videos. The method is based on the observations that current DeepFake algorithm can only generate images of limited resolutions, which need to be further warped to match the original faces in the source video. Such transforms leave distinctive artifacts in the resulting DeepFake videos, and the research shows that they can be effectively captured by convolutional neural networks (CNNs). Compared to previous methods which use a large amount of real and DeepFake generated images to train CNN classifier, this method does not need DeepFake generated images as negative training examples since it targets the artifacts in affine face warping as the distinctive feature to distinguish real and fake images. The advantages of this method are two-fold: (1) Such artifacts can be simulated directly using simple image processing operations on an image to make it as negative example. Since training a DeepFake model to generate negative examples is time-consuming and resource-demanding, this method saves a plenty of time and resources in training data collection; (2) Since such artifacts are general existed in DeepFake videos from different sources, the method is more robust compared to others. The method is evaluated on two sets of DeepFake video datasets for its effectiveness in practice.

Overview of Current Media

News Web sites have recently become a major way of acquiring news. According to a report published by Pew Research Center for People and the Press (2008), 40% of the Americans get their national and international news from the Internet. The increasing use of online networks and their global diffusion raise questions regarding their biases that could affect our perceptions of the world. Together with better abilities to express local and national views, popular news Web sites may reinforce, for example, dominant American or Western views.

On the other hand, there are also consistent empirical indications to support Wallerstein's (1974) World System Theory (WST) that argues for the presence of an asymmetrical international system of core-, semi-peripheral-, and peripheral-countries. Various scholarly work (Galtung & Ruge, 1965; Mowlana, 1985; Schramm, 1964), and these were substantially examined and found to play a significant role in international news flows (Chang, Lau, & Hao, 2000; Chang & Lee, 1992; Chang, Shoemaker, & Brendlinger, 1987; Chang et al., 2005; Golan & Wanta, 2003; Peng, 2004; Riffe, 1996; Wu, 2000) show that media in general is still predominantly centered either on US or those with political or economic ties to the U.S.

The Flow of Fake News

The bias of international news and the dominance of certain actors are often related to the political economy of news production, i.e. the strong influence of certain international news agencies and the one-directional news flow, reflecting mostly the interests of large news producing countries (Galtung & Ruge, 1965; Schramm, 1964). The model of international communication presented by Mowlana (1985) differentiates between the news sources, messages, distribution, and destination on the one hand, and the communication hardware and software on the other. Thus, in order to grasp the complexity of the flow of fake news, he suggests looking at the international network of actors that involve in the process of news production, dissemination, and consumption, as well as the technological means employed.

The content of international news is heavily biased, where certain countries are totally neglected from the imaginary world constructed by the news. Chang, Himelboim, and Dong (2009) found support for the core-periphery model when studying the structure of hyperlinks in news Web sites.

Other studies that examined the core-periphery structure of nations in communication terms (Barnett, 2001; Barnett, Jacobson, Choi, and Sun-Millers, 1996; Chase-Dunn & Hall, 1994; see below) provide empirical support for the significance of the economic dimension. The core-periphery structure that is common to the WST and to Mowlana's model of international communication presents some drawbacks as well.

Although European and American countries are still the main exporters of media content, Tunstall (2008) argues that in many countries content becomes predominantly local. He differentiates between big and small population countries, indicating that the latter produce less local content than the former, and import relatively more from the U.S., the U.K., and France, and from their larger neighbors. Subsequently, he divides the centers of media production and dissemination into several self-sufficient regions, based on geography, religion, culture, and language (or group of languages). Given this flow of international news, it should not be a surprise that fake news could be easily spread on a global scale.

Definitions

Given the recent spread of fake news and the controversies surrounding the impact of social networking sites such as Facebook, this study is concerned with expressive online participation; i.e. a form of online participation that entails the public expression of political orientations. From this perspective, not all online conversations entail participatory behavior.

It should be clear that all forms of political conversation have important political consequences, but distinguishing between background conversations and the public expression of individual views makes sense theoretically, and has been supported empirically. In my view, then, expressive online participation is a subset of political participation–political participation with a dimension of public expressiveness. To be clear: (1) there is no consensus in the literature on the conceptualization of expressive online participation; (2) the distinction between online conversation and expressive online participation is sometimes blurred in certain strands of the literature (Boyle et al., 2006); and (3) certain theorists discount expressive online participation as form of participation altogether (Verba, Schlozman, & Brady, 1995).

In previous literature, expressive online participation has sometimes been conceptualized as political expression, political attitude expression, political participation involving public expression, and opinion expression. Certain strands of the literature blur the distinction between political conversation and expressive online participation. For example, Boyle et al. (2006) coin the term expressive action to include talking to friends and family about politics, sending letters to the editor, contacting public officials, and attending rallies. Moreover, Verba et al. (1995) explicitly do not consider online discussion among friends, letters to the editor, or calls to a live show as forms of political participation since “the target audience is not a public official” (p. 40), and rather consider these to be at the border of political activity. It is not the purpose of this article to fully sort out these controversies that are intimately tied to the definition of participation under which one operates, but it is critical to make them explicit in order to place the findings within the broader context of spread of fake news within the political sphere.

This work adopts this expanded view of political participation, and contend that expressive online participation constitutes a sub-dimension of political participation, one that is particularly critical for societies in transition, in which democratic institutions are not fully established. This study conceptualizes expressive political participation as a dimension of this broader construct–one that is particularly relevant for the study of societies in transition and the emergence of democratic institutions–and treat conversation and news media use as antecedents of expressive political participation.

Fake News within the Political Sphere

The results of the first wave of research on the effects of fake news on political participation provided mixed results, in part because some studies employed access or time spent rather than specific uses; also certain samples were considered not to be representative of the population; and causality and endogeneity problems were still being sorted out (for a summary of this debate see Nie, 2001). Since then, a new wave of studies has mostly refuted dystopian views of new communication technologies, establishing positive relationships between informational uses of the Internet and social capital (Shah, Kwak, & Holbert, 2001), political participation (Shah, Schmierbach, Hawkins, Espino, & Donovan, 2002), and civic engagement (Jennings & Zeitner, 2003).

Wellman, Quan-Haase, Witte, and Hampton (2001) provided evidence that online interaction supplements interpersonal relations and results in increased voluntary association membership and increased political participation (see also Wellman et al., 2003). Even Kraut has revisited his earlier study and claimed that the negative effects of Internet use had “dissipated.”

Furthermore, online information seeking has been linked to increases in online interactive civic messaging that ultimately result in higher levels of civic participation (Shah et al., 2005). Kavanaugh, Reese, Carroll, and Rosson (2005) provide evidence that this happens through an interaction effect with weak ties (Granovetter, 1973), which results both in increased face-to-face contact and political participation.

Understanding Deep Fake Videos

Researchers in artificial intelligence should always reflect on the dual use nature of their work, allowing misuse considerations to influence research priorities and norms. Given the severity of the malicious attack vectors that deepfakes have caused, this paper presents a novel solution for the detection of this kind of video (Figure 1.0):

Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 7 Issue 10, October - 2020

- First, the paper proposes a two-stage analysis composed of a CNN to extract features at the frame level followed by a temporally-aware RNN network to capture temporal inconsistencies between frames introduced by the face-swapping process.
- Second, it makes use of a collection of 600 videos to evaluate the proposed method, with half of the videos being deepfakes collected from multiple video hosting websites.
- Third, it shows experimentally the effectiveness of the described approach, which allows use to detect if a suspect video is a deepfake manipulation with 94% more accuracy than a random detector baseline in a balanced setting.

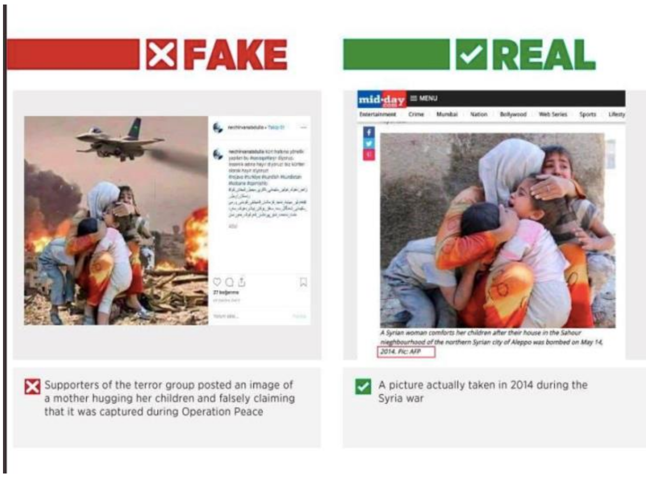

Figure 1. Example of a fake video

The field of digital media forensics aims to develop technologies for the automated assessment of the integrity of an image or video. Both feature-based and CNN-based integrity analysis methods have been explored in the literature. For video-based digital forensics, the majority of the proposed solutions try to detect computationally cheap manipulations, such as dropped or duplicated frames or copy-move manipulations. Techniques that detect face-based manipulations include methods that distinguish computer generated faces from natural ones such as Conotter et al. (2015) or Rahmouni et al. (2010). In biometry, Raghavendra et al. (2016) recently proposed to detect morphed faces with two pre0trained deep CNNs and Zhou et al. (2018) proposed detection of two different face swapping manipulations using a two-stream network. Of special interest to practitioners is a new dataset by Rossler et al. (2019), which has about half a million edited images that have been generated with feature-based face editing (2016).

Face-based Video Manipulation Methods

Multiple approaches that target face manipulations in video sequences have been proposed since the 1990s. Thies et al. demonstrated the first real-time expression transfer for faces and later proposed Face2Face (2005), a real-time facial reenactment system, capable of altering facial movements in different types of video streams. Alternatives to Face2Face have also been proposed. Several face image synthesis techniques using deep learning have also been explored as surveyed by Lu et al. (2013). Generative adversarial networks (GANs) are used for aging alterations to faces (2018), or to alter face attributes such as skin color (2010). Deep feature interpolation shows remarkable results in altering face attributes such as age, facial hair or mouth expressions. Similar results of attribute interpolations are achieved by Lample et al. (2015).

Moreover, Long Short Term Memory (LSTM) networks are a particular type of Recurrent Neural Network (RNN), first introduced by Hochreiter and Schmidhuber [20] to learn long-term dependencies in data sequences. When a deep learning architecture is equipped with a LSTM combined with a CNN, it is typically considered as “deep in space” and “deep in time” respectively, which can be seen as two distinct system modalities. CNNs have achieved massive success in visual recognition tasks, while LSTMs are widely used for long sequence processing problems. Because of the inherent properties (rich visual description, long-term temporal memory and end-to-end training) of a convolutional LSTM architecture, it has been thoroughly studied for other computer vision tasks involving sequences (e.g. activity recognition or human re-identification in videos) and has led to significant improvements.

The new generation of AI-based video synthesis algorithms are based on the recent developments in new deep learning models, especially the generative adversarial networks (GANs). A GAN model consists of two deep neural networks trained in tandem. The generator network aims to produce images that cannot be distinguished from the training real images, while the discriminator network aims to tell them apart. When training completes, the generator is used to synthesize images with realistic appearance.

The creation of a DeepFake video starts with an input video of a specific individual (’target’), and generates another video with the target’s faces replaced with that of another individual (’source’), based on a GAN model trained to translate between the faces of the target and the source.

More recently, Zhu et al. (2010) proposed cycle-consistent loss to push the performance of GAN, namely Cycle-GAN. Bansal et al. [2017] stepped further and proposed Recycle-GAN, which incorporated temporal information and spatial cues with conditional generative adversarial networks. StarGAN learned the mapping across multiple domains only using a single generator and discriminator.

The artifacts introduced by the DeepFake production pipeline is in essence due to affine transforms to the synthesized face. In the literature of digital media forensics, detecting transforms or the underlying resampling algorithm has been extensively studied; yet, the performance of these methods are affected by the post-processing steps which are not subject to simple modeling. Besides, these methods usually aim to estimate the exact resampling operation from whole images, but for this research's purpose, a simpler solution can be obtained by just comparing regions of potentially synthesized faces and the rest of the image – the latter are expected to be free of such artifacts while the existence of such artifacts in the former is a telltale cue for the video being a DeepFake.

Zhou et al. (2010) proposed two-stream CNN for face tampering detection. NoisePrint (2009) employed CNN model to trace device fingerprints for forgery detection. Recently, detecting GAN generated images or videos has also made progress. Li et al. (2010) observed that DeepFake faces lack realistic eye blinking, as training images obtained over the Internet usually do not include photographs with the subject's eyes closed. The lack of eye blinking is detected with a CNN/RNN model to expose DeepFake videos. However, this detection can be circumvented by purposely incorporating images with closed eyes in training.

Yang et al. (2017) utilized the inconsistency in head pose to detect fake videos. The wor exploited the color disparity between GAN generated images and real images in non-RGB color spaces to classify them. The work also analyzed the color difference between GAN images and real images. However, it is not clear if this method is extensible to inspecting local regions as in the case of DeepFake.

Afchar et al. (2018) trained a convolutional neural networks namely MesoNet to directly classify real faces and fake faces generated by DeepFake and Face2face. The work extended to temporal domain by incorporating RNN on CNN. While it shows promising performance, this holistic approach has its drawback. In particular, it requires both real and fake images as training data, and generating the fake images using the AI-based synthesis algorithms is less efficient than the simple mechanism for training data generation in our method.

Methodology

This paper describes a new deep learning based method that can effectively distinguish DeepFake videos from the real ones. The method is based on a property of the DeepFake videos: due to limitation of computation resources and production time, the DeepFake algorithm can only synthesize face images of a fixed size, and they must undergo an affine warping to match the configuration of the source's face. As such, this artifacts can be used to detect DeepFake Videos.

Before going into technical details let's first see how a deepfake video is generated to understand why these anomalies are introduced in the videos and how we can exploit them.

Creating Deepfake Videos

In terms of the deep learning techniques, the auto-encoder has been applied for dimensionality reduction, compact representations of images, and generative models learning. Thus, auto-encoders are able to extract more compressed representations of images with a minimized loss function and are expected to achieve better compression performance than existing image compression standards. The compressed representations or latent vectors that current convolutional auto-encoders learn are the first cornerstone behind the face-swapping capabilities. Next comes the use of two sets of encoder-decoders with shared weights for the encoder networks.

In terms of training, two sets of training images are required:

- The first set only has samples of the original face that will be replaced, which can be extracted from the target video that will be manipulated. This first set of images can be further extended with images from other sources for more realistic results.

- The second set of images contains the desired face that will be swapped in the target video. To ease the training process of the auto-encoders, the easiest face swap would have both the original face and target face under similar viewing and illumination conditions. However, this is usually not the case. Multiple camera views, differences in lightning conditions or simply the use of different video codecs makes it difficult for auto-encoders to produce realistic faces under all conditions. This usually leads to swapped faces that are visually inconsistent with the rest of the scene.

This frame-level scene inconsistency will be the first feature that will be exploited with this approach. It is also important to note that if one trains two auto-encoders separately, they will be incompatible with each other. If two auto-encoders are trained separately on different sets of faces, their latent spaces and representations will be different. This means that each decoder is only able to decode a single kind of latent representations which it has learnt during the training phase. This can be overcome by forcing the two set of auto-encoders to share the weights for the encoder networks, yet using two different decoders. In this fashion, during the training phase these two networks are treated separately and each decoder is only trained with faces from one of the subjects. However, all latent faces are produced by the same encoder which forces the encoder itself to identify common features in both faces. This can be easily accomplished due to the natural set of shared traits of all human faces (e.g. number and position of eyes, nose, etc).

When the training process is complete, one can pass a latent representation of a face generated from the original subject present in the video to the decoder network trained on faces of the subject we want to insert in the video. The decoder will try to reconstruct a face from the new subject, from the information relative to the original subject face present in the video. This process is repeated for every frame in the video where we want to do a face-swapping operation. It is important to point out that for doing this frame-level operation, first a face detector is used to extract only the face region that will be passed to the trained auto-encoder. This is usually a second source of scene inconsistency between the swapped face and the reset of the scene.

The third major weakness to be exploited is inherent to the generation process of the final video itself. Because the auto-encoder is used frame-by-frame, it is completely unaware of any previous generated face that it may have created. This lack of temporal awareness is the source of multiple anomalies. The most prominent is an inconsistent choice of illuminants between scenes with frames, with leads to a flickering phenomenon in the face region common to the majority of fake videos. Although this phenomenon can be hard to appreciate to the naked eye in the best manually-tuned deepfake manipulations, it is easily captured by a pixel-level CNN feature extractor. The phenomenon of incorrect color constancy in CNN-generated videos is a well known and still open research problem in the computer vision field. Hence, it is not surprising that an auto-encoder trained with very constrained data fails to render illuminants correctly.

Technical Overview of the Suggested Detection System

This section presents the overview of an end-to-end trainable recurrent deepfake video detection system. The proposed system is composed by a convolutional LSTM structure for processing frame sequences. There are two essential components in a convolutional LSTM:

1. CNN for frame feature extraction

2. LSTM for temporal sequence analysis

The system learns and infers in an end-to-end manner and, given a video sequence, outputs a probability of it being a deepfake or a pristine video. It has a convolutional LSTM subnetwork, for processing the input temporal sequence. Given an unseen test sequence, a set of features for each frame which are generated by the CNN are obtained. Afterwards, the features of multiple consecutive frames are concatenated and passed to the LSTM for analysis. Finally, an estimate of the likelihood of the sequence being either a deepfake or a non-manipulated video is produced.

Given an image sequence, a convolutional LSTM is employed to produce a temporal sequence descriptor for image manipulation of the shot frame. Aiming at end-to-end learning, an integration of fully-connected layers is used to map the high-dimensional LSTM descriptor to a final detection probability. Specifically, the shallow network consists of two fully-connected layers and one dropout layer to minimize training over-fitting. The convolutional LSTM can be divided into a CNN and a LSTM, which will be describe separately in the following paragraphs.

CNN for Feature Extraction

Inspired by its success in the IEEE Signal Processing Society Camera Model Identification Challenge, we adopt the InceptionV3 [36] with the fully-connected layer at the top of the network removed to provide a deep frame representation by means of the ImageNet pre-trained model. There is no fine-tuning of the network. The 2048-dimensional feature vectors after the last pooling layers are then used as the sequential LSTM input (Figure 2).

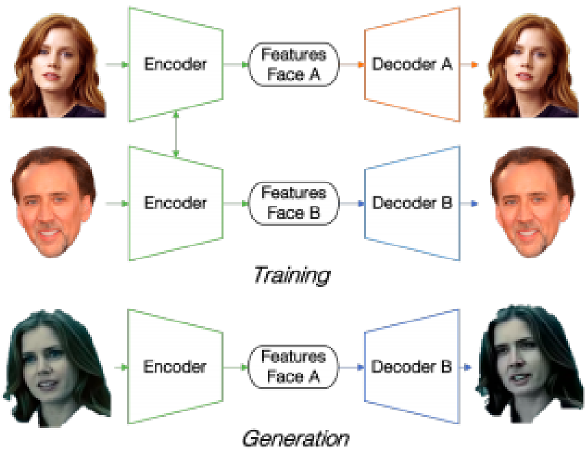

Figure 1. Sample feature extraction

LSTM for Sequence Processing

Let us assume a sequence of CNN feature vectors of input frames as input and a 2-node neural network with the probabilities of the sequence being part of a deepfake video or an untampered video. The key challenge is the design of a model to recursively process a sequence in a meaningful manner. For this problem, a 2048-wide LSTM unit with 0.5 chance of dropout can be used. More particularly, during training, the LSTM model takes a sequence of 2048-dimensional ImageNet feature vectors. The LSTM is followed by a 512 fully-connected layer with 0.5 chance of dropout. Finally, a softmax layer is used to compute the probabilities of the frame sequence being either pristine or deepfake. Note that the LSTM module is an intermediate unit in the pipeline, which is trained entirely end-to-end without the need of auxiliary loss functions (Figure 3).

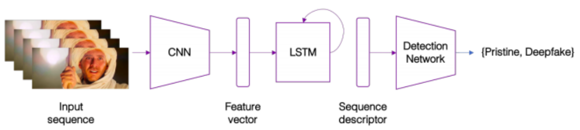

Figure 3. Sample LSTM Sequencing

Experiments

This section includes the details about the experiments. First, the dataset is described. Then, details of the experimental settings are provided to ensure reproducibility and end up by analyzing the reported results.

Dataset

For this work, deepfake videos from multiple video-hosting websites can gathered. HOHA dataset can be used as the source of pristine videos since it contains a realistic set of sequence samples from famous movies with an emphasis on human actions. Given that a considerable number of the deepfake videos are generated using clips from major films, using videos from the HOHA dataset further ensures that the overall system learns to spot manipulation features present in the deepfake videos, instead of memorizing semantic content from the two classes of videos present in the final dataset.

Parameter Settings

First, a random 70/15/15 split can be used to generate three disjoints sets, used for training, validation and test respectively. A balanced splitting can be done, i.e., splitting first for the first half of deepfake videos and then repeating the process for the remaining nonmanipulated videos. This guarantees that each final set has exactly 50% videos of each class, which allows to report the results in terms of accuracy without having to take into account biases due to the appearance frequency of each class or the need of using regularizing terms during the training phase.

In terms of data preprocessing of the video sequences, the following steps are undertaken:

• Subtracting channel mean from each channel.

• Resizing of every frame to 299×299. • Sub-sequence sampling of length N controlling the length of input sequence – N = 20, 40, 80 frames.

This allows to see how many frames are necessary per video to have an accurate detection.

• The optimizer is set to Ad for end-to-end training of the complete model with a learning rate of 1e−5 and decay of 1e−6. 5.3.

Results

It is not unusual to find deepfake videos where the manipulation is only present in a small portion of the video (i.e. the target face only appears briefly on the video, hence the deepfake manipulation is short in time). To account for this, for every video in the training, validation and test splits, continuous subsequences of fixed frame length are extracted that serve as the input of the system.

The experimental results using a large collection of manipulated videos will show whether using a simple convolutional LSTM structure an accurate prediction can be made if a video has been subject to manipulation or not with as few as 2 seconds of video data. This research offers a powerful first line of defense to spot fake media created using the tools describe. Future work should explore how to increase the robustness of the system against manipulated videos using unseen techniques during training.

Discussion

Norris (2000) has proposed that the "process of political communication can be understood as a 'virtuous circle,' a ratcheting process that over the long term gradually reinforces the activism of the active" (p. 309). A more plausible model to understand these relations is one of asymmetrical reciprocal causation (Rojas, 2006). In an asymmetrical reciprocal model, communication variables would have primacy over political variables, i.e. participating in politics might make you more likely to talk about politics in the future, but the relationship between talking about politics today and participating in the future is stronger. Empirical evidence supporting this notion has been reported for political efficacy (Semetko & Valkenburg, 1998), civic participation, (Shah et al., 2005), and political participation (Rojas, 2006). This notion of asymmetrical reciprocal causation is congruent with findings reported by McLeod and colleagues under the rubric of communication mediation (McLeod et al., 1996; E. Puig-i-Abril & H. Roja1999; McLeod, Scheufele, Moy, Horowitz, et al., 1999), and those reported by Ball-Rokeach and colleagues under the notion of "storytelling neighborhood" or communication infrastructure (Ball-Rokeach, Kim, & Matei, 2001; Matei & Ball-Rokeach, 2003; Matei, Ball-Rokeach, & Qiu, 2001).

Future research, of course, needs to strive for multiple measures over time- given the increasing trend of the use of deepfake videos- in a way that provides more solid grounding for the asymmetrical reciprocal causation that underlies the creation of fake videos presented in this study. In addition, future research should strive to provide empirical evidence that sheds light on the multiple explanations that have been discussed here, as well as explore the relationship between political knowledge and online expressive political participation in terms of fake video creation. Is this relationship an artifact of the diffusion stage of the Internet, or does it signal a lowering of the bar for political debate? In conclusion, this paper provides further evidence for the hypothesis that online news media use including creation of fake videos and social interactions facilitate online political engagement. Most importantly, the findings will suggest that whether these online practices come along with increased political participation.

Conclusion

As the technology behind DeepFake keeps evolving, there is a need to improve the detection method. First, it is suggested to evaluate and improve the robustness of the detection method with regards to multiple video compression. Second, despite the current use of a predesigned network structure for this task (e.g., resnet or VGG), for more efficient detection, one should explore dedicated network structure for the detection of DeepFake videos.

References

Alesina, A., & La Ferrara, E. (2000). Participation in heterogeneous communities. Quarterly Journal of Economics, 115, 847-904.

Althaus, S., & Tewksbury, D. (2000). Patterns of Internet and traditional news media use in a networked community. Political Communication, 17, 21-45.

Ball-Rokeach, S. J., Kim, Y. C., & Matei, S. (2001). Storytelling neighborhood: Paths to belonging in diverse urban environments. Communication Research, 28, 392-428.

Berelson, B. R., Lazarsfeld, P. F., & McPhee, W. N. (1954). Voting: A study of opinion formation in a presidential election. Chicago: University of Chicago Press.

Bogart, L., & Orenstein, F. E. (1965). Mass media and community identity in an interurban setting. Journalism Quarterly, 42, 179-188.

Boyle, M. P., Schmierbach, M., Armstrong, C. L., Cho, J., McCluskey, M. R., McLeod, D. M., et al. (2006). Expressive responses to news stories about extremist groups: A framing experiment. Journal of Communication, 56, 271-288.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. (1960). The American voter. New York: Wiley.

Capella, J. N., & Jamieson, K. H. (1997). Spiral of cynicism: The press and the public good. New York: Oxford University Press.

Delli-Carpini, M. X., & Keeter, S. (1996). What Americans know about politics and why it matters. New Haven, CT: Yale University Press.

Dutta-Bergman, M. J. (2005). The Antecedents of community-oriented Internet use: Community participation and community satisfaction. Journal of Computer Mediated Communication, 11(1). Retrieved October 2, 2006, from http://jcmc.indiana.edu/vol11/issue1/dutta_bergman.html

Dutta-Bergman, M. J. (2006). Community participation and Internet use after September 11: Complementarity in channel consumption. Journal of Computer-Mediated Communication, 11(2). Retrieved September 20, 2006, from http://jcmc.indiana.edu/vol11/issue2/dutta-bergman.html

Endersby, J. W., & Towle, M. J. (1996). Tailgate partisanship: Political and social expression through bumper stickers. The Social Science Journal, 33, 307-319. E. Puig-i-Abril & H. Rojas / International Journal of Internet Science 2 (2007), 28-44 40

Granovetter, M. S. (1973). The strength of weak ties. American Journal of Sociology, 78, 1360-1380.

H. Farid. Photo Forensics. MIT Press Ltd, 2016.

I. Goodfellow et al. Generative adversarial nets. Advances in Neural Information Processing Systems, pages 2672–2680, Dec. 2014. Montreal, Canada. ´ 1

Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5967–5976, July 2017. Honolulu, HI.

J. Donahue et al. Long-term recurrent convolutional networks for visual recognition and description. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(4):677–691, Apr. 2017. 2

Jennings, M. K., & Zeitner, V. (2003). Internet use and civic engagement: A longitudinal analysis. Public Opinion Quarterly, 67, 311-334.

Kavanaugh, A. L., Reese, D. D., Carroll, J. M., & Rosson, M. B. (2005).Weak ties in networked communities. The Information Society, 21, 119-131.

Kim, S. H., & Han, M. (2005). Media use and participatory democracy in South Korea. Mass Communication & Society, 8, 133-153.

Kraut, R., Kiesler, S., Boneva, B., Cummings, J., Helgeson, V., & Crawford, A. (2002). The Internet paradox revisited. Journal of Social Issues, 58, 49-74.

Kraut, R., Patterson M., Lundmark V., Kiesler, S., Mukopadhyay, T., & Scherlis W. (1998). Internet paradox: A social technology that reduces social involvement and psychological well being? American Psychologist, 53, 1017-1031.

Krueger, B. S. (2002). Assessing the potential of Internet political participation in the United States. American Politics Research, 30, 476-498.

Krueger, B. S. (2005). Government surveillance and political participation on the Internet. Social Science Computer Review, 23, 439-452.

Kwak, N., Poor, N., & Skoric, M. M. (2006). Honey, I shrunk the world! The relation between Internet use and international engagement. Mass Communication and Society, 9, 189-213.

Lample et al. Fader networks: Manipulating images by sliding attributes. Advances in Neural Information Processing Systems, pages 5967–5976, Dec. 2017. Long Beach, CA. 2

Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, June 2008. Anchorage, AK.

Lu, Y.-W. Tai, and C.-K. Tang. Conditional cyclegan for attribute guided face image generation. arXiv:1705.09966, May 2017. 2.

N. McLaughlin, J. M. d. Rincon, and P. Miller. Recurrent convolutional network for video-based person reidentification. Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition, pages 1325–1334, June 2016. Las Vegas, NV.

Noelle-Neumann, E. (1993). The spiral of silence. Public opinion–our social skin (2nd ed.). Chicago, IL: The University of Chicago Press.

Norris, P. (2000). A virtuous circle: Political communications in post-industrial democracies. Cambridge: Cambridge University Press.

Norris, P. (1996). Does television erode social capital? A reply to Putnam. PS: Political Science & Politics, 293, 474-480. Pizano, L. (2003). Bogotá y el Cambio: Percepciones sobre la Ciudad y la Ciudadanía. Bogotá: IEPRI - CESO.

Polat, R. K. (2005). The Internet and political participation: Exploring the explanatory links. European Journal of Communication, 20, 435-459.

Putnam, R. D. (1995). Bowling alone: America's declining social capital. Journal of Democracy, 1, 65-78.

Putnam, R. D. (1996). The strange disappearance of civic America. The American Prospect, 24, 34-48.

Putnam, R. D. (2000). Bowling alone: The collapse and revival of American community. New York: Simon & Schuster.

Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch. Transferable deep-cnn features for detecting digital and print-scanned morphed face images. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 1822–1830, July 2017. Honolulu, HI. 2

Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen. Distinguishing computer graphics from natural images using convolution neural networks. Proceedings of the IEEE Workshop on Information Forensics and Security, pages 1–6, Dec. 2017. Rennes, France. 2

Rossler et al. Faceforensics: A large-scale video dataset ¨ for forgery detection in human faces. arXiv:1803.09179, Mar. 2018. 2

Tewari et al. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. Proceedings of the IEEE International Conference on Computer Vision Workshops, pages 1274–1283, Oct. 2017.

Scheufele, D. A., Shanahan, J., & Lee, E. (2001). Real talk: Manipulating the dependent variable in spiral of silence research. Communication Research, 28, 301-24.

Stanyer, J. (2005). The British public and political attitude expression: The emergence of a self-expressive political culture? Contemporary Politics, 11, 19-32.

Taveesin, N. J., & Brown, W. J. (2006). The use of communication technology in Thailand's political process. Asian Journal of Communication, 16, 59-78.

Walsh, K. C. (2003). Talking about politics: Informal groups and social identity in American life. Chicago and London: The University of Chicago Press.

Wellman, B., Quan-Haase A., Boase, J., Chen, W., Hampton, K., Isla de Diaz, I., et al. (2003). The Social affordances of the Internet for networked individualism. Journal of Computer Mediated Communication, 8(3). Retrieved March 10, 2007, from http://jcmc.indiana.edu/vol8/issue3/wellman.html

Wellman, B., Quan-Haase, A., Witte, J., & Hampton, K. (2001). Does the Internet increase, decrease or supplement social capital? American Behavioral Scientist, 45, 436-455.

Wuthnow R. (1994). Sharing the journey: Support groups and America's new quest for community. New York: Free Press.

Wyatt, R. O., Katz, E., & Kim, J. (2000). Bridging the spheres: Political and personal conversation in public and private spaces. Journal of Communication, 50, 71-92.