Lighten the Face Forgery Detection Network by using MobileNet

Shibo Wang
wangshibo1993@tamu.edu

Chih-Peng Wu
chinuy@tamu.edu

1. Introduction

The maturity of the deep learning techniques make it possible to "fake" videos that are almost impossible to be identified by human eyes. The so-called *Deepfake* is a technique for image and video synthesis that allows users to replace the face of a celebrity (or anyone else) in a video with another face from another person, regardless of gender, hair style, race, and other seemingly impossible shapes, as Fig. 1 demonstrated.

The *Deepfake* is so accessible already that there are productions (e.g., FaceApp [1] and FakeApp [2]) that can easily be acquired by general users who don't know how to control computer programs and can create highly "real" video and images. Those video generated by *Deepfake* have cause many social problems. For example, female celebrities were transformed into porn videos as porn stars [10]. There are more penitential misbehaved usage, such as fake news, fake surveillance videos, pornographic.

According to the malicious report by Brundage et al. [5], the growing use of AI systems has led to **Introduction to new threats**, and researchers should be also responsible to study how to mitigate the threats.

Figure 1. Example of *Deepfake* video (source: [14])

2. Related work

An authenticated digit media may have tampered when users share to a different website, social media and so on. Using deep convolutional neural networks (CNNs) to identify the authentic of camera images was proved feasible [15][7][8]. For video, progress was made through finding computationally cheap manipulations, for example, dropped or duplicated frames [16] or copy-move manipulations [4].

Several recent works have already given different neural networks to make the classification of real and fake pictures and videos. Rossler et al. created and maintained a large scale video dataset for forgery detection [12], providing over 0.5 million edited images. And they used the Xception Net which is a complicated traditional CNN to approach the highest classification accuracy to date. Besides CNNs, another new and more complicated neural network – Capsule network forensics [11] – was also proposed and can have a better performance than some works based on traditional CNNs [6] [3].

3. Challenges

There are two challenging problems in current fake face detection. The first problem is there has already existed and maybe emerge more different mechanisms that can transfer one face to another, like *Deepfake*, *Face2Face*, and *FaceSwap*. Thus, the training dataset needs to maintain enough diversity. This problem seems to be solved to some degree since many open-source codes provide their dataset and also Google has just released a bunch of fake videos to help people construct their dataset. And the second problem seems to be more severe. Those networks with outstanding accuracy are always complicated and need to cost much computing sources. Since several relevant mobile Apps of *Deepfake* have already been launched, our *Deepfake*-detection should also start paying attention to mobile apps and provide some lightweight networks.

4. Action

To lighten the network, we propose using the MobileNet [9][13] to replace traditional CNN, which splits the traditional CNN into depth-wise convolution and point-wise convolution to build lightweight deep neural networks. The structure of MopbileNet is shown in Fig. 2. Based on [9], the ratio of the computational volume of MobileNet to the

traditional CNN is only about 0.1189. Besides using depthwise separable convolution to reduce the number of calculations and parameters, a recent and more advanced version, MobileNet V2 [13], also introduces the shortcut that has already succeeded in many popular networks like Resnet and Densenet. To adapt the MobileNet to shortcut, two tricks, inverted residuals and linear bottlenecks, are introduced to mitigate feature degradation. The difference between the MobileNet V2 and the ResNet are shown in Fig. 3.

5. Resolution

Our current work based on the dataset in MesoNet [3]. We will then collect more fake faces cut from different existed datasets including the Google-*Deepfake*-detection video set. After that, we will evaluate the proposed method on assigned validation and test set. Currently, we have replicated the traditional CNN called Mesonet [3] successfully. In the future, we will replace CNN with MobileNet V2, and then further tune and expand the network. Finally, we will evaluate the structures using Top-1 and Top-5 accuracy and submit the network structure with the best performance as the final results.

References

[1] Faceapp. `https://www.faceapp.com/`. Accessed: October 2019.

[2] Fakeapp. `https://www.fakeapp.org/`. Accessed: October 2019.

[3] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen. Mesonet: a compact facial video forgery detection network. In *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pages 1–7. IEEE, 2018.

[4] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro. Local tampering detection in video sequences. In *2013 IEEE 15th International Workshop on Multimedia Signal Processing (MMSP)*, pages 488–493. IEEE, 2013.

[5] M. Brundage, S. Avin, J. Clark, H. Toner, P. Eckersley, B. Garfinkel, A. Dafoe, P. Scharre, T. Zeitzoff, B. Filar, et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. *arXiv preprint arXiv:1802.07228*, 2018.

[6] D. Cozzolino, G. Poggi, and L. Verdoliva. 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*, pages 159–164. ACM, 2017.

[7] D. Güera, Y. Wang, L. Bondi, P. Bestagini, S. Tubaro, and E. J. Delp. A counter-forensic method for cnn-based camera model identification. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1840–1847. IEEE, 2017.

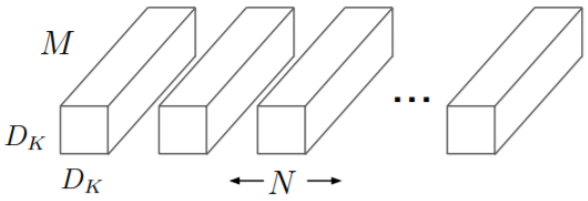
[8] D. Güera, F. Zhu, S. K. Yarlagadda, S. Tubaro, P. Bestagini, and E. J. Delp. Reliability map estimation for cnn-based camera model attribution. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 964–973. IEEE, 2018.

[9] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

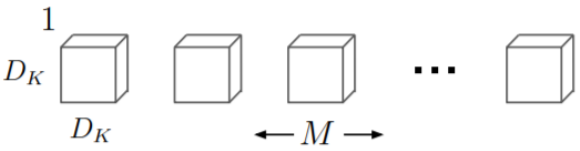
[10] L. Matsakis. Artificial intelligence is now fighting fake porn. `https://www.wired.com/story/gfycat-artificial-intelligence-deepfakes/`, 2018. Accessed: October 2019.

[11] H. H. Nguyen, J. Yamagishi, and I. Echizen. Capsule-forensics: Using capsule networks to detect forged images and videos, 2018.

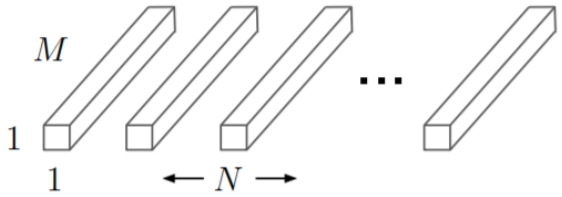
[12] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner. Faceforensics: A large-scale video dataset



(a) Standard Convolution Filters



(b) Depthwise Convolutional Filters



(c) 1×1 Convolutional Filters called Pointwise Convolution in the context of Depthwise Separable Convolution

Figure 2. The architecture of MobileNet [9]. The standard convolutional filters in (a) are replaced by two layers: depthwise convolution in (b) and pointwise convolution in (c) to build a depthwise separable filter.

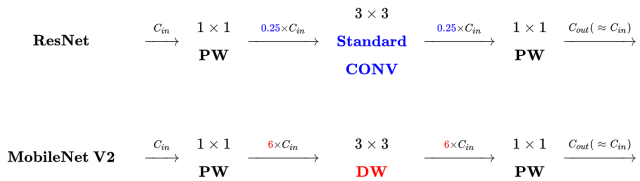


Figure 3. ResNet and MobileNet V2 comparison

for forgery detection in human faces. *arXiv preprint arXiv:1803.09179*, 2018.

[13] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4510–4520, 2018.

[14] H. Schellmann. Deepfake videos are getting real and that's a problem. `https://www.wsj.com/articles/deepfake-videos-are-ruining-lives-is-democracy-next-1539595787`, 2018. Accessed: October 2019.

[15] A. Tuama, F. Comby, and M. Chaumont. Camera model identification with the use of deep convolutional neural networks. In *2016 IEEE International workshop on information forensics and security (WIFS)*, pages 1–6. IEEE, 2016.

[16] W. Wang and H. Farid. Exposing digital forgeries in interlaced and deinterlaced video. *IEEE Transactions on Information Forensics and Security*, 2(3):438–449, 2007.