Intelligent Automation & Soft Computing
DOI:10.32604/iasc.2021.018854
Article

Tech Science Press

Duplicate Frame Video Forgery Detection Using Siamese-based RNN

Maryam Munawar and Iram Noreen*

Department of Computer Science, Bahria University, Islamabad, Lahore Campus, 54782, Pakistan
*Corresponding Author: Iram Noreen. Email: iram.bulc@bahria.edu.pk
Received: 22 March 2021; Accepted: 23 April 2021

Abstract: Video and image data is the most important and widely used format of communication today. It is used as evidence and authenticated proof in different domains such as law enforcement, forensic studies, journalism, and others. With the increase of video applications and data, the problem of forgery in video and images has also originated. Although a lot of work has been done on image forgery, video forensic is still a challenging area. Videos are manipulated in many ways. Frame insertion, deletion, and frame duplication are a few of the major challenges. Moreover, in the perspective of duplicated frames, frame rate variation and loop detection are also key issues. Identification of forged duplication frames for large videos with variant frame rates in real-time is not applicable due to computational limitations, lack of generalization, and low-performance accuracy. This research has investigated the problem of frame duplication with varied frame rates using a deep learning approach. A novel deep learning framework consisting of Inflated 3D (I3D) and Siamese-based Recurrent Neural Network (RNN) is proposed to resolve the aforementioned issues. The first step in the proposed framework is to extract the features and convert videos into frames. I3D network receives an original and a forged video to detect frame-to-frame duplication. Then multiple frames are merged to create a sequence. This sequence is passed to Siamese-based RNN which is used for the sequence to sequence forgery detection in video. Media Forensic Challenge (MFC) is a relatively new dataset with various frame rates, and a huge volume of videos. MFC and Image Retrieval and Analysis Tool (VIRAT) datasets are used for training and validation of the proposed model. The accuracy of the proposed method with the VIRAT dataset is 86.6% and with the MFC dataset 93%. The comparative analysis with state-of-the-art approaches has shown the robustness of the proposed approach.

Keywords: Duplicate frame; video forgery; deep learning; RNN; Siamese; I3D

1 Introduction

The rapid advent of computationally cheap and cross-platform video editing software has enabled the huge volume of video content available to a large number of users via the Internet [1]. In recent years, the abundance of video data, AI techniques, and readily available, high-performance easy to use video editing tools have given rise to fake videos. Fraudulent activities are carried out using fake images and videos to bypass facial

authentication, to publish fake news, and for entertainment as well. Video forgery is continuously increasing in the digital world due to breaches of information security, consequently establishing a scenario for image and video content monitoring for forgery identification [2–4]. The spread of fake videos raises security risks and anarchy in society. Video forgery detection has applications in media science, forensic analysis, digital investigations, and authenticity verification of a video. The purpose of video forensic technology is to extract features to distinguish fake content frames from original videos [5].

Video forgery detection is mainly categorized as inter-frame forgery and copy-move forgery Copying and pasting content within the same frame is referred to as copy-move forgery [6–8]. Copying and pasting content within the same frame is referred to as copy-move forgery. Whereas Inter frame refers to frame deletion and frame duplication. The frame duplication is examined in three different ways, duplication of frame sequences in successive video frames, duplication of same and different dimensions from other videos, and duplication of different lengths at many locations using correlation coefficient and coefficient of variation. Machine learning have been used for video forger detection using a number of approaches. Initially, most of the work for image forgery detection has been done using Support Vector Machine (SVM). Later, neural network-based techniques also addressed the problem of video forgery for frame duplication detection [1].

Recently, Deep learning methods and specifically Convolution Neural Networks (CNNs) have gained tremendous success due to its powerful ability of automatic learning of features for large-scale video classification [7,9,10]. Copy move forgery problem is investigated a lot, however, inter-frame duplication is not explored much and still is not applicable in real-time due to computational limitations and robustness issues for real-time scenarios. Existing approaches in the literature suffer from low accuracy rates, low efficiency, and high computational complexity. Further, most of the existing approaches are trained on datasets with a limited sample size; which is not enough to unleash the potential of deep learning performance. Moreover, most of the existing work does not address variable frame rates for sample videos. Which is a challenging issue as multiple platforms with different hardware and software sources exist today. The purpose of this study is to investigate deep learning method for video forgery detection on large datasets with varied frame rates. A deep neural network approach is presented to classify the forged videos by finding duplicated frames in a video. The silent features of the proposed approach are as following.

- The proposed deep learning approach has integrated the functionalities of two neural network-based methods, i.e., I3D and Siamese-based Recurrent Neural Network (RNN). Off-the self, fine-tuned I3D model integration has improved the speed of the process to handle huge data size. Siamese-based RNN provides a robust forgery detection mechanism.
- The proposed approach is trained and tested on the most cited VIRAT [11] dataset and newly published MFC dataset [12]. MFC dataset is a recently published benchmark video dataset and is more challenging, rich, and offers varied frame rates than VIRAT. Very limited work is reported to explore its potential for forgery detection.
- The proposed approach has shown improved performance on both VIRAT and MFC datasets than existing approaches.

The rest of the paper is organized as follows. Section 2 describes the prominent related research in the domain. Section 3 presents the dataset description and Section 4 presents proposed approach. Section 5 presents results discussion and state of art comparison followed by conclusion in Section 6.

2 Related Work

In literature, a number of machine learning approaches have been introduced for video forgery detection. Saddique et al. [13] proposed a forgery detection technique using localized texture analysis in consecutive

frames. Christlein et al. [4] evaluated the performance of features for copy-move forgery detection using a number of manual feature extraction techniques like Speeded up Robust Features (SURF) and Scale Invariant Feature Transform (SHIFT), the block-based features Kernel Principal Component Analysis (KPCA), Principal Component Analysis (PCA), Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT).

Amiano et al. [14] proposed a patch-match based copy-move detection method using the near neighbor dense field concept. However, this approach did not perform well for huge data. Wu et al. [15] proposed an approach for frame duplication and frame deletion detection using velocity and discontinuity peaks observations in vector flow image sequence. Bidokhti [16] presented a video forgery copy-move method of detection in MPEG (Moving Picture Expert Group) videos. Their method divides every video frame into suspicious cleared portions to calculate optical flow coefficient for every part. It further detects forgery when an unfamiliar trend in the optical flow coefficient object is detected. Singh et al. [17] presented a passive blind scheme to detect duplicated frames using correlation coefficient and coefficient of variation.

Afchar et al. [18] presented two network architectures to identify forgeries efficiently with low computational cost using a deep learning approach. Their approach focuses on Deep fake and Face2Face the two approaches are used to generate hyper-realistic forged videos. Nguyen et al. [19] introduced a capsule network to do replay attack forgery detection, face-swapping detection, and facial reenactment detection on recorded videos and published images. Hosler et al. [20] presented a new standard database, video ACID (Atomicity, Consistency, Isolation, Durability), designed to study multimedia forensics on videos. Ulutas et al. [21] proposed a Bag-of-Words (BoW) model for frame duplication detection method. The proposed method compares Peak Signal-To-Noise Ratio (PSNR) values of frames and feature vectors to detect forgery. Chen et al. [8] proposed copy-move forgery regions to detect invariant moments design using Scale-Invariant Feature Transform (SIFT) and region growing strategy. Pavlovi [6] presented a novel method for CMFD (Copy-Move Forgery Detection), based on a multi-fractal spectrum. Liu et al. [22] presented a K-mean clustering and super-pixel segmentation technique for image forgery detection. Dong-Ning Zhao et al. [23] presented a passive blind scheme based on resemblance analysis to detect an inter-frame forgery in shots using histogram variation in two adjacent frames, and SURF feature extraction. M. Aloraini et al. [24] explored the issue of video forgery for object removal and introduced a new approach based on sequential and patch analyses to localize forged regions and detect video forgery by visualizing a removed object movement. They modeled video sequences as normal and anomalous patches to define the distribution of each patch on the video level. Nguyen et al. [25] presented a method of detecting inter-frame forgeries using Convolution Neural Network (CNN) by retraining the available ImageNet dataset-trained CNN models to detect inter-frame forgeries, and to exploit spatial-temporal relationships in a video. Their CNN model showed 81% accuracy.

Due to large-scale datasets availability recently, the potential of deep learning is being explored in this domain. Recently, Bakas et al. [1] presented a deep learning approach for detection of inter-frame forgeries in videos based on motion vector analysis. However, they have also not addressed the issue of variant frame rate. Varied frame rate is a challenging issue as multiple platforms with different hardware and software sources exist today. Similarly, the behavior of duplicate forgery detection for variable frame rate videos such as available in MFC dataset is not studied much. Similarly, in comparison to copy-move forgery, limited work exists for inter-frame forgery. Carreira et al. [2] introduced a Long-Short-Term-Memory (LSTM) with two-stream 3D-ConvNet to check video forgery producing 80% accuracy. Jia et al. [3] defined the coarse-to-fine technique based on video parameters and Over Flow (OF) features of frame to detect copy-move forgery with high accuracy and low computation cost under different common attacks. Avino et al. [26] used Auto Encoder integrated with RNN for forgery detection. Recently, Long et al. [10] extracted compact features from digital videos containing both temporal and spatial information of

video segmentation using I3D and ResNet with 84.05% accuracy on the VIRAT dataset and 85.88% accuracy on the MFC dataset.

3 Dataset Description

Two benchmark datasets are applied in this study. First one is Video Image Retrieval and Analysis Tool (VIRAT) action recognition dataset [11,27] which offers 29 hours of videos describing 23 different events. As, this dataset is not for frame duplication identification, therefore, in pre-processing stage different steps are applied to prepare it for frame duplication identification. Since our objective is to find duplicate frame sequences, therefore, annotations available for VIRAT are ignored. After preprocessing video clips, they will have duplicated frame sequences. Another dataset is Media Forensic Challenge (MFC) [12] which is a recent benchmark dataset for video forgery evaluation. It offers 300,000 video clips and 11,000 high-provenance (HP) videos with 4,000 manipulated videos and over 500 video manipulation journals with historical statistics and annotation information for manipulation. MFC dataset offers varied frame rates in data samples. Detail of both datasets is described in Tab. 1.

Table 1: Dataset description

Dataset	Year Published	Total Video frame samples	Frame rate	Format
MFC [12]	2019	3 million frames (100 videos)	10 fps, 29–30 fps, 60 fps, 240 fps	MP4, MOV
VIRAT [11]	2016	8000 frames (60 videos)	23 fps	MP4

4 Proposed Framework

Deep learning has gained immense popularity due to its potential of high performance with large-scale data. The internet age has has provided abundant video data with easy access, hence a rise in the forgery and fake video content as well. The proposed framework is a deep learning approach based on Inflated 3D (I3D) network and Siamese-based Recurrent Neural Network (SRNN) for frame duplication. Elements of proposed approach are described as following.

4.1 Preprocessing

Pre-processing steps are crucial before the model's training such as analyzing and normalization of the dataset. Following pre-processing activities are performed on both datasets.

4.1.1 Conversion of Video to Frames

Video clips are transformed to frame sequences by a pre-processing script written using opencv library, which captures a frame and saves it in a directory. This step is executed in an iterative process and is continued till all frames are extracted from video clips. Similarly, all video clips in the dataset are processed.

4.1.2 Generation of Frames Duplication

Second step is to generate duplicate frames in original sequence. 16 frame sequences are created from original video, and in each sequence after first 8 frames initial 4 frames are repeated and then next 4 frames after 8th frame are added which form a total of 16 frames sequence. Same method is used here for each clip and 16 frame sequences are produced for whole video clip. Therefore, in each sequence frames 1 to 4 are duplicated at position 9 to 12. For each clip these sequences are stored separately in a new directory.

Additionally, each frame size is changed to 224 × 224 × 3 which is default input size for I3d network. A total of 64 sequences are generated for all videos.

4.1.3 Feature Extraction

64 frame rate is selected for sequence to sequence detection. 64 frame sequences from videos are generated and then forged sequence from original sequence is generated.

4.2 Siamese-based RNN Integrated with ID3

A Siamese network consists of twin neural networks which accept separate inputs but are connected at the top by an energy function. This function calculates a metric between the representations of the highest level feature on each side. The parameters are tangled up between the twin networks [26]. The network is symmetric, such that the top conjoining layer will measure the same metric if we present two distinct images to the twin networks, as if we were to present the same two images, but to the opposite twins. Siamese network help us to follow the deep learning approach that mostly follow the sequence to sequence detection method. Sequence method means a set of frames combine in a group that is called a sequence.

Recurrent Neural Networks (RNN) are a class of neural networks for processing sequential data. It recalls previous data from the input series for forecast using feedback loops. Long Short Term Memory (LSTM) is a type of RNN used in deep learning domain [10]. It includes a special cell called 'memory cell' to maintain information for longer periods of time.

Proposed framework comprises of I3D (Inflated 3 Dimension) model followed by Siamese model. Two 16 frame sequences are created, one is forged as described in pre-processing steps, and the other is non-forged. Both of these sequences are passed through I3D model and 7,1,1,1024 feature tensor is obtained from it for both sequences. These features are saved in ".npy" format which are then used as inputs for fine classification by Siamese RNN network. Thus, videos are converted into frames by I3D model to create a 64 frames sequence, i.e., multiple of 16 frames. After forming 64 frames sequences, one original input sequence, i.e., without frame duplication and the forged input sequence having 16 frames duplicated, both are fed to I3D network to obtain 1024 vector features from its off the shelf model (last layers which classify input are removed). Positive and negative pairs are formed from generated sequences using a pair function. A positive pair is, if both input sequences are of same class which is either forged or non-forged, and a negative pair is when one input sequence in the pair is forged and the other is non-forged. After forming pairs, dataset is split into train and test set. Additionally, shape of feature tensor is changed to 2, 192, 98, 98 where 2 represent pair (frame height and width is reduced to 98 from 224 to speed up processing and avoid memory shortage).

Paired data is fed to Siamese RNN model for fine classification or for verification that input frames are forged or not. Siamese-based RNN network is used for sequence to sequence forgery detection. It consists of two similar networks which share exactly the same parameters, each receives one of the two input frames. Siamese architecture of the neural network learns to distinguish between two frames in the given pair. Siamese network verifies that video frames are forged or not by finding similarity score between frames sequence. If similarity score is high then it means that frames are forged and duplicated. In the Siamese network 2 convolution layers along with 3 recurrent neural network layers are used which are followed by dense/hidden layers. A vector is produced by last layer in base model for each input sequence and similarity score is measured through Euclidean distance. If similarity score is below 0.5 then video frames are predicted as forged or duplicated. It performs fine level search in order to find the duplication between frames. After getting this feature vector for each input sequence, cosine distance or L2 distance is used to measure distance between frames. If distance is less than threshold then it implies the sequence has duplicated frames. The overall methodology and detailed architecture of Siamese-based RNN is demonstrated in Fig. 1.

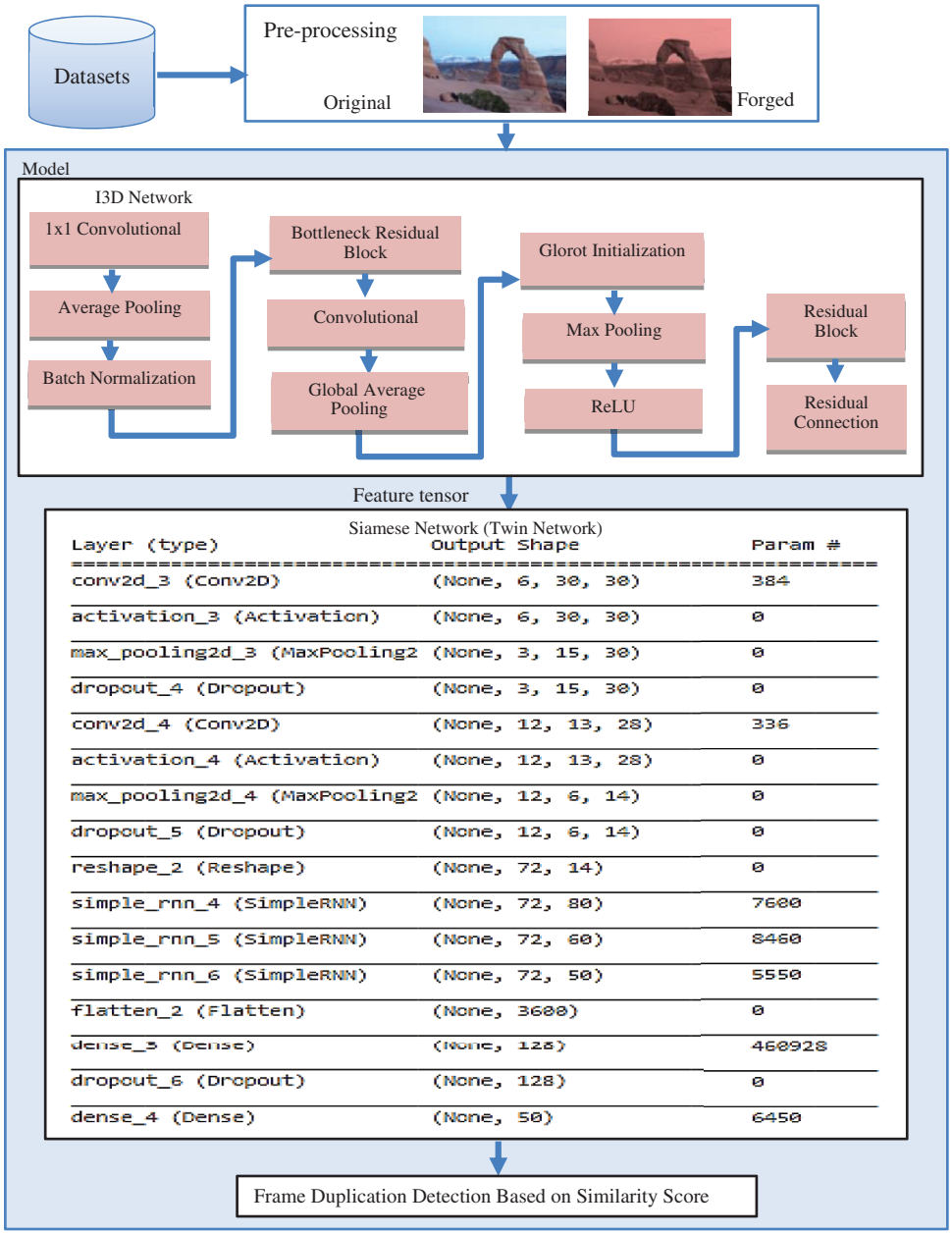

Figure 1: Proposed approach

4.3 Implementation and Setup

The combination of I3D (Inflated 3 Dimension) and Siamese-based RNN is used to build proposed solution using CV2, TensorFlow [28] and Keras libraries [29] for different phases of model construction such as Layers, Activation, Dense, Input, Batch Normalization, Cov3D, Maxpooling3D, Averagepooling3D, Dropout, Reshape, Lambda and Global Averagepooing3D. Google COLAB Python3 is used as development environment to build, train and test the proposed model. VIRAT and MFC both datasets are used for training of proposed architecture. During training, Leave-One-Out (LOO) cross validation is used to avoid overfitting and obtain performance. It is a special case of K-fold cross validation when number of logical limit K is equal to N, i.e., number of folds equal to number of instances in data set. So the learning

algorithm is applied once for each instance and uses all other instances as the training set and the selected instance as a single item test set. The model is trained using 64 batch size and 0.01 learning rate. ADAM optimizer is applied. A dropout of 25% was used for Siamese layers and a dropout of 20% was used in RNN layers. The network parameters details and hyper-parameters finalized after fine tuning are described in Tab. 2.

Table 2: Network parameters

Network Parameters	Values	Network Parameters	Values
I3D Total Parameter	12,308,882	Batch size	64
I3D Trainable Parameters	12,294,332	Learning Rate	0.01
Siamese Total Parameter	489,708	Dropout rate	Siamese layer 25% RNN layer 20%,
Siamese Trainable Parameters	489,708	Optimizer	ADAM
Epochs	6	Same padding	Stride size 1

5 Results Discussion

The Accuracy score is not the true representative of a model's performance. Therefore, a number of metrics are used to evaluate the performance such Precision, Recall and F1 score. Moreover, area under the ROC curve known as AUC-Score is also calculated to measure the performance. The ROC curve is a trade-off between sensitivity and precision and tells us about the true-positive rate and false-positive rate. If it is closest to the diagonal, then the curve is not fine. A better output is indicated by classifiers that offer curves closer to the top-left corner. The closest the curve gets to the space of the ROC 45-degree diagonal, the less precise the model is. All performance metrics are presented in Eqs. (1)–(4).

$$Recall = TP = TP / (TP + FN) \tag{1}$$

$$Precision = TP / (TP + FP) \tag{2}$$

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \tag{3}$$

$$F1 - Score = 2 * (Precision * Recall) / (Precision + Recall) \tag{4}$$

The confusion matrix of the proposed approach is presented in Fig. 2 for both the VIRAT and the MFC datasets. Confusion matrix of proposed methodology with VIRAT dataset shows that the true-negative is 3362 which is the percentage of actual negative which are correctly identified, true-positive is 2809 which is the percentage of actual positive which are correctly identified, false-positive is 755 which incorrectly predicts the positive class and false-negative is 202 which incorrectly predicts the negative class. As shown by confusion matrix the true-negative is 0.81 which is the percentage of actual negative which are correctly identified, true-positive is 0.93 which is the percentage of actual positive which are correctly identified, false-positive is 0.18 percent which is incorrectly predict the positive-false and false-negative is 0.06 percent which is incorrectly predict the negative class.

The Confusion matrix of the proposed approach with MFC dataset presents the true negative as 0.92 which is the percentage of actual negative which are correctly identified, true-positive as 0.94 which is the percentage of actual positive which are correctly identified. False positive is 0.07 which incorrectly predicts the positive class and false-negative is 0.05 which incorrectly predicts the negative class. The confusion matrix of the proposed approach with MFC dataset presents the true negative as 3710 which is the percentage of actual negative which are correctly identified, and true positive as 3595 which is the

percentage of actual positive which are correctly identified. False-positive is 321 which incorrectly predicts the positive class and false-negative is 206 which incorrectly predicts the negative class.

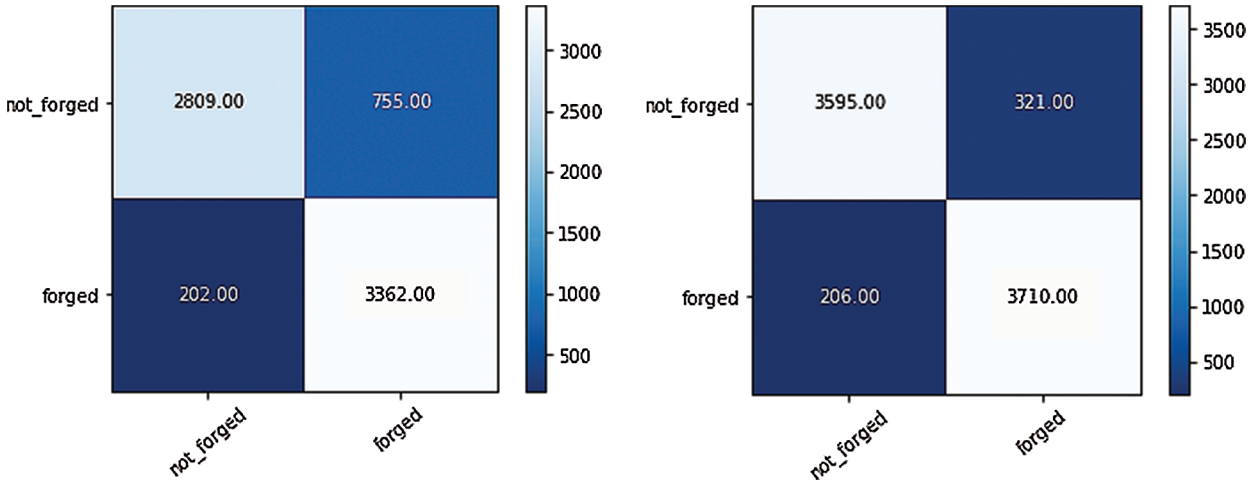

Figure 2: Confusion matrix of proposed approach with VIRAT dataset (left), MFC dataset (right)

ROC curves of the proposed approach with VIRAT and MFC datasets are presented in Figs. 3 and 4, respectively. The AUC score on VIRAT dataset for *'not forged'* data is 88.3% and it shows true positive and false positive rate of proposed methodology. Which implies that there is 88.3% chance that the model will be able to distinguish between negative class and positive class. The AUC score for *'forged'* data is 84.8% and it shows true positive and false positive rate of the proposed methodology. Which implies that there is 84.8% chance that the model will be able to distinguish between negative class and positive class. AUC score of the proposed methodology on MFC dataset for *'not forged'* data is 96% and it shows true-positive and false positive rate of proposed methodology. Which implies that there is 96% chance that the model will be able to distinguish between negative class and positive class. The AUC Score for *'forged'* data is 90% and it shows true-positive and false-positive rate of the proposed approach. Which implies that there is 90% chance that the model will be able to distinguish between negative class and positive class.

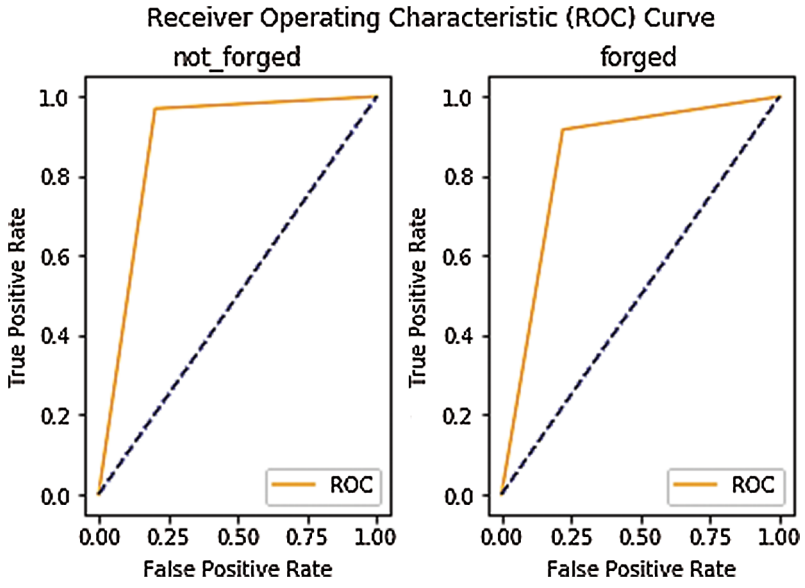

Figure 3: ROC curve of the proposed approach with VIRAT dataset

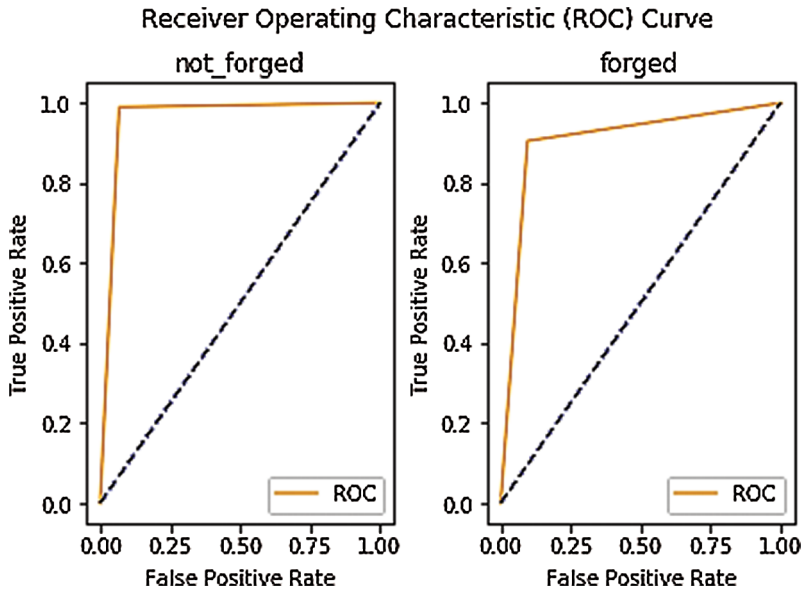

Figure 4: ROC curve of the proposed approach with MFC dataset

Tab. 3 presents the summary of evaluation scores of Siamese network and proposed Siamese-based RNN+I3D network for both datasets as per different performance metric. Siamese network accuracy with VIRAT dataset is 83.6%, precision is 85.1%, recall is 83.6% F1 score is 83.4% and MCC is 68.7%. Siamese network accuracy with MFC dataset that is 89.5%, precision is 90.7%, recall is 89.5% F1.score is 89.5% and MCC is 80.3%. The proposed methodology accuracy with MFC dataset that is 93.3%, precision is 93.3%, recall is 93.3% F1 score is 93.3% and MCC is 86.6%. ROC curve of the proposed approach for both datasets is above the threshold diagonal plot which shows the robustness and effective training of the proposed approach.

Table 3: Performance metric detail of the proposed approach

Parameters	VIRAT Dataset	MFC Dataset
Accuracy	86.6%	93.3%
Precision	87.5%	93.3%
Recall	86.6%	93.3%
F1 Score	86.5%	93.3%
MCC	71.7%	86.65%
AUC Score (Forged)	85%	91%
AUC Score (Not Forged)	88%	97%

Tab. 4 describes the state of the art comparison with the proposed approach. In previous approaches Siamese network is performed without linking with any other neural network. However, Long et al. [17] presented Siamese-based ResNet with 84.05% accuracy in VIRAT dataset and 85.88% accuracy with MFC dataset. Jabeen et al. [17] provided 81% accuracy using Inception V3. Instead of using Siamese-based ResNet, we have applied Siamese-based RNN and improved the accuracy and F1 score by on both datasets VIRAT and MFC.

Table 4: Performance comparison with state-of-the-art

Methods (MFC Dataset)	Accuracy	Methods (VIRAT dataset)	Accuracy
C. Long et al. [10], 2019	85.88%	C. Long et al. [17] 2019	84.05%
Jabeen et al. [9], 2020	81%	Bhattacharya et al. [30]	78%
Proposed Methodology	**93%**	**Proposed Methodology**	**86.6%**

6 Conclusion

In this study, a deep learning approach is presented to detect duplicate frame video forgery for varied frame rate with improved performance. The proposed approach is a Siamese (twin) RNN integrated with I3D and identifies frame duplication from the group of forged sequence frames. The proposed approach is tested using MFC and VIRAT datasets. The proposed approach demonstrated 93.3% F1 score on MFC and 86.5% on VIRAT dataset. The results are comparable with state of art approaches. Future work recommendation is to adapt transfer learning in proposed approach for enhanced and automated feature extraction and fast training process. In this context, behavior investigation of many pre-trained CNN models is required to acquire high performance and robustness for real-time applications.

Funding Statement: University research funding will be available for publishing of research article for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] J. Bakas, R. Naskar and S. Bakshi, "Detection and localization of inter-frame forgeries in videos based on macroblock variation and motion vector analysis," *Computers & Electrical Engineering*, vol. 89, pp. 106929, 2021.

[2] A. J. Carreira and Zisserman, "A new model and the kinetics dataset," in *Conf. on Computer Vision and Pattern Recognition*, pp. 6299–6308, 2017.

[3] S. Jia, Z. Xu, H. Wang, C. Feng and T. Wang, "Coarse-to-fine copy-move forgery detection for video forensics," *IEEE Access*, vol. 6, pp. 25323–25335, 2018.

[4] V. Christlein, C. Riess, J. Jordan, C. Riess and E. Angelopoulou, "An evaluation of popular copy-move forgery detection approaches," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1841–1854, 2012.

[5] D. Cozzolino, G. Poggi and L. Verdoliva, "Extracting camera-based fingerprints for video forensics," in *Conf. on Computer Vision and Pattern Recognition*, pp. 130–137, 2019.

[6] A. Pavlovi, "Copy-move forgery detection based on multifractals," *Multimedia Tools and Applications*, vol. 78, no. 15, pp. 20655–20678, 2019.

[7] J. Bakas, A. K. Bashaboina and R. Naskar, "MPEG double compression based intra-frame video forgery detection using CNN," in *Int. Conf. on Information Technology*, pp. 221–226, 2018.

[8] C. Chen, W. Lu and C. Chou, "Rotational copy-move forgery detection using SIFT and region growing strategies," *Multimedia Tools and Applications*, vol. 78, no. 13, pp. 18293–18308, 2019.

[9] S. Jabeen, U. G. Khan, R. Iqbal, M. Mukherje *et al.,* "A deep multimodal system for provenance filtering with universal forgery detection and localization," *Multimedia Tools and Applications*, vol. 11, no. 6, pp. 6165, 2020.

[10] C. Long, A. Basharat and A. Hoogs, "A coarse-to-fine deep convolutional neural network framework for frame duplication detection and localization in forged videos," in *CVPR Workshops*, pp. 1–10, 2019.

[11] S. Oh, A. Hoogs, A. Perera, N. Cuntoor, C. Chen *et al.,* "A large-scale benchmark dataset for event recognition in surveillance video," in *CVPR*, pp. 3153–3160, 2011.

[12] J. F. H. Guan, M. Kozak, E. Robertsion, Y. Lee, A. N. Yates *et al., MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation*. IEEE Winter Applications of Computer Vision Workshops, pp. 63–72, 2019.

[13] M. Saddique, K. Asghar, U. I. Bajwa, M. Hussain and Z. Habib, "Spatial video forgery detection and localization using texture analysis of consecutive frames," *Advances in Electrical and Computer Engineering*, vol. 19, no. 3, pp. 97–108, 2019.

[14] L. D'Amiano, D. Cozzolino, G. Poggi and L. Verdoliva, "A patchmatch-based dense-field algorithm for video copy-move detection and localization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 3, pp. 669–682, 2019.

[15] Y. Wu, X. Jiang, T. Sun and W. Wang, "Exposing video inter-frame forgery based on velocity field consistency," in *IEEE Int.Conf. on Acoustics, Speech and Signal Processing*, pp. 2674–2678, 2014.

[16] A. Bidokhti, "Detection of regional copy/move forgery in MPEG videos using optical flow," in *The Int. Sym. on Artificial Intelligence and Signal Processing*, pp. 13–17, 2015.

[17] G. Singh and K. Singh, "Video frame and region duplication forgery detection based on correlation coefficient and coefficient of variation," *Multimedia Tools and Applications*, vol. 78, no. 9, pp. 11527–11562, 2019.

[18] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, "MesoNet: A compact facial video forgery detection network," in *IEEE International Workshop on Information Forensics and Security*, pp. 1–7, 2018.

[19] H. H. Nguyen, J. Yamagishi and I. Echizen, "Capsule-Forensics: Using capsule networks to detect forged images and videos," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 2307–2311, 2019.

[20] B. C. Hosler, X. Zhao, O. Mayer, C. Chen, J. A. Shackleford *et al.,* "The video authentication and camera identification database: A new database for video forensics," *IEEE Access*, vol. 7, no. 1, pp. 76937–76948, 2019.

[21] G. Ulutas, B. Ustubioglu, M. Ulutas and V. V. Nabiyev, "Frame duplication detection based on BoW model," *Multimedia Systems*, vol. 24, no. 5, pp. 549–567, 2018.

[22] Y. Liu, H. Wang, Y. Chen, H. Wu and H. Wang, "A passive forensic scheme for copy-move forgery based on superpixel segmentation and K-means clustering," *Multimedia Tools and Applications*, vol. 79, no. 1, pp. 1–24, 2019.

[23] D. N. Zhao, R. K. Wang and Z. M. Lu, "Inter-frame passive-blind forgery detection for video shot based on similarity analysis," *Multimedia Tools and Applications*, vol. 77, no. 19, pp. 25389–25408, 2018.

[24] M. Aloraini, M. Sharifzadeh and D. Schonfeld, "Sequential and patch analyses for object removal video forgery detection and localization," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 3, pp. 917–930, 2021.

[25] X. H. Nguyen, Y. Hu, M. A. Amin and K. G. Hayat, "Detecting video inter-frame forgeries based on convolutional neural network model," *International Journal of Image, Graphics and Signal Processing*, vol. 12, no. 3, pp. 1–12, 2020.

[26] D. D'Avino, D. Cozzolino, G. Poggi and L. Verdoliva, "Autoencoder with recurrent neural networks for video forgery detection," *Electronic Imaging*, vol. 2017, no. 7, pp. 92–99, 2017.

[27] Q. Ji, X. Wang, L. Davis, H. Lee, J. K. Aggarwal *et al.,* "AVSS 2011 demo session: A large-scale benchmark dataset for event recognition in surveillance video," in *8th IEEE Int. Conf. on Advanced Video and Signal Based Surveillance*, pp. 527–528, 2011.

[28] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis *et al.,* "Tensorflow: A system for large-scale machine learning," in *12th {USENIX} sym. on operating systems design and implementation ({OSDI} 16)*, 2016.

[29] N. Ketkar, "Introduction to keras," *Deep Learning with Python*, pp. 97–111, 2017.

[30] S. Bhattacharya, R. Sukthankar, R. Jin and M. Shah, "A probabilistic representation for efficient large scale visual recognition tasks," Conference on Computer Vision and Pattern Recognition, pp. 2593–2600, 2011.