Copy Move and Splicing Image Forgery Detection using CNN

Devjani Mallick[1, *], *Mantasha* Shaikh[2], *Anuja* Gulhane[3] and *Tabassum* Maktum[4]
[1,2,3,4]Ramrao Adik Institute of Technology, D. Y. Patil deemed to be University, Navi Mumbai, Maharashtra, India.

Abstract. The boom of digital images coupled with the development of approachable image manipulation software has made image tampering easier than ever. As a result, there is massive increase in number of forged or falsified images that represent incorrect or false information. Hence, the issue of image forgery has become a major concern and it must be addressed with appropriate solution. Throughout the years, various computer vision and deep learning solutions have emerged with a purpose to detect forgery in case of digital images. This paper presents a novel approach to detect copy move and splicing image forgery using a Convolutional Neural Network (CNN) with three different models i.e. ELA (Error Level Analysis), VGG16 and VGG19. The proposed method applies the pre-processing technique to obtain the images at a particular compression rate. These images are then utilized to train the model and further the images are classified as authentic or forged. The paper also presents the experimental results of the proposed method and performance evaluation in terms of accuracy.

Keywords- *Image, forgery, CNN, ELA, VGG16, VGG19, CASIA v2.O, NC2016*

1 Introduction

Image forensics is field in technology which aims to categorize authentic and forged images using various techniques like CNN etc., which is otherwise impossible by naked human eye to make out the difference. The recent growth of research in this field is mainly due to number of free image manipulation software available to common man, which can lead to various crimes, riots, misunderstandings, etc. that might create turbulence in the smooth conduct of public order and safety of citizens. We have witnessed an unprecedented growth of digital images, a trend which has been made possible by the numerous devices around, such as smart phones and tablets along with cheap and fast internet. Moreover, the development of user-friendly image manipulation software that is available at reasonable prices, has made the manipulation of such content easier than ever. Three of the most common manipulations in literature are: (1) Copy-move forgery, in which a specific region from the image is copy pasted within the same image. (2) Removal, in which an image region is removed and the removed part is then in-painted. (3) Splicing, in which a region from an authentic image is copied into a different image.

There have already been numerous approaches in the literature that address the task of image forgery detection. The traditional computer vision approaches have focused mainly on developing algorithms that tackle some of the aforementioned sub-tasks in isolation. The more recent research has focused on applying deep learning techniques for detecting image forgeries. Furthermore, the Convolutional Neural Network can learn how to extract the necessary features irrespective of the manipulation method. However, while some of these methods seem to solve the image forgery detection problem, the accuracy achieved raises suspicions about their robustness. Studies have shown that CNN network architecture achieves an impressive accuracy of approximately 70% on the CASIA v2.0 dataset for detection of copy move forgery. Following that, the same network architecture is trained and tested on the Media Forensics Challenge 2016 dataset whose tampered images are significantly more difficult to recognize. Finally, the effect of hyper parameter tuning and data augmentation on the network performance is analyzed in the paper.

The main goal of this study is to evaluate how the performance of an image forgery detection using CNN varies based on the sample difficulty. In order to address this issue the classification pipeline is developed in the proposed method. The existing approaches achieve high accuracy as the CAISA datasets used to test their network are manipulated in a way that is easy to recognize by humans. Therefore, the behavior of such a CNN on a more challenging dataset is analyzed in this paper. The performance of the CNN model may degrade significantly and to validate this intuition, two datasets are selected which are used to train the CNN. The CASIA v2.0 dataset is chosen over the v1.0 as it is considered more challenging and also offers nearly seven times more samples. This is particularly important given that previous studies have shown that a high number of training samples is a requirement for a reasonably well-trained CNN.

The rest of the paper is organized as follows:

The section 2 gives the survey of existing systems for detection of image forgeries. The section 3 demonstrates the proposed system to detect copy move and splicing image forgery presented. The results of the proposed system are presented in section 4. Finally, the conclusion and future work is expounded in Section 5.

[*] Corresponding author: anuja.gulhane@gmail.com

2 Related Work

In the paper [1], a reliable deep learning-based approach for detecting image forgeries in the context of double image compression is presented. To train the model, the difference between the original and recompressed versions of an image is used. The forged image is recompressed, and the difference between it and the original image is determined. The forged part is now highlighted due to the difference in the source of the forged part and the original part of the image. The proposed approach is compact, and its performance reveals that it outperforms current techniques. The experiment's results are promising, with a validation accuracy of 92.23 percent overall.

The paper [2] proposes a system architecture based on ResNet50v2 as basic convolutional model with five stages initialised with YOLO CNN weights for the purpose of object detection and transfer learning. In this paper, it first performs batch normalisation and then it applies activation function to update and optimise the weights. This paper addresses the degradation problem by means of deep learning framework.

The paper [3] proposes a blind image splicing detection system that uses a backbone of deep convolutional residual networks, followed by a fully connected classifier network that distinguishes between genuine and altered images. The paper [4] presents “a robust copy move forgery classification using end to end convolution neural network”, which provides a deep neural network-based technique with good outputs for categorizing photos based on whether they include any copy move forgeries. The following work seeks to classify all photos that have copy move forgeries and are scaled, rotated, or compressed at different levels. They proposed a novel CNN model that achieved a 93-95 percent accuracy with datasets of the same type or hybrid datasets.

A novel image splicing detection and localization approach based on deep convolutional neural networks is presented in [5]. The first layer of the CNN model is initialized using an optimal combination of the 30 fundamental high-pass filters used in the spatial rich model (SRM) for image steganalysis to suppress the effects of picture contents and extract more diversified and expressive residual features for RGB color images. The paper [6] proposes using a convolutional neural network to extract information from the spliced image automatically. After that, the discrete wavelet transform (dwt) is applied. Later, for classification, a support vector machine is used. Additional tests are carried out, in which the discrete cosine transform is used instead of dwt, and principal component analysis is used.

To deal with tampering and segmentation of visual data, paper [7] suggests a CNN model. To identify tampering, the model is designed to preserve local and global properties of data. Their approach is an inception-based module CNN network that uses mask r CNN to detect tampering. We demonstrated that our proposed technique can find out how to identify various photo tampering methods without relying on pre-selected highlights or any pre-preparing and segment out modified region through a series of tests. CAISA v1.0 and v2.0 datasets yielded 98.76 percent and 97.92 percent accuracy, respectively.

The goal of the paper [8] is to provide an efficient splicing detection and CMFD pipeline architecture whose major goal is to detect the traces left by various splicing and copy-move forgery post-processing operations, such as jpeg compression, noise addition, blurring, contrast adjustment, and so on. The accuracy of the COMOFOD and BOSSBASE datasets is 95.97% and 94.26 percent, respectively.

The paper [9] provides a unique strategy based on neural networks and deep learning to improve copy-move forgery detection, concentrating on the convolutional neural network architectural approach. To achieve satisfactory results, the suggested method uses a CNN architecture with pre-processing stages. The paper [10] uses a fusion processing technique that combines a deep convolutional model and an adversarial model to detect copy-move forgeries. A total of 4 datasets are used. The results show that the deep learning CNN and discriminator forgery detectors have a considerably high detection rate (95 percent). A two-branch design and a fusion module are used to build the network. Through CNN and GAN, the two branches are used to locate and identify copy-move forgery regions.

The paper [11] provides a new deep learning-based image fraud detection system for automatically learning hierarchical representations from input RGB color photographs. Image splicing and copy-move forgeries can be detected using the suggested CNN. The fundamental high-pass filter set employed in the spatial rich model (SRM) is utilized to establish the weights at the first layer of our network, which serves as a regularizer to efficiently suppress the effect of picture contents and capture the subtle artefacts created by the tampering operations. The pre-trained CNN is used as a patch descriptor to retrieve dense features from the test images, and the final discriminative features for SVM classification are obtained via a feature fusion technique. The suggested CNN-based model outperforms various state-of-the-art approaches, according to experimental results on multiple public datasets.

The authors in [12] offer a shallow convolutional neural network (SCNN) that can identify manufactured region boundaries from original edges in low-resolution images. SCNN was created to make use of chroma and saturation data. Two techniques based on SCNN, termed sliding windows detection (SWD) and fast SCNN, have been developed to detect and identify image forgery regions.

3 Proposed Work

The overall system architecture is shown in Figure 1. The major components of the proposed methodology are, pre-processing, error level analysis and CNN. These modules are elaborated in detail next in this section.

(a) Data pre-processing: pre-processing involves normalization of images. the purpose of normalization is to make sure that all the images have similar data distribution. For normalization, the whole dataset is resized into 128*128 pixels.

(b) Error level analysis: To convert an image to ELA, the pre-processed photos must first be resaved at a certain quality level. The image is whitened or brightened as a result of this technique. In order to resave the images, consider both genuine and falsified images that have been pre-processed and then resave photos at a specific compression level. Finally, Images that have been pre-processed and images that have been re-saved are compared to see how much of a difference there is. The modified parts of the image in the forged ELA-generated image are brighter than the corresponding original components. Having ELA in the preprocessing tends to have a huge advantage in further processing of the neural network as the ELA converted image contains only non-redundant information and has similar intensity contrast as compared to the nearby pixels. Due to this reason, our neural network training optimizes in only 8-9 epochs with a learning rate of 0.0001.

The next step involves changing the image size. We want each RGB value to lie between 0 and 1 so we divide each cell value by 255.0 to normalize the training of the neural network so that it converges faster. After that each image is categorically labelled where 0 stands for authentic image and 1 stands for forged image. Subsequently, the images are divided into two sets – 80% of the images are taken for the training set and the remaining 20%for the validation set for the working of the convolutional neural network.

(c) Convolutional neural network: Convolutional layer is used as feature extractor that learns feature representation. This is from the image that is input to CNN. Meanwhile, the pooling layer shrinks the convolution layer's output map and prevents overfitting. In general, there are stacks of numerous convolutional and pooling layers before the fully connected layer that serve to extract a more abstract feature representation. Pooling layer usually decodes the image (such as 2x2) in the aggregation into a single unit. The first layer of the CNN is a convolutional layer with a kernel size of 5x5 and a total of 32 filters. The second layer CNN consists of a convolutional layer with a kernel size of 5x5 and 32 filters, as well as a max pooling layer with a size of 2x2. After that, the max pooling layer adds dropouts for 0:25 to avoid overfitting. The flatten layer changes the feature map by flattening it into the feature vector and passing it to a fully connected layer after images are trained from a finetuned pre-processed model. In the last dense layer of models, the fully connected layer is employed for pattern recognition, and the SoftMax activation function is used to convert the feature vector in a probabilistic manner. The training set is compared to the test set using SoftMax activation, and a probability distribution on actual and forged images is returned.

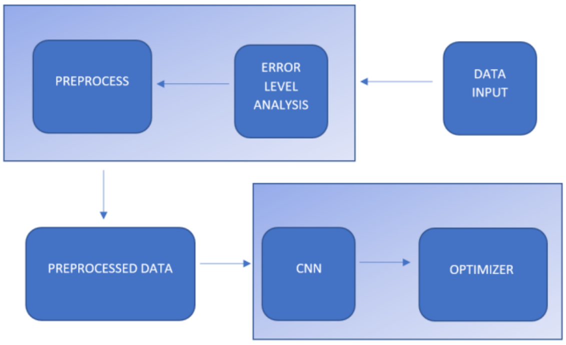


Figure 1. System Architecture

Optimization applied during the training is RMS Prop Optimizer in which the learning rate for each parameter is automatically adjusted without our intervention which helps in optimizing several parameters like number of features, number of training samples, target MSE, number of hidden layers, etc.

For the training of the neural network, two popular neural architectures have been utilized mainly VGG16 and VGG 19.

A. VGG16

VGG 16 is a CNN consisting of 22 neural layers out of which the end one is a SoftMax output classifier (4 dense layers at the end). The ELA converted training set images are then passed to finetuned VGG-16 model consisting of 16 layers to generate the model which is trained on a training set of 13,000 images out of which 6,500 are authentic and the rest are forged images.

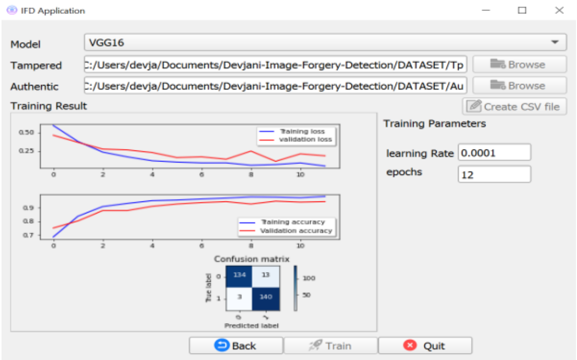
VGG16 architecture consists of two convolutional layers which consists of mainly 64 channels of size 3 by 3 kernel, one maxpool layer of size and stride 2 by 2, two convolutional layers of 128 channels, one maxpool layer of size and stride 2 by 2, three convolutional layers of 256 channels, one maxpool layers of 2 by 2 pool, three convolutional layers of 512 channels and finally maxpool layers. RELU layers are added at each step so as to filter all the negative values from previous layers, which is finally passed on to dense layer to flatten the output and then add dense layers of 4096 units and dense softmax layers of 2 units. The SoftMax layer will show values between 0 and 1 based on the model that which category the input belongs to. So, with learning rate as 0.0001 and 12 epochs, from the confusion matrix we get the training accuracy for VGG16 as 94.4%.

B. VGG-19

VGG 19 is neural network having 24 layers with similar chronology distribution as compared to VGG 16. Pre-processed output after ELA is then passed on to VGG19 model which has total of 19 layers. Afterwards the images are classified based on SoftMax activation where training set id utilized to make comparison with the test set and return the probability distribution of different classed. The VGG19 architecture consists of two convolutional layers, one pooling layer, two more convolutional layers, again pooling layer, two more convolutional layers, and two more stacks of three convolutional layers separated by maxpool layer, followed by SoftMax layer at output. So, with learning rate as 0.0001 and 12 epochs, from the confusion matrix we get the training accuracy for VGG19 as 95%.

4 Results & Discussion

For training images, for every model, a mix of CASIA-2.0 and NC2016 dataset was considered. The dataset consists of RGB images from two these two datasets, which are split into training and testing datasets, and then passes on to our CNN models to classify them into two classes i.e. authentic and forged. First step is to divide the dataset into 2 categories: original and fake images. Therefore a total of 13000 images were trained using ELA, VGG16 and VGG19 model with ratio of authentic -tampered images being 50-50.And similarly for testing, a total of 1000 images with 500 authentic and 500 tampered were considered for each model. VGG16 Training result is shown in Figure 2.


IFD Application

Model: VGG16

Tampered: C:/Users/devja/Documents/Devjani-Image-Forgery-Detection/DATASET/Tp

Authentic: C:/Users/devja/Documents/Devjani-Image-Forgery-Detection/DATASET/Au

Training Result

Training Parameters

learning Rate: 0.0001

epochs: 12

Figure 2. VGG16 Training Curve

So, with learning rate as 0.0001 and 12 epochs, from the confusion matrix we get the training accuracy for VGG16 as 94.4%. VGG19. The training result is shown in Figure 3.

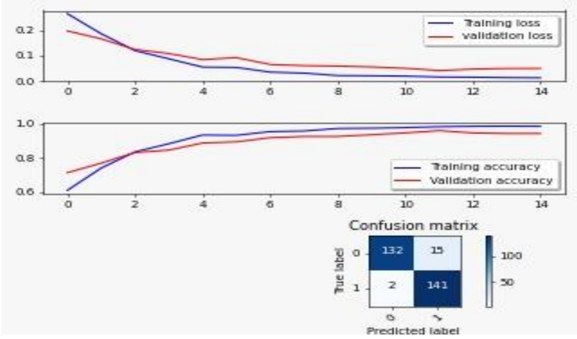
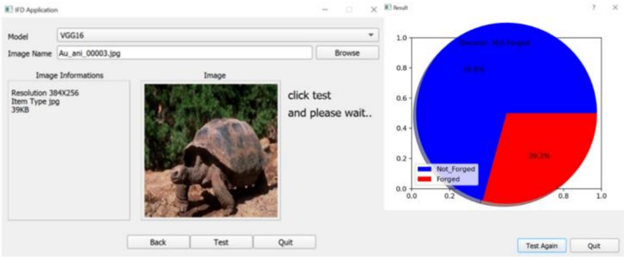


Figure 3. VGG19 Training Curve

So, with learning rate as 0.0001 and 12 epochs, from the confusion matrix we get the training accuracy for VGG19 as 95%. After training is completed, testing part begins on all three models i.e. ELA, VGG16 and VGG19. Also the threshold for all three models was kept as 50% , i.e. if CNN model predicts accuracy of authenticity as greater than 50% , then it will be labelled as authentic else forged. The example of one such test image on VGG16 model has been shown in Figure 4.


IFD Application

Model: VGG16

Image Name: Au_ani_00003.jpg

Image Informations

Resolution 384X256
Item Type jpg
39KB

click test
and please wait..

Figure 4. VGG16 Test Result

The overall comparison of different models along with their accuracy has been summarized in the Table 1. The accuracy of different models with respect to separate authentic and tampered images and hence overall accuracy have been presented with the help of bar chart in Figure 5.

Table 1:Comparison of Different Models

model	ELA	VGG 16	VGG 19
true positive	315	347	329
true negative	185	153	171
false positive	391	369	400
false negative	109	131	100
accuracy	70.6%	71.6%	72.9%

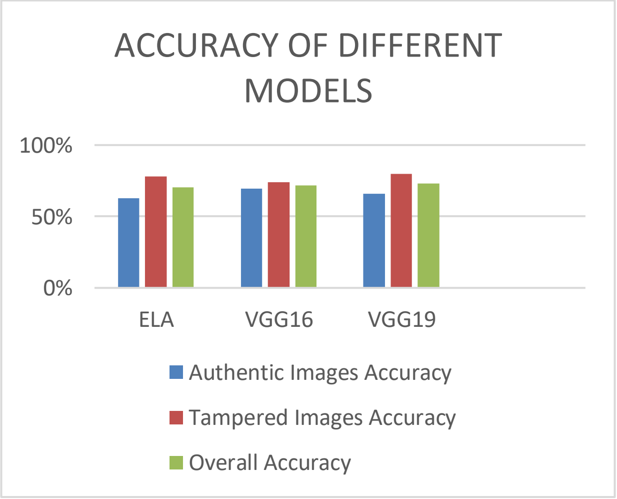


Figure 5. Accuracy comparison of different models

As can be inferred from the above bar-chart, VGG19 is the best model when it comes to accuracy of predicting forged images as forged whereas the same is true for

VGG16 in case of authentic images. The overall accuracy VGG19 comes out to be slightly better than VGG16 if not similar.

5 Conclusion

In this paper a method to detect copy move image forgery using CNN is presented. The proposed method uses CNN network to extract features from two datasets of varying difficulty, namely CASIA v2.0 and NC2016. The experimental results validate that the classification performance decreases when the samples are more challenging. However, the implemented architecture does not easily generalize to datasets with different underlying distributions. The proposed model is based on three different models i.e. ELA, VGG16, and VGG19 is able to achieve good results, with their respective accuracies as 70.6%, 71.6% and 72.9% , on a set of images from CASIA2.0 and NC2016 datasets. Nevertheless, there is surely a lot of work still to be done in the image forgery detection domain and neural networks will be able to detect tampered images regardless of their difficulty. In future there is scope to improvise the VGG19 training model by increasing the dataset and using high computational power systems.

References

[1] S.S. Ali, I.I. Ganapathi, N.S. Vu, S.D. Ali, N. Saxena, N. Werghi, *Image Forgery Detection Using Deep Learning by Recompressing Images*, Electronics 11(3), 403 (2022).

[2] E.U.H. Qazi, T. Zia, A. Almorjan, *Deep Learning-Based Digital Image Forgery Detection System*, Appl. Sci., 12, 2851 (2022).

[3] S. Nath, R. Naskar, *Automated image splicing detection using deep CNN learned features and ANN-based classifier*, Springer-Verlag London ltd., part of springer nature (2021).

[4] S. Kumar, S.K. Gupta, *A robust copy-move forgery classification using end to end convolution neural network*, 8th international conference on reliability, Infocom Technologies and optimization (Trends and Future Directions) (ICRITO) (June 2020).

[5] Y. Rao, J. Ni, *Deep learning local descriptor for image splicing detection and localization*, IEEE access, v**ol. 8** (2020).

[6] Eman I. Abd El-Latif, A. Taha, Hala H. Zayed, *A passive approach for detecting image splicing using Deep Learning and Haar Wavelet Transform*, Arabian journal for science and engineering, **vol. 6** (2020).

[7] S. Saleem, A. Dilawariand U.G. khan, *Multimedia Forensic: An approach for splicing detection based on deep visual features*, 2019 international conference on robotics and automation in industry (ICRAI) (2019).

[8] R. Thakur, R. Rohilla, *Copy-Move forgery detection using residuals and convolutional neural network framework: a novel approach*, 2019 international conference on robotics and automation in industry (ICRAI) (2019)

[9] Y. Abdalla 1, M.T. Iqbal, M. Shehata, *Convolutional Neural Network for copy-move forgery detection*, conference: computer vision and pattern recognition Doi: 10.11.19/ICPR (Nov 2019).

[10] Y. Abdalla 1, M.T. Iqbal, M. Shehata, *Copy-Move forgery detection and localization using a generative adversarial network and convolutional neural network*, article - department of computer science, math, physics, and statistics, university of British Columbia, Kelowna, BC v6t 1z4, Canada (2019).

[11] Y. Rao, J. Ni, *A deep learning approach to detection of splicing and copy-move forgeries in images*, 2016 IEEE international workshop on information forensics and security (WIFS) (2016).

[12] Zhang, Zhongping & Zhang, Yixuan & Zhou, Zheng & Luo, *Boundary-based image forgery detection by fast shallow CNN*, conference: computer vision and pattern recognition Doi: 10.11.09/ICPR (2009).