IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 7, July 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.542

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907056 |

Implementation of a Discriminator to Detect Fake Facial Images

B. Avinash[1], A. Prasanna Sai[2], B.Y Amrutha Valli[3], G. Yasaswini[4], J. Prathyusha[5]

Assistant Professor, Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India[1]

B. Tech Student, Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Andhra Pradesh, India[2,3,4,5]

ABSTRACT: The rapid growth of digital image processing technologies and editing software has given rise to the creation of large amounts of tampered images and circulating them in our daily lives by invading privacy of people. This undermines credibility and trustworthiness of real images and also creates false beliefs in many real-world situations. Hence it is generating a great demand for automatic forgery detection algorithms in order to determine the authenticity of an image. Many techniques and tools have been implemented to detect such type of forgery with the real image but because of increasing editing software every day, the problem is not solved yet. In this project, we have first generated realistic fake images using Generative Adversarial Network (GAN). GAN can be used to generate tampered images of specific people that may affect personal safety. And then we have developed a Deep Forgery Discriminator (DeepFD) by introducing contrastive loss and triplet loss to efficiently and effectively detect the fake and real computer-generated image.The discriminator successfully detected 94% fake images generated by DCGAN.

KEYWORDS: Digital Image Processing, Generative Adversarial Network, Deep Forgery Discriminator, Forgery detection, tampered images, contrastive loss.

I INTRODUCTION

With the advancement in technology, it is now possible to create representations of human faces in a seamless mannerWith the advancement in technology, it is now possible to create representations of human faces in a seamless manner for fake media using different kinds of editing software's and image processing technologies by invading privacy of people. Along with that, the rapid growth of deep learning techniques for image generation and processing, made the image synthesis/generation much easy, the improper and malicious use of such techniques and technologies brings hidden concerns to people's security. For example, the Generative Adversarial Network (GAN) can be used to generate the realistic image, can also be used to generate a tampered video for specific people and inappropriate events, create images that are harmful to a particular person, and may even affect their personal safety. This kind of misuses of technology will cause serious problems on the society, political, commercial and people activities. The Forged images are extensively used within the media, either deliberately or accidentally. Detecting manipulation and forgery within these images is therefore of the utmost importance.Traditional works on image forgery detection are mostly based on extracting simple features that are specific for detecting some particular type of forgery. Recently, works on forgery detection based on neural networks have proved to be very efficient in detecting image forgery. Neural networks are capableof extracting complex hidden features of an image, thus givingbetter accuracy. Contrary to the traditional methods of forgery detection, a deep learning model automatically builds the required features, hence it has become the new area of research in image forgery.

Image forgery is a term that refers to manipulating or tampering the original image to hide some useful information or to showcase some false information. The purpose behind creating forged images could range from earning money, spreading rumors, or making false claims in one's favor.

Due to the widespread use of images and serious consequences of forged images, many types of research have been done to detect manipulated images. Active approaches for forgery detection which include digital signatures and watermarking suffer from the drawback of inserting the watermark or the signature beforehand in the image which limits the scenarios where this technique would work. Various approaches have been proposed for passive based forgery detection. Popescu Farid [1] described a method for detecting duplicate regions present in an image by applying PCA to blocks of images and then sorting the blocks lexicographically. Amerini et al, [2] has used a SIFT based approach that solves the two-fold purpose of detecting the copy move forgery and to fetch the geometric transformation applied to build the forged image. W ei et al. [3] proposed a rotation angle estimation method that can detect splicing

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907056 |

forgery. The algorithm is capable of blind detection of geometrical operations performed and to detect foreign areas present in the forged image. Keet al.[4] used a technique to detect forged images by determining the consistency of shadow. It assumes the tampered image has an inconsistent shadow due to splicing. However, traditional forgery detection techniques are hard to detect the generated images by GANs since their image content are made by deep neural network directly. Therefore, it does not exist any unusually statistical property in the intrinsic features of the received images, leading to traditional forgery detection approach fails to detect the generated images. To solve this shortcoming, we propose a deep neural network called deep forgery discriminator (DeepFD) based strategy to effectively and efficiently detect the generated / fake images synthesized by GANs or other advanced networks.

The training of a deep neural network classifier is easier by collecting a huge amount of fake and real images and train the classifier to differentiate the fake and real images. However, the trained classifier may not distinguish the fake and real images, if the new images provided to the classifier are synthesized from any of the GANs. In order to make that classifier to give accurate results we need to re-train the classifier with the new GAN synthesized images. So, to avoid such overwork we have introduced contrastive loss to into the networking framework [5]. Contrastive loss learns such joint features from heterogeneous training images by introducing the pairwise information so that the DeepFD should be able to effectively distinguish any fake image generated by any GAN.

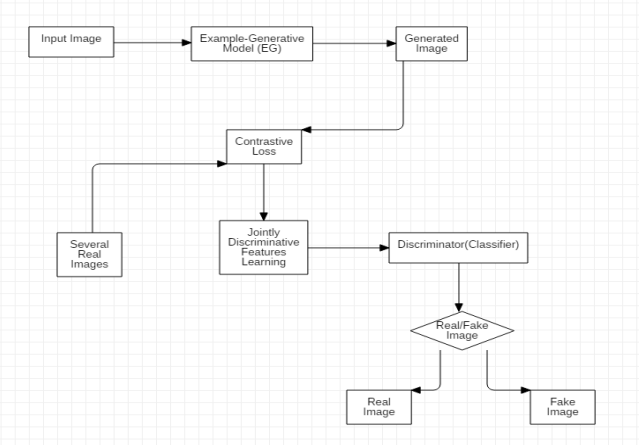

Fig 1 Architectural design of Proposed Discriminator

II PROPOSED DEEP FORGERY DISCRIMINATOR

The proposed deep forgery discriminator localizes unrealistic details of the fake images based on fully convolutional architecture. Figure 1 represents the system architecture which shows the discriminator prosed along with the generation of synthesized images from the real images and followed by the introduction of contrastive loss to the images and provide them to the DeepFD classifier. First, we generate a lot of fake images from Example-Generative Model (EG) using DCGAN. This phase generates a hug number of fake images from the real images we have provided. Later, these synthesized images are combined with real images for contrastive loss. Afterward, the discriminator which is the DeepFD uses the previous images to further distinguish the fake images from real images. The process of proposed system is explained further:

A.Deep Convolutional GAN

The Deep Convolutional GAN is an approach to generative modeling using deep learning methods, such as convolutional neural networks. DCGAN is an extension of the GAN architecture for using deep convolutional neural networks for both the generator and discriminator models and configurations for the models and training that result in the stable training of a generator model. To train the discriminator, we need both true image and false image.

The Fake images were created by following two steps:

1.Create Fake images model using DCGAN.

2. Create Fake images using DCGAN based on learned model.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907056 |

To generate the training samples, the CelebaA dataset will be used in this experiment. The images in CelebA cover large pose variations and background clutter including 10,177 number of identities and 202,599 aligned face images. The DCGAN generates 200,000 fake images of size 64X64 into the fake image pool. These 200,000 fake images are combined with real images. The total of 216,780 images are used for training and 1,000 are testing images.

Create Fake images model using DCGAN: GANs are a framework for teaching a model to capture the training data's distribution so we can generate new data from that same distribution. They are made of two distinct models, a generator and a discriminator. The job of the generator is to spawn 'fake' images that look like the training images. The job of the discriminator is to look at an image and output whether or not it is a real training image or a fake image from the generator. During training, the generator is constantly trying to outsmart the discriminator by generating better and better fakes, while the discriminator is working to become a better detective and correctly classify the real and fake images.

To train the Fake images model we have divided the dataset into batches and trained the model in a total of 8 epochs.

The Batch size is a hyperparameter that defines the number of samples to work through before updating the internal model parameters.A training dataset can be divided into one or more batches.The number of epochs is a hyperparameter that defines the number times that the learning algorithm will work through the entire training dataset.

One epoch means that each sample in the training dataset has had an opportunity to update the internal model parameters. An epoch is comprised of one or more batches. We have built this model in TensorFlow, the batch size and epoch we have chosen are 64 and 8 respectively. While the model was getting trained all the processed images were saved as a 10X10 grid of a single sample image for every100 processed images. The sample image was shown below:

Fig 2 Sample Processed Images

This process creates checkpoints for every 1000 processed images. Checkpoints capture the exact value of all parameters used by a model. Checkpoints do not contain any description of the computation defined by the model and thus were typically only useful when source code that will use the saved parameter values is available. The Checkpoint with highest epoch can be used as the final Fake image model to generate Fake images.

B. Jointly Discriminative Feature Learning

The main drawback of supervised learning is that it is hard to identify the subject that is excluded from the training phase. To enhance the performance of the proposed method, we have introduced contrastive loss.To obtain the jointly discriminative feature for our task, we adopt pairwise information to guide the feature learning in two epochs.

C. Deep Forgery Discriminator

The Implementation of Deep Forgery Discriminator starts with setting up the TensorFlow environment. Later, the process starts with training and validation datasets we got from the previous process. Initialize the model for training set and validation sets based on the ResNet.

A Residual Neural Network (ResNet) is an artificial neural network (ANN) of a kind that builds on constructs known from pyramidal cells in the cerebral cortex. Residual neural networks do this by utilizing skip connections, or shortcuts to jump over some layers. Typical ResNet models are implemented with double- or

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907056 |

triple- layer skips that contain nonlinearities (ReLU) and batch normalization in between. An additional weight matrix may be used to learn the skip weights; these models are known as HighwayNets.Models with several parallel skips are referred to as DenseNets. In the context of residual neural networks, a non-residual network may be described as a plain network.

Deep Residual Network is almost similar to the networks which have convolution, pooling, activation and fully-connected layers stacked one over the other. The only construction to the simple network to make it a residual network is the identity connection between the layers. The screenshot below shows the residual block used in the network. You can see the identity connection as the curved arrow originating from the input and sinking to the end of the residual block.

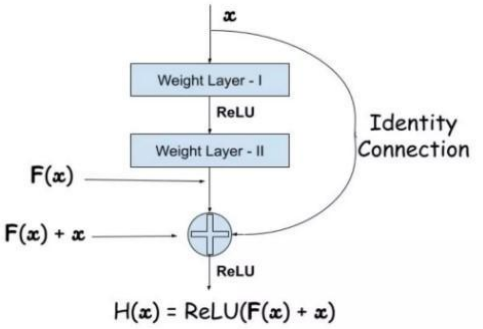

Fig 3 Residual Block of ResNet

After getting the 2d convolution computations the process was continued by feature extraction with a kernel size of 3 X 3 and 5 X 5 using Residual block function. In our project in residual function we have used only two convolution layers.We know neural networks are universal function approximators and that the accuracy increases with increasing number of layers. But there is a limit to the number of layers added that result in accuracy improvement. So, if neural networks were universal function approximators then it should have been able to learn any simplex or complex function.if we still keep increasing the number of layers, we will see that the accuracy will start to saturate at one point and eventually degrade. And, this is usually not caused due to overfitting. So, it might seem that the shallower networks are learning better than their deeper counterparts and this is quite counter-intuitive.

After initialization and forming the hard triplet loss sampler, we have then trained the fake detector using ResNet based on the triplet loss. This process is done batch wise and for every iteration we have found the Loss, Triplet Loss, Training Accuracy with a fixed learning rate. And for every 20 iterations we validate the model to find Validation Accuracy, Precision and Recall. After the optimization the final model was saved at the desired Location.It is clear that the proposed DeepFD is easily converged and reach higher performance.

To demonstrate the effectiveness of the proposed DeepFD, we separate the fake images generated by one of the collected GAN methods from the training pool. A generalized DeepFD should be able to detect the fake images even they are not used in the training. The proposed DeepFD is more generalized and effective than others. Therefore, the purely supervised approach (i.e., the proposed method without contrastive loss) cannot well capture the common features for the fake image. The proposed DeepFD is easier to extract the jointly discriminative feature for all kinds of the fake images, leading to higher performance.

Since the proposed DeepFD is designed to be a fully convolutional network, the feature maps can be visualized to localize the unrealistic details in the fake images.

III.EXPERIMENTAL RESULTS

The proposed DeepFD is easier to extract the jointly discriminative feature for all kinds of the fake images, leading to higher performance. The proposed DeepFD is designed to be a fully convolutional network, the feature maps can be visualized to localize the unrealistic details in the fake images.

The experimental results proved that the proposed discriminator can detect **94%** of fake images generated by DCGAN.

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.542 |

|| Volume 9, Issue 7, July 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0907056 |

The DeepFD discriminator was trained with the learning rate of 0.00010 (1e-4). The parameters for this discriminator used are learning rate which is set to 1e-4, the marginal value m is set to 0.8, with batch size 128. The Adam optimizer [6] is used for training of the Discriminator. The precision and recall rate for the proposed system were 94.3% and 94.2% respectively. Our experimental results demonstrate that the DeepFD discriminator outperforms other baseline approaches in terms of precision and recall rate.

IV.CONCLUSIONS

With this paper, we have tried to develop a Deep Forgery Discriminator (DeepFD) by embedding contrastive loss to the images, to detect fake/synthesized images generated by DCGAN. In this proposed system the main achievement is the introduction of contrastive loss to the images, which was used to distinguish well the fake/synthesized images generated by DCGAN. The experimental results demonstrate that the proposed method outperforms other baseline approaches in terms of precision and recall rate.

REFERENCES

[1] Popescu, A. C., & Farid, H. (2004). Exposing Digital Forgeries by Detecting Duplicated Image Regions. Technical Report, TR2004-515, Department of Computer Science, Dartmouth College, Hanover, New Hampshire, 2000, 1–11. http://os2.zemris.fer.hr/ostalo/2010 marceta/Diplomski files/102.pdf

[2] Amerini, I., Ballan, L., Caldelli, R., Del Bimbo, A., Serra,G. (2011). A SIFT-based forensic method for copy-move attack detection and transformation recovery. IEEE Transactions on Information Forensics and Security, 6(3 PART 2), 1099–1110. https://doi.org/10.1109/TIFS.2011.2129512

[3] Wei, W., Wang, S., Zhang, X., & Tang, Z. (2010). Estimation of image rotation angle using interpolation-related spectral signatures with application to blind detection of image forgery. IEEE Transactions on Information Forensics and Security, 5(3), 507–517.https://doi.org/10.1109/TIFS.2010.2051254

[4] Ke, Y., Qin, F., Min, W., & Zhang, G. (2014). Exposing image forgery by detecting consistency of shadow. The Scientific World Journal, 2014(d), 1–9. https://doi.org/10.1155/2014/364501

[5] E. Simo-Serra, et al. "Discriminative learning of deep convolutional feature point descriptors," Computer Vision (ICCV), 2015 IEEE International Conference on. IEEE, 2015

[6] I. Sutskever, et al. "On the importance of initialization and momentum in deep learning." International conference on machine learning. 2013

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 7.542

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details