Inteligencia Artificial 26(72) (2023), 1-14
doi: 10.4114/intartif.vol26iss72pp,1-14

INTELIGENCIA ARTIFICIAL

http://journal.iberamia.org/

Deep Learning Approach for Detecting Fake Images Using Texture Variation Network

Haseena S [1], Saroja S [2], Shri Dharshini D [3], Nivetha A [4]

[1,3,4] Department of Information Technology, Mepco Schlenk Engineering College, India
[2]Department of Computer Applications, National Institute of Technology, Trichy, India

Abstract. Face manipulation technology is rapidly evolving, making it impossible for human eyes to recognize fake faces in photos. Convolutional Neural Network (CNN) discriminators, on the other hand, can quickly achieve high accuracy in distinguishing fake from real face photos. In this paper, we investigate how CNN models distinguish between fake and real faces. According to our findings, face forgery detection heavily relies on the variation in the texture of the images. As a result of the aforementioned discovery, we propose a deep texture variation network, a new model for robust face fraud detection based on convolution and pyramid pooling. Convolution combines pixel intensity and pixel gradient information to create a stationary representation of composition difference information. Simultaneously, multi-scale information fusion based on the pyramid pooling can prevent the texture features from being destroyed. The proposed deep texture variation network outperforms previous techniques on a variety of datasets, including Faceforensics++, DeeperForensics-1.0, Celeb-DF, and DFDC. The proposed model is less susceptible to image distortion, such as JPEG compression and blur, which is important in this field.

Keywords: Fake images, Deep Learning, Texture Variation Network, CNN, forgery

1 Introduction

The rapid development of face modification technology has fueled the rise in forgery face images and videos [1]. DeepFakes [2], Face2Face [3], and particularly learning-driven generative models like Generative Adversarial Nets (GAN) [4] may produce lifelike forgeries of facial photos and videos that are impossible to distinguish even with human eyes. Figure 1 shows several compelling examples from four datasets: Faceforensics++, DeeperForensics-1.0, Celeb-DF and DFDC. Deep learning is utilized to the full extent in all these datasets' as part of manipulation methods. However, it is dangerous when these approaches are used for evil purposes, such as fake news, reputation infringement, or even political purposes.

ISSN: 1137-3601 (print), 1988-3064 (on-line)
©IBERAMIA and the authors

Figure 1: Examples of real and fake images across four datasets, Faceforensics++, DeeperForensics-1.0, Celeb-DF and DFDC. The real images are in top row, and fake images are in bottom row.

As a result, developing more effective ways to detect face forgeries is critical. For the most popular classical operations, such as splicing, copy-move, and removal, many approaches [5]–[7] have been developed in earlier research. Most works use hand-crafted features such as illumination color [5, 6], color filter array patterns [6, and blur type inconsistency [7] to classify a specific patch of an image as tampered or not. These features emphasize the differences between actual and fraudulent photos while concentrating on a specific tampering method. These methods have inspired deep learning-based face forgery detection and given a solid theoretical platform for picture forensics research. Direct use of these strategies, however, may be futile in the face of today's massive fraud.

In recent years, deep learning has become widely used in computer vision applications such as object detection [8]. Deep learning-based approaches for detecting face forgeries have also been developed by researchers [9–12]. An explainable and resilient model must adapt to this issue, considering the influencing elements (distortion and compression) in the real world. Our effort is motivated by two factors. On the one hand, counterfeit clues might show up in a variety of places on the face. On the other hand, we're curious about the Convolutional Neural Network (CNN) approach to identifying fake and real facial photos. In terms of the first motive, we offer some examples in Fig. 1 that show how counterfeit clues can occur in a variety of places on the face.

Under these two motivations, this work starts with the learning focus on CNN's behaviour in the face forgery detection task, and we find that texture variation network information plays an important role in face forgery detection. Thus, a texture variation network model is proposed for face forgery detection. First, we use Cascaded Convolutional Neural Networks (CCNN) [13] to crop faces from video frames and use a crop (extended by a factor of 1:3) around the centre of the tracked face. The cropped faces are then fed into a module that extracts the composition difference features. The pixel intensity information and the pixel gradient information are used to represent composition difference information. A special convolution operation is used to combine the intensity and gradient information. Next, we extract multi-scale information through a module. Finally, we fuse the extracted composition difference features at different scales for classification. The major contributions of our work can be outlined as follows:

- To present a stationary description of composition difference information, we use pixel intensity and pixel gradient information. We take advantage of a particular convolution operation based on convolution [14] to extract the practical characteristics. To the best of our knowledge, this is the first attempt to introduce specific convolution techniques for feature extraction and information fusion in the domain of face forgery detection.

- For face forgery detection, we propose a texture variation network. The goal of our model is to extract and merge texture variation networks. To evaluate our technique, we ran extensive tests on fur benchmark datasets: Faceforensics++ [10], DeeperForensics-1.0 [16], Celeb-DF [17], and DFDC [18]. The results of experiments show that our strategy outperforms others. Furthermore, with realistic data with strong compression and mixed distortion, our technique performs better. The robustness and feasibility of a face forgery detection system are significantly improved by the proposed method.

2 Related Works

This section describes related work in detail.

A. Facial Alteration Techniques:

There are numerous public datasets of genuine faces available nowadays, including CelebA, CASIA-WebFace, and others. Face data in these datasets is derived from publicly available data for research purposes and does not infringe on personal privacy. Researchers have built some public facial manipulation datasets as deep learning-based facial manipulation algorithms become more advanced. Few datasets like Faceforensics++, DeeperForensics-1.0, Celeb-DF, DFDC consists of fake images. The quality of these facial alteration datasets varies, as do the facial manipulation techniques used in each case. They are, however, all based on data provided by the facial alteration techniques listed below.

Deepfakes is the name given to a specific facial manipulation technique. The goal of this technique is to replace the target sequence's faces with the faces of the source movies or photographs. This approach is currently implemented in a number of open-source projects, including GitHub [2] and Fakeapp. A common encoder and two automatic decoders are used in this system. The automatic decoder is trained to reconstruct the training images of the source and target faces, while the shared encoder is trained to encode the features of the source and target faces. Following that, the decoders corresponding to the source and target faces are swapped, and the resulting model may swap the source and target faces.

To display face replication, NeuralTextures [19] is a facial manipulation technology based on the neural texture rendering method. It extracts the target face's neural texture from the raw video data and trains the rendering network that corresponds to it. The network employs a photometric reconstruction loss in conjunction with an adversarial loss during the training process. To achieve a flawless reconstruction effect, the database Faceforensics++ employs a patch-based GAN-loss similar to Pix2Pix [20].

FaceSwap [1] is a graphics-based approach for transferring facial regions from one image to another. It extracts facial sections based on facial landmarks that are sparsely recognized. These landmarks are used in the procedure, which mixes shapes to suit the 3D template model. Researchers have developed some ways that mix GAN with other methods in order to achieve better outcomes. Researchers improved the generation effect and solved the image blur and video jitter by adding adversarial loss and perceptual loss to the

Variational Auto-Encoder (VAE) [21]. Wang et al. [22] improved the Pix2Pix approach with CycleGAN [23], which considerably improved the quality and details of generated face images.

Face2Face [3] is a face reproduction system that can reproduce the facial expressions of the characters in the source video onto the faces of the characters in the target video while maintaining the identity information of the target character. In other words, Face2Face can exchange only the facial expressions of the characters [60] [61]. This method works by selecting the key frames in the video stream and then extracting the face information of the character. It then transfers the expression through the mapping relationship and then re-synthesizing the human face under different conditions (lighting and expression)[62].

B. Digital Evidence Methods:

Scholars have developed some methods based on common statistical features to detect facial tampering, inspired by the field of classic picture forensics and steganalysis. Face forgery detection systems based on deep learning have proved their benefits in recent years. The majority of classic picture forensics approaches are based on statistical data or hand-designed routines. Scholars have developed a number of efficient detection methods for the traditional picture manipulation outlined in [24]. To detect manipulated photos, Lukas et al. [25] leveraged the uniqueness of the association between the camera's fixed pattern noise and the source device. Cozzolino et al. [26] advocated using summed noise statistics to detect image splicing. To conceal picture content, Stamm and Bayar [27] presented limited convolutional neural networks. This strategy laid a solid foundation for future forensic investigations. In [28], light sources were utilized to detect the direction of scene lighting, and lighting discrepancies were used to assess whether the image had been manipulated. Other approaches looked at JPEG compression artefacts [29], [30], and resampling traces [31].Fridrich and Kodovsky [32] developed a custom function to scan features with a pixel radius of 2 as well as the image's horizontal and vertical directions using a high-pass filter, and then used these features to train a linear Support Vector Machine (SVM) classifier. Setting noise patterns on EXIF entries [33] or directly searching for unqualified noise [34] are two examples of neural network-based detection approaches. Zhou et al. [35] suggested a dual-stream network that can detect tampered photos and distinguish between the methods used to tamper with them. These methods have inspired deep learning-based forged picture detection and given a solid theoretical platform for image forensics research. The application of existing methods in deep forgery detection is becoming increasingly difficult as GANs advance. Some studies [36], [37] [59] used deep networks to learn discriminative features or uncover manipulation traces to detect face fraud as forgery faces have gotten more realistic. According to Li et al. [38], GAN-based films did not blink. In [39], a rudimentary deep-fake detection network called MesoNet was constructed, which yielded some interpretable results. As a binary face forgery picture detector, Rossler et al. [10] introduced the successful Xception-Net. Li et al. [40] focused on artefacts or splicing traces formed when altered photos were created, and they had excellent results. Amerini et al. [41] used the difference in optical flow field as a cue to distinguish between deepfakes and real footage, taking into account probable abnormalities in the sequence's time dimension. Durall et al. [42] shifted the research focus from the spatial to the frequency domain, employing the power spectrum as a forensic image forensic characteristic. Frequency-aware Decomposition [43] presents two characteristics based on frequency domain design (FAD).

Tolosana et al. [44] provided a comprehensive analysis of first and second-generation deepFakes about facial regions and false detection performance, providing a good reference for further research. Liu et al. [45] focus on the overall texture and gamut module in the network. This method is remarkable because it improves endurance and provides direction for recent research. Kumar et al. [46] propose an approach based on measurement methodology, which provides an important basis to look up the next classifier in face spoofing detection. Some recent studies have used biological information [47], such as heart rate information [48], for deep video detection.

3 Proposed Method

The block diagram of the proposed work is shown in Figure 2. In this work, we present a multi-scale texture difference model that is motivated by the ability of CNN models to predict the authenticity of a facial image.

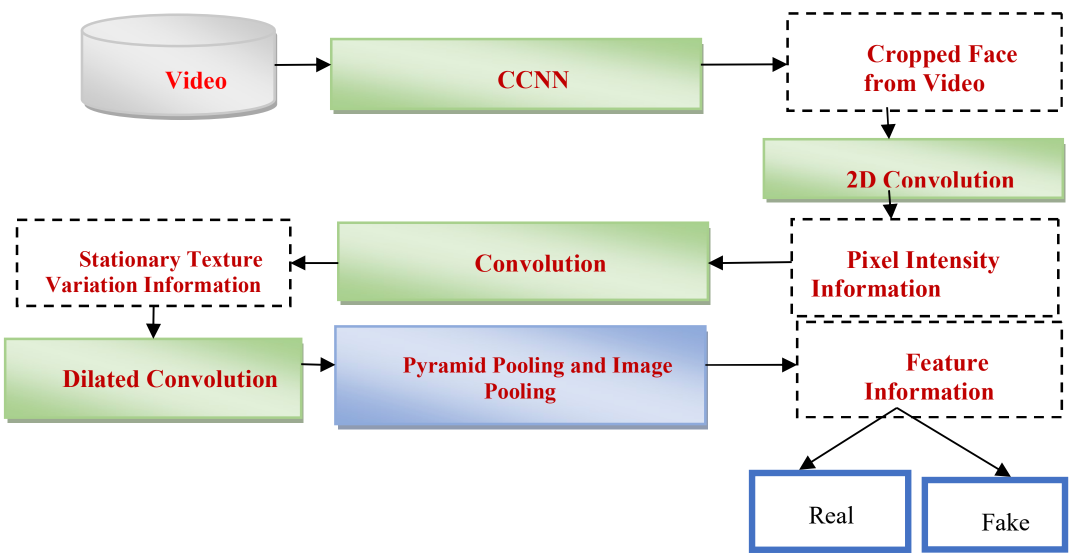


Figure 2: Block Diagram of the proposed work

A. Composition difference Features Extraction:

The extraction of composition difference features tries to make use of textural discriminative information. Gram-Net [45] previously used the gram matrix to extract global composition difference information from a GAN-generated complete image. However, a new definition of composition difference information has been developed. is required for the face of locally generated images. In our case, we use the pixel in our work, following the principle of convolution [14]. To supply the pixel gradient information and the intensity information, the composition difference information is described in a stationary manner. The pixel intensity information is extracted using the most basic 2D convolution. The following is the 2D convolution technique for obtaining the next-level feature map M_{l+1} from the feature map M_l:

$$M_{l+1}(s_c) = \sum_{s \in r} \omega(s_r) M_l(s_c + s_r) \tag{1}$$

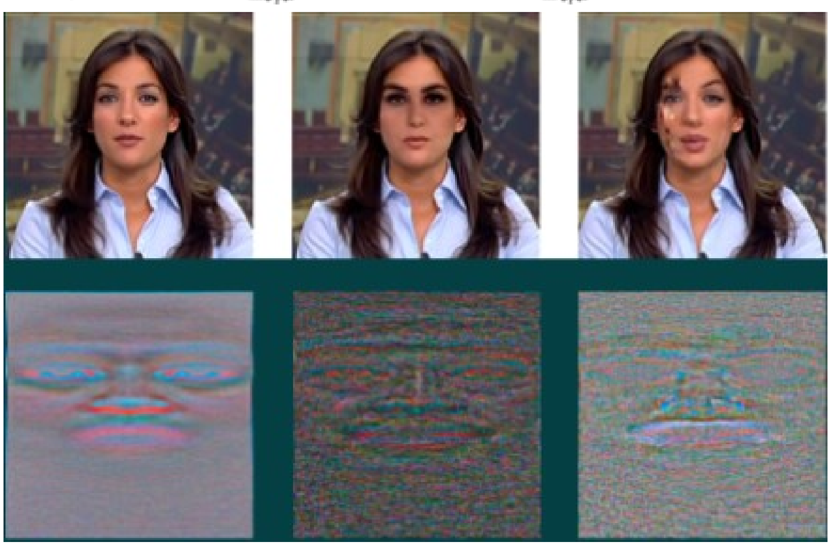
On both the input feature map M_l and the feature map M_{l+1}, S_c represents the current position. For 2D convolution, r is the local receptive field region, and S_r enumerates the positions in r. For example, S_r (1, 1), (1, 0), (0, 1), (1, 1) with the 3x3 kernel and dilation 1, and W (S_r) represents the weight.

We use convolution to provide a stationary description for the pixel gradient information. This combines the idea of difference with the convolution operation to enhance the ability of features to express pixel gradient information. The following equation shows how convolution extracts the next-level feature map M_{l+1} from the feature map M_l.

$$M_{l+1}(s_c) = \sum_{s_r \in r} \omega(s_r)(M_l(s_c + s_r) - M_l(s_c)) \tag{2}$$

Then, with a parameter of [0, 1], we combine 2D convolution and the convolution.

$$M_{l+1}(s_c) = (1 - \alpha) \sum_{s_r \in r} \omega(s_r) M_l(s_c + s_r) + \alpha \sum_{s_r \in r} \omega(s_r)(M_l(s_c + s_r) - M_l(s_c)) \tag{3}$$



(a) (b) (c)

Figure 3: The visual image obtained using the method in [52]. We can clearly see that the extracted features are different. From left to right are the fakeimage, the features extracted by convolution, and the features extracted by 2DC. (a) Full face. (b) Edge of scarf (real area). (c) Eyes (fake area).

After decomposition and merging, the formula is as follows:

$$M_{l+1}(s_c) = \sum_{s_r \in r} \omega(s_r) M_l(s_c + s_r) - \alpha \sum_{s_r \in r} \omega(s_r) M_l(s_c) \tag{4}$$

We refer to the work [14] in face anti-spoofing (FAS) and designate the generalised defined by Equ. 4 the full form of convolution. It's worth noting that during the specific installation procedure, convolution does not increase the number of network parameters.ResNet-18's convolutions were replaced with central difference convolutions, and the model was trained using the same settings as in Section III-A. In the convergent model, we extracted the features of the last convolutional layer. The fake image, the features extracted by convolution, and the features extracted by 2DC are displayed in Figure 3(a), (b), and (c), from left to right. As shown in Figure 3(b), the edges of the character's scarf represent the genuine region in the image, and the properties of convolution are stronger than those of 2 DC. In Figure 3(c), the character's eyes indicate the fake area, and the features of convolution mirror the image's genuine information more than the features of 2DC. The convolution-extracted feature map can more accurately reflect the fundamental aspects of forged facial photos.

B. Scale Extraction:

Multi-scale information is supposed to take advantage of discriminative information from multiple scales. To extract different-scale features from the feature map FT, which is the result of the composition difference features module, we employ a unique convolution operation called dilated convolution [53]. The following is the procedure for dilated convolution:

$$DC_d(s) = \sum_{t\epsilon r} M_T(s + dt)\omega(t) \tag{5}$$

We refer to the work [15], which introduces PP and image pooling.

$$DC_{Image} = U(\delta(M_T)) \tag{6}$$

$$\mathrm{DC} = \delta(\mathrm{concat}(\mathrm{DC}_{\mathrm{Image}}, [\delta(\mathrm{M_T}), \mathrm{DC}_6, \mathrm{DC}_{12}, \mathrm{DC}_{18}])) \tag{7}$$

C. The Networks for Face Manipulation Detection:

The goal of our research is to develop a function, Yfake/real = P(xface), that can accurately identify whether an input face image is real or fake. Because the shortcuts can mix intuitive characteristics at the bottom layer and abstract features near the top layer, the composition difference feature module is built on a simple 18-layer ResNet. To make full use of the composition difference features, convolution replaces the convolutions of the basic ResNet-18.In addition, we change the size of the feature map by setting dilation rate to 2 in block2 to extract enough features for face forging. In the following module, the extracted features are incorporated.

We propose a multi-scale information module, which comprises of PP blocks, to extract multi-scale information after the composition difference feature module. The module's structure is based on the work of [15]. We employ global average pooling to squeeze the spatial information into channel statistics after the multi-scale information module, and then send the feature information to the fully connected layer for final classification. We can observe that, when compared to the basic ResNet-18 network, the proposed texture variation network uses a wider range of features to make judgments, including the genuine and false areas. Texture variation network has a good effect on face forgery detection.

We employ random initialization for the layers and blocks in face forgery detection, which is distinct from the specified initialization used for the dilated convolution in image segmentation [53], due to the differences between the two tasks. Adam [54] is used to optimise the networks. We employ the Cosine [55] learning rate scheduler with a base learning rate of 0.001 and a momentum of 0.9. For around 50 epochs of training, the batch size is fixed to 64. The cross-entropy function, which is commonly employed as the loss of binary classification task functions, is chosen as the loss function.

$$\mathrm{L}(\mathrm{x_t}, \mathrm{x_s}) = -(\mathrm{x_t} \times \log(\mathrm{x_s}) + (1 - \mathrm{x_t})\log(1 - \mathrm{x_s})) \tag{8}$$

The real label of x is x_t, and the probability of x_t = 1 is x_s, where label 0 and 1 are assumed.

4 Experiments and Results

Extensive experiments are carried out in this section to verify the efficacy of our proposed strategy. First, we'll go over the implementation specifics in detail as implemented. Finally, we present and discuss the findings of our experiments.

A. Experimental Setups:

The experimental conditions of our method are described in this subsection so that other researchers can replicate our findings.

I. Evaluation Metrics:

The Accuracy Score (ACC) is used as the evaluation measures. Faceforensics++ and DeeperForensics-1.0 are both evaluated using the frame-level ACC. The following is the ACC formula:

$$ACC = S_{tp}/ALL_{test} \quad (9)$$

where S_{tp} is the number of correctly classified images, and ALL_{test} is the total number of images in the test.

B. Performance Evaluation:

In this section, the performance of the proposed method is investigated and compared to that of other cutting-edge methods. We tested the proposed strategy on four datasets.

In the comparative section, we used some traditional frame-level detection approaches. Durall et al. [42] created a method for training SVM for identification that makes use of the high-frequency information difference. Rahmouni et al. [36] used various CNN designs with a global pooling layer that computes four statistics (mean, variance, maximum, and minimum). MesoNet [39] is a face forgery detection network based on CNN.

Table I: Comparative Analysis of Detection Performance

Method	Face Forencics ++ c23				Face Forencics ++ c40				Deeper Forencics 1.0		Celeb DF	DFDC
	DF	F2F	FS	NT	DF	F2F	FS	NT	Random	Std	[ACC]	[ACC]
Durall et.al	81.7	89.2	90.3	72.7	71.6	65.6	65.4	59.3	79.5	85.6	0.7846	-
Rahmouni. et.al	82.1	93.4	92.5	75.1	73.2	62.3	67.0	62.6	80.3	87.5	0.8127	-
Bondi.et.al	97.8	98.7	98.2	93.5	94.9	91.3	94.2	87.7	99.3	99.7	0.9980	0.9190
Bonettini. et.al	**98.9**	99.3	98.0	94.5	96.1	92.9	94.0	88.1	99.7	99.8	**0.9991**	**0.9226**
Proposed	98.6	**99.7**	**98.4**	**94.6**	**97.8**	**96.8**	**96.8**	**88.4**	**99.9**	**99.9**	0.9987	0.9197

XceptionNet [10] is a standard CNN based on separable convolutions with residual connections that was trained on ImageNet. To address multi-scale challenges, DSP-FWA [58] used a dual spatial pyramid technique, and the utilization of multi-scale data gives vital inspiration for future study. The gramm module was introduced into the network by Liu et al. [45], who focused on global texture.

Qian et al. [43] proposed a frequency domain approach. We do not use any one-of-a-kind training methods, such as specific loss, for the fair comparison in this paper. These algorithms take a single face image as input and do not require any other information, such as masks or other metadata. We conducted our own experiments for all methods whose source codes are available to the public.

On the Faceforensics++ database, the proposed texture variation network performed admirably. The ACC outperforms the reference approaches in c23 in several categories, including DeepFakes (DF), Face2Face (F2F), FaceSwap (FS), NeuralTextures (NT). The ACC outperforms the reference approaches in all categories in c40. The performance of the texture variation network-superior Net on low-quality data could explain this result. Furthermore, when the compression level is increased from c23 to c40, the ACC of the proposed method decreases less than that of the other reference methods.

The proposed technique produces excellent results on the other four datasets. On the DeeperForensics-1.0 database, our texture variation network achieves the best results in std/random, implying that the proposed technique is more resistant to distorted data. Our texture variation network ranks second among all methods on Celeb-DF and DFDC, indicating that our method has good generalization performance.On Celeb-DF and DFDC, texture variation network ranks second among all methods, indicating that our method has a high generalization performance in face forgery detection. Table I demonstrates that the proposed texture variation network has the best overall performance. Figure 4 and Figure 5 shows the comparison of existing methods with the proposed method in Face Forensic ++ dataset with c23 and c40. Figure 6 shows the accuracy of the proposed methods applied across four datasets.

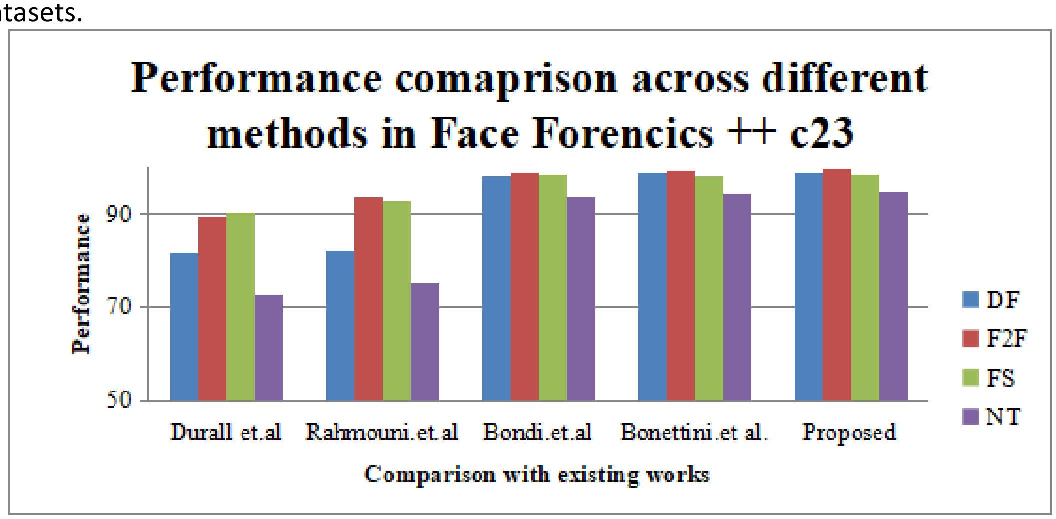


Figure 4: Comparison of existing methods with the proposed method in Face Forensic ++ dataset with c23

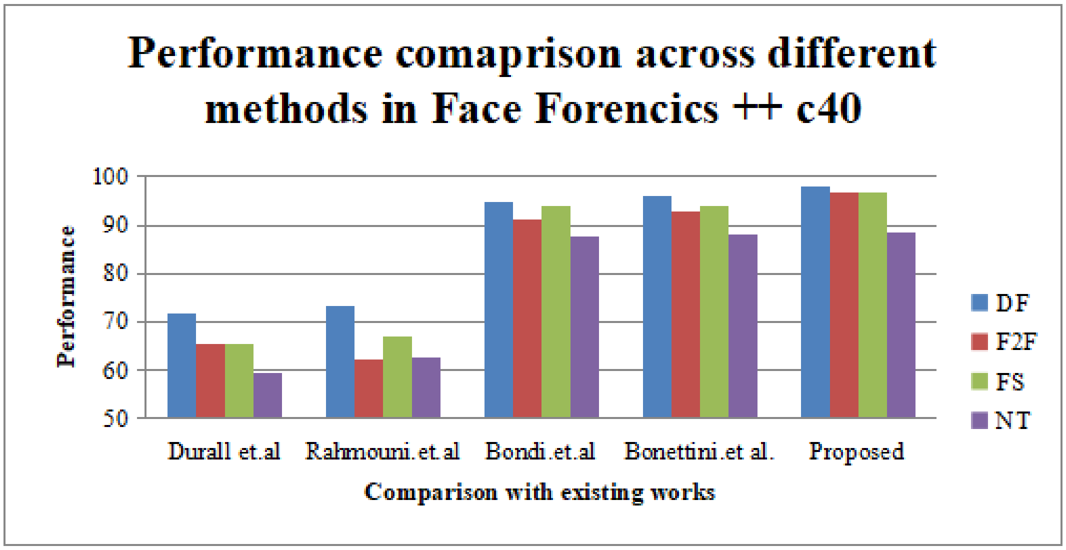


Figure 5: Comparison of existing methods with the proposed method in Face Forensic ++ dataset with c40

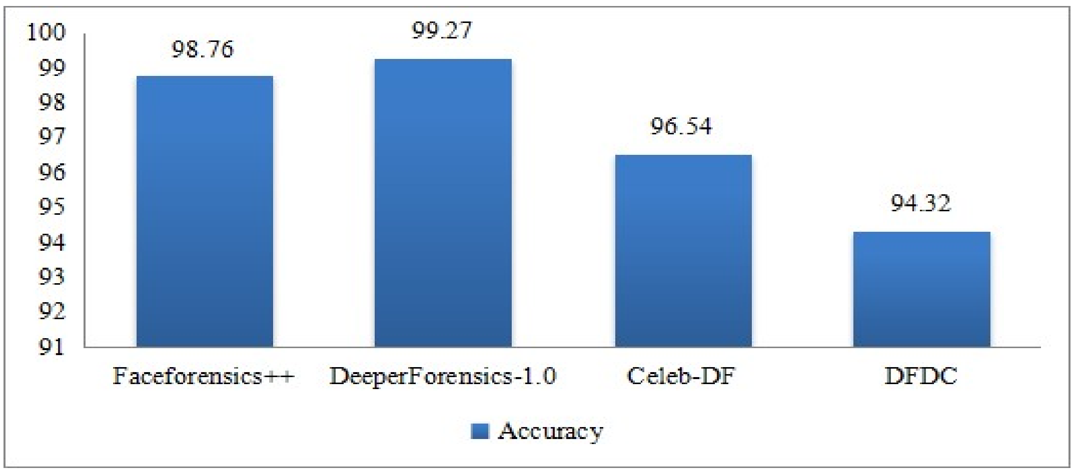


Figure 6: Accuracy of the proposed method across four different datasets.

5 Conclusion and Prospectives

We developed a method for detecting face counterfeiting based on texture variation network data. The texture discrepancies between the real and fake facial photos were discovered using the standard texture representation GLCM. Based on our findings, we attempted to introduce specific convolution operations—convolution for feature extraction and information fusion. Meanwhile, multi-scale data was combined using the advantages of PP and image-level pooling. Experiments show that our proposed method is capable of detecting fake facial images with high accuracy while minimising distortion. However, for an unknown face manipulation method, a new model must be trained. In future studies, we will commit to developing a universal technique and promoting face forgery detection.

6 References

1. Faceswap.(2019).DeepFakes.[Online]. Available:https://www.github.com/MarekKowalski/FaceSwap/
2. DeepFakes GitHub. (2019). DeepFakes. [Online]. Available: https://github.com/deepfakes/faceswap
3. J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner,“Face2Face: Real-time face capture and reenactment of RGB videos,” Commun. ACM, vol. 62, no. 1, pp. 96–104, 2018.
4. I. J. Goodfellow et al., “Generative adversarial networks,” in Proc. Adv. NeuralInf. Process. Syst., vol. 3, 2014, pp. 2672–2680.
5. T. J. de Carvalho, C. Riess, E. Angelopoulou, H. Pedrini, and A. Rocha,“Exposing digital image forgeries by illumination color classifification,” IEEE Transactions on Information Forensics and Security, vol. 8, no. 7, pp. 1182-1194, July 2013, doi: 10.1109/TIFS.2013.2265677.
6. P. Ferrara, T. Bianchi, A. D. Rosa, and A. Piva, “Image forgerylocalization via fine-grained analysis of CFA artifacts,” IEEE Trans. Inf.Forensics Security, vol. 7, no. 5, pp. 1566–1577, Oct. 2012.
7. K. Bahrami, A. C. Kot, L. Li, and H. Li, “Blurred image splicinglocalization byexposing blur type inconsistency,” IEEE Trans. Inf.Forensics Security, vol. 10, no. 5, pp. 999–1009, May 2015.
8. J. Wen, J. Yang, B. Jiang, H. Song, and H. Wang, “Big data drivenmarine environment information forecasting: A time series predictionnetwork,” IEEE Trans. Fuzzy Syst., vol. 29, no. 1, pp. 4–18, Jan. 2021.
9. D. Cozzolino, G. Poggi, and L. Verdoliva, “Recasting residual-basedlocal descriptors as convolutional neural networks: An application toimage forgery detection,” in Proc. 5th ACM Workshop Inf. HidingMultimedia Secur., Jun. 2017, pp. 159–164.
10. A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, andM. Niessner, “FaceForensics++: Learning to detect manipulated facialimages,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019,pp. 1–11.
11. F. Matern, C. Riess, and M. Stamminger, “Exploiting visual artifacts toexpose DeepFakes and face manipulations,” 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), 2019, pp. 83-92, doi: 10.1109/WACVW.2019.00020.
12. J. Yang, S. Xiao, A. Li, G. Lan, and H. Wang, “Detecting fakeimages by identifying potential composition difference,” Future Gener.Comput. Syst., vol. 125, June 2021.
13. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, “Joint face detection andalignment using multitask cascaded convolutional networks,” IEEESignal Process. Lett., vol. 23, no. 10, pp. 1499–1503, Oct. 2016.
14. Z. Yu et al., “Searching central difference convolutional networks forface anti-spoofing,” in Proc. IEEE/CVF Conf. Comput. Vis. PatternRecognit. (CVPR), Jun. 2020, pp. 5295–5305.
15. L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” Jun. 2017,arXiv:1706.05587. [Online]. Available: http://arxiv.org/abs/1706.05587
16. L. Jiang, R. Li, W. Wu, C. Qian, and C. C. Loy, “DeeperForensics-1.0:A large-scale dataset for real-world face forgery detection,” in Proc.IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020,pp. 2889–2898.

17. Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, “Celeb-DF: A large-scalechallenging dataset for DeepFake forensics,” in Proc. IEEE/CVF Conf.Comput. Vis. Pattern Recognit. (CVPR), Seattle, WA, USA, Jun. 2020,pp. 3207–3216.
18. B. Dolhansky et al., “The DeepFake detection challenge (DFDC)dataset,” Jun. 2020, arXiv:2006.07397. [Online]. Available:http://arxiv.org/abs/2006.07397
19. J. Thies, M. Zollhöfer, and M. Nießner, “Deferred neural rendering:Image synthesis using neural textures,” ACM Trans. Graph., vol. 38,no. 4, pp. 1–12, Jul. 2019.
20. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translationwith conditional adversarial networks,” in Proc. IEEE Conf. Comput. Vis.Pattern Recognit. (CVPR), Jul. 2017, pp. 1125–1134.
21. D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” Dec. 2013, arXiv:1312.6114. [Online]. Available: https://arxiv.org/abs/ 1312.6114
22. T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, “High-resolution image synthesis and semantic manipulation with conditional GANs,” inProc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 8798–8807.
23. J. -Y. Zhu, T. Park, P. Isola and A. A. Efros, "Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks," 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244.
24. J. A. Redi, W. Taktak, and J.-L. Dugelay, “Digital image forensics: A booklet for beginners,” Multimedia Tools Appl., vol. 51, no. 1, pp. 133–162, Jan. 2011.
25. J. Lukäš, J. Fridrich, and M. Goljan, “Digital camera identifification from sensor pattern noise,” IEEE Trans. Inf. Forensics Security, vol. 1, no. 2, pp. 205–214, Jun. 2006.
26. D. Cozzolino, G. Poggi, and L. Verdoliva, “Splicebuster: A new blind image splicing detector,” in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS)
27. B. Bayar and M. C. Stamm, “Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection,” IEEE Trans. Inf. Forensics Security, vol. 13, no. 11, pp. 2691–2706, Nov. 2018.
28. F. Matern, C. Riess and M. Stamminger, "Gradient-Based Illumination Description for Image Forgery Detection," IEEE Transactions on Information Forensics and Security, vol. 15, pp. 1303-1317, 2020, doi: 10.1109/TIFS.2019.2935913.
29. T. H. Thai, R. Cogranne, F. Retraint, and T.-N.-C. Doan, “JPEG quantization step estimation and its applications to digital image forensics,”
30. T. Bianchi, A. Piva, and F. Perez-Gonzalez, “Near optimal detection of quantized signals and application to JPEG forensics,” in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Nov. 2013, pp. 168–173.
31. M. Kirchner, “Linear row and column predictors for the analysis of resized images,” in Proc. 12th ACM Workshop Multimedia Secur. (MM&Sec), 2010, pp.1
32. J. Fridrich and J. Kodovský, “Rich models for steganalysis of digital images,” IEEE Trans. Inf. Forensics Security, vol. 7, no. 3, pp. 868–882, Jun. 2012.
33. M. Huh, A. Liu, A. Owens, and A. A. Efros, “Fighting fake news: Image splice detection via learned self-consistency,” in Computer Vision— ECCV, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 106–124.

34. O. Mayer and M. C. Stamm, “Learned forensic source similarity for unknown camera models,” in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP),Apr. 2018, pp. 2012–2016.
35. P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, “Learning rich features for image manipulation detection,” in Proc. IEEE/CVF Conf. Comput. Vis. PatternRecognit. (CVPR), Jun. 2018, pp. 1053–1061.
36. N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen, “Distinguishingcomputer graphics from natural images using convolution neural networks,” in Proc. Inf. Forensics Secur., 2018, pp. 1–6.
37. R. Wang et al., “FakeSpotter: A simple yet robust baseline for spotting AI-synthesized fake faces,” in Proc. 29th Int. Joint Conf. Artif. Intell. (IJCAI), C.Bessiere, Ed., Jul. 2020, pp. 3444–3451.
38. Y. Li, M.-C. Chang, and S. Lyu, “In Ictu oculi: Exposing AI created fake videos by detecting eye blinking,” in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2018, pp. 1–7.
39. D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, “MesoNet: A compact facial video forgery detection network,” in Proc. IEEE Int. Workshop Inf. Forensics Secur. (WIFS), Dec. 2018, pp. 1–7.
40. L. Li et al., “Face X-ray for more general face forgery detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 5001–5010.
41. I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, “DeepFake video detection through optical flflow based CNN,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW), Oct. 2019, pp. 1–3.
42. R. Durall, M. Keuper, F.-J. Pfreundt, and J. Keuper, “Unmasking DeepFakes with simple features,” Nov. 2019, arXiv:1911.00686. [Online]. Available: http://arxiv.org/abs/1911.00686
43. Y. Qian, G. Yin, L. Sheng, Z. Chen, and J. Shao, “Thinking in frequency: Face forgery detection by mining frequency-aware clues,” Jul. 2020, arXiv:2007.09355. [Online]. Available: http://arxiv.org/abs/2007.09355
44. R. Tolosana, S. Romero-Tapiador, J. Fierrez, and R. Vera-Rodriguez, “DeepFakes evolution: Analysis of facial regions and fake detection performance,”Apr. 2020, arXiv:2004.07532. [Online]. Available: http://arxiv.org/abs/2004.07532
45. Z. Liu, X. Qi, and P. H. S. Torr, “Global texture enhancement for fake face detection in the wild,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 8057–8066.
46. A. Kumar, A. Bhavsar, and R. Verma, “Detecting DeepFakes with metric learning,” in Proc. 8th Int. Workshop Biometrics Forensics (IWBF), Apr. 2020
47. J. Hernandez-Ortega, R. Tolosana, J. Fierrez, and A. Morales, “DeepFakesON-phys: DeepFakes detection based on heart rate estimation,” Oct. 2020, arXiv:2010.00400. [Online]. Available: http://arxiv.org/abs/2010.00400
48. U. A. Ciftci, I. Demir, and L. Yin, “FakeCatcher: Detection of synthetic portrait videos using biological signals,” IEEE Trans. Pattern Anal. Mach. Intell., early access, Jul. 15, 2020, doi: 10.1109/TPAMI.2020.3009287.
49. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual explanations from deep networks via gradient-basedlocalization,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 618–626.

50. K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016,
51. R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classifification,” IEEE Trans. Syst., Man, Cybern., vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.
52. V. Dumoulin and F. Visin, “A guide to convolution arithmetic for deep learning,” Mar. 2016, arXiv:1603.07285. [Online]. Available: http://arxiv.org/abs/1603.07285
53. F. Yu and V. Koltun, “Multi-scale context aggregation by dilated convolutions,”in Proc. ICLR, 2016, pp. 1–13.
54. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” Dec.2014, arXiv:1412.6980. [Online]. Available: https://arxiv.org/abs/1412.6980
55. I. Loshchilov and F. Hutter, “SGDR: Stochastic gradient descent with warm restarts,” in Proc. 5th Int. Conf. Learn. Represent. (ICLR), 2017, pp. 1–16.
56. K.-Y. Lin and G. Wang, “Hallucinated-IQA: No-reference image quality assessment via adversarial learning,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 732–741.
57. J. Yang et al., “No reference quality assessment for screen content images using stacked autoencoders in pictorial and textual regions,” IEEE Trans. Cybern., early access, Oct. 15, 2020, doi: 10.1109/TCYB.2020.3024627.
58. Y. Li and S. Lyu,“Exposing DeepFake videos by detecting face warpingartifacts,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)
59. S. Haseena, S. Bharathi, I. Padmapriya, and R. Lekhaa, “Deep learning based approach for gender classification,” in 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 1396–1399, Coimbatore, India, March 2018.
60. S. Haseena, S. Saroja, and T. Revathi, “A fuzzy approach for multi criteria decision making in diet plan ranking system using cuckoo optimization,” Neural Computing and Applications, vol. 34, no. 16, pp. 13625–13638, 2022.
61. H. Sikkandar and R. Thiyagarajan, “Deep learning based facial expression recognition using improved cat swarm optimization,” Journal of Ambient Intelligence and Humanized Computing, vol. 12, no. 2, pp. 3037–3053, 2021.
62. H. Sikkandar and R. Thiyagarajan, “Soft biometrics-based face image retrieval using improved grey wolf optimisation,” IET Image Processing, vol. 14, no. 3, pp. 451–461, 2020.