Fake Image and Video Detection Using Capsule Network

Mrs. Smitha P[1], B G Sumith Kumar[2], Chaithra K M[3], Deepika R[4], Varun S Naik[5]
[1]Assistant Professor, Department of Information Science and Engineering, East West Institution of Technology, Visvesvaraya Technological University, Karnataka, India
[234]Student, Department of Information Science and Engineering, East West Institution of Technology, Visvesvaraya Technological University, Karnataka, India

***Abstract* - Attackers can now easily manufacture forged photos and movies thanks to recent developments in media generating algorithms. The development of a forged version of a single film collected from a social network can be done in real time using state-of-the-art technology. Although several methods for identifying faked photos and videos have been developed, they are often aimed at specific areas and quickly become obsolete as new types of attacks emerge. The method shown in this project use a capsule network to detect various types of spoofs, ranging from replay attacks utilizing printed images or recorded films to deep convolutional neural networks-based computer-generated videos. It broadens the scope of capsule networks' application beyond their original goal of tackling inverse graphics difficulties.**

***Index Terms* - forged, state-of-the-art, capsule network, spoofs, convolutional neural network.**

I.INTRODUCTION

People have been interested in modifying photographs since the discovery of photography, primarily to rectify or emphasize flaws in the images. However, we have progressed much beyond these fundamental alterations, including the addition of unreal individuals and the insertion and removal of items. Digital photography has made the process of image manipulation much easier, especially with the help of professional tools like Adobe Photoshop.

As a result, persons with bad intent can readily use these modern technologies and data to produce phony photographs and films, which they can then extensively distribute on social media or use to overcome face authentication.

We offer a method for detecting forged images and videos in a variety of forging scenarios, including replay attack detection and (both fully and partially) computer-generated image/video detection, that employs a capsule network.

II. RELATED WORK

On the basis of the features employed and the target, we divide forgery detection methodologies into replay attack detection and computer-generated image/video detection in this section. It's worth noting that certain tactics are two-fold, while others are limited to specific types of attacks. We also cover the basics of capsule networks and the dynamic routing mechanism that made them possible.

LBP approaches were the principal defense against replay attacks prior to the present deep learning age [7, 8]. Kim et al. [9] proposed an approach based on local patterns of diffusion speed (local speed patterns) that provides higher accuracy than LBP-based methods. The capacity to detect replay attacks has considerably increased since the introduction of deep learning. A support vector machine is used by Yang et al. [10] to categorize features retrieved by a pre-trained convolutional neural network. Menotti et al. [11] utilize a similar technique but enhance the filters in a high-performance CNN architecture that is already available. Alotaibi and Mahmood [12] apply nonlinear diffusion in their own CNN, which is based on an additive operator splitting scheme. Ito et al. [13] recently published an approach that uses a pre-trained CNN and uses the entire image rather than just the extracted facial region.

For example, a deep fake methodology for face swapping [6], the Face2Face method for facial reenactment [1], or the deep video portraits technique [2] are all state-of-the-art ways for detecting images or videos made by a computer for the purpose of forging. Fridrich and Kodovsky [14] proposed a steganalysis approach based on hand-crafted feature noise that may also be used for forgery detection. A CNN variant of this strategy was implemented by Cozzolino et al. [15].

III. IMPLEMENTATION

This research describes a method for detecting forged images and videos using a capsule network in a variety of forgery scenarios, including replay attack detection and (both fully and partially) computer-generated image/video detection. This is ground-breaking work in the use of capsule networks to digital forensics difficulties. Capsule networks were originally created to handle computer vision problems. The higher performance of the suggested strategy was demonstrated by a comprehensive assessment of state-of-the-art related work and intensive comparisons utilizing four significant datasets.

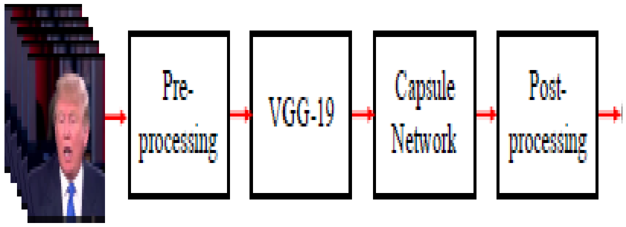

Fig 1. Architecture of proposed work

Capsule-forensics

Both photos and videos can be used with the suggested technique (Fig. 1). In the pre-processing phase for video input, the video is split into frames. The frames are then used to obtain the classification results (posterior probabilities). To reach the final result, the probabilities are averaged in the post-processing phase. The remaining components are built in the same way that they are when the input is an image.

Faces are recognized and scaled to 128 x 128 in the pre-processing step. We extract the latent features, which are the inputs to the capsule network, using a component of the VGG-19 network [27], as we did in our earlier work [12]. We take the output of the third maxpooling layer instead of three outputs before the ReLU layers, as we did in our earlier work.

Capsule Design

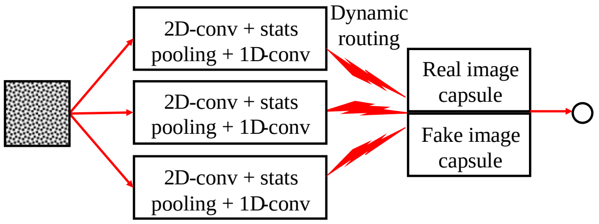
2D-conv + stats pooling + 1D-conv
2D-conv + stats pooling + 1D-conv
2D-conv + stats pooling + 1D-conv
Dynamic routing
Real image capsule
Fake image capsule

Fig. 2. Overall design of capsule-forensics network.

Three primary capsules and two output capsules, one for actual and one for fake images, make up the proposed network (Fig. 2). The inputs, which are dispersed to the three major capsules, are latent properties recovered by part of the VGG-19 network [27]. (Fig. 3). Statistical pooling, which is critical for forgery detection, is applied, as it was in our prior work [12].

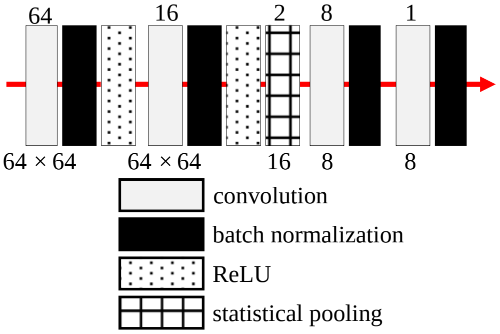

Fig. 3. Detailed design of primary capsule. Upper numbers indicate number of filters (depth) while lower number indicate size of outputs of corresponding filters.

Algorithm

procedure ROUTING(uj|i,W,r) Wˆ ← W + rand(size(W)) uˆj|i ← Wˆisquash(uj|i) . Wi ∈ Rm×n for all input capsule i and all output capsules j do bij ← 0

for r iterations do

for all input capsules i do ci ← softmax(bi) for all output capsules j do sj ← Pi cijuˆj|i for all output capsules j do vj ← squash(sj) for all input capsules i and output capsules j do

bij ← bij + uˆj|i · vj

return vj

Algorithm 1 was used to route the data to the output capsules (vj) for r iterations. The network has about 2.8 million parameters, which is a small quantity for a network of this size. Via adding Gaussian random noise to the 3-D weight tensor W and applying one more squash (equation 1) before routing by iteration, we enhanced the technique of Sabour et al. [15]. The additional noise reduces overfitting, while the additional equation keeps the network stable.

IV. RESULTS

We examined the proposed method with and without random noise to determine the benefit of employing random noise (Capsule-Forensics-Noise and Capsule-Forensics). The random noise was created using a normal distribution N(0,0.01) and was only utilised for training. The dynamic routing technique employed two rounds (r = 2). We used the half total error rate (HTER) $\left(\frac{FRR+FAR}{2}\right)$ and accuracy $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$ as metrics.

Replay Attack Detection

On the well-known Idiap REPLAY-ATTACK dataset [7], we compared the performance of the proposed technique to that of eight state-of-the-art detection methods to establish its capacity to detect replay attacks. The proposed technique with random noise (Capsule-Forensics-Noise), as well as our earlier method [12], both had an HTER of zero, as indicated in Table 1.

Method	HTER (%)
Chigovska et al. [7]	17.17
Pereira et al. [8]	08.51
Kim et al. [17]	12.50
Yang et al. [18]	02.30
Menotti et al. [19]	00.75
Alotabib et al. [20]	10.00
Ito et al. [21]	00.43
Nguyen et al. [12]	00.00
Capsule-Forensics	00.28
Capsule-Forensics-Noise	00.00

Table 1. Half total error rate of state-of-the-art detection methods on REPLAY-ATTACK dataset [7].

V. CONCLUSION

Fake news media can be created using forged photos and videos to go around facial authentication. With the development of complex network topologies and the utilization of vast amounts of training data, the quality of modified photos and videos has significantly improved. We offer a Capsule-Forensics approach for detecting computer-manipulated/generated images and videos, as well as detecting presentation attacks, in digital images and video forensics. The proposed method's capacity to withstand adversarial machine attacks, particularly the proposed random noise at test time, will be evaluated and improved in future study. It will also concentrate on making the suggested method resistant to mixed attacks, detecting abnormalities, and promoting awareness of this important subject among researchers.

REFERENCES

[1] “H. Kim, P. Garrido, A. Tewari, W. Xu, J. Thies, M. Nießner, P. P´erez, C. Richardt, M. Zollh¨ofer, and C. Theobalt, “Deep video portraits,” in International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH). ACM, 2018”.

[2] “O. Fried, A. Tewari, M. Zollh¨ofer, A. Finkelstein, E. Shechtman, D. B. Goldman, K. Genova, Z. Jin, C. Theobalt, and M. Agrawala, “Text-based editing of talking-head video, in International Conference and Exhibition on Computer Graphics and Interactive Techniques (SIGGRAPH). ACM, 2019”.

[3] “D. Cozzolino, G. Poggi, and L. Verdoliva, “Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection,” in Workshop on Information Hiding and Multimedia Security (IH&MMSEC). ACM, 2017.”

[4] H. H. Nguyen, J. Yamagishi, and I. Echizen, “Capsuleforensics: Using capsule networks to detect forged images and videos,” in International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 2307–2311.

[5] C. Xiang, L. Zhang, Y. Tang, W. Zou, and C. Xu, “Mscapsnet: A novel multi-scale capsule network,” IEEE Signal Processing Letters, vol. 25, no. 12, pp. 1850– 1854, 2018.

[6] M. T. Bahadori, “Spectral capsule networks,” in International Conference on Learning Representations (ICLR), 2018

[7] Ivana Chingovska, Andre Anjos, and S ´ ebastien Marcel, ´ “On the effectiveness of local binary patterns in face anti-spoofing,” in BIOSIG, 2012.

[8] Tiago de Freitas Pereira, Andre Anjos, Jos ´ e Mario ´ De Martino, and Sebastien Marcel, “Can face anti- ´ spoofing countermeasures work in a real world scenario?,” in ICB. IEEE, 2013.

[9] Wonjun Kim, Sungjoo Suh, and Jae-Joon Han, “Face liveness detection from a single image via diffusion speed model,” IEEE TIP, 2015

[10] Jianwei Yang, Zhen Lei, and Stan Z Li, “Learn convolutional neural network for face anti-spoofing,” arXiv preprint arXiv:1408.5601, 2014.

[11] David Menotti, Giovani Chiachia, Allan Pinto, William Robson Schwartz, Helio Pedrini, Alexandre Xavier Falcao, and Anderson Rocha, “Deep representations for iris, face, and fingerprint spoofing detection,” IEEE TIFS, 2015.

[12] Aziz Alotaibi and Ausif Mahmood, “Deep face liveness detection based on nonlinear diffusion using convolution neural network,” Signal, Image and Video Processing, 2017.

[13] Koichi Ito, Takehisa Okano, and Takafumi Aoki, “Recent advances in biometrics security: A case study of liveness detection in face recognition,” in APSIPA ASC. IEEE, 2017.

[14] Jessica Fridrich and Jan Kodovsky, “Rich models for steganalysis of digital images,” IEEE TIFS, 2012.

[15] Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva, “Recasting residual-based local descriptors as convolutional neural networks: an application to image forgery detection,” in IH&MMSEC. ACM, 2017.