JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# A Deep Learning Framework for Tampering Detection and Localization towards Digital Image Forensics.

**Siva Prasad Patnayakuni**
Senior Data Engineer, H-E-B, TX, USA

**Abstract**

Deep learning became very significant to solve problems associated with computer vision applications. In this paper, we proposed a deep learning framework that exploits three deep learning models for automatic detection of image tampering detection towards digital forensics. The framework exploits pre-trained models such as ResNet50, VGG16 and Inception. These models are widely used in solving different real world problems. They are known for efficiency in image processing. A supervised learning approach is used to have training and testing phases. Out of the three models we enhanced ResNet50 model with transfer learning. CASIA 2 is the dataset used in the experiments. This dataset contains benchmark samples used for image forgery detection research. The proposed framework has mechanisms and algorithm to detect tampered images automatically. To realize our framework, we proposed an algorithm known as Learning based Image Tampering Detection (LbITD). It CASIO 2 dataset as input, performs pre-processing to generate train and test datasets. Then the algorithm proceeds with deep learning model including model creation, model compilation and model training. Our experimental results showed that ResNet50 outperforms other models with 92.45% accuracy.

***Keywords*** *–Deep Learning, Image Tampering Detection, VGG16, ResNet50, Inception*

## 1. INTRODUCTION

Multimedia content is being generated in contemporary era with unprecedented pace. In other words, with the cloud innovation, companies are preserving image content like never before. Even in computer vision applications and government applications, it became very important to deal with images. Images are widely used in different applications for purposes such as identification, comparison and so on. Human identification is an important issue to be considered. There are commercial companies producing images with copyrights. In this context, it is important to develop a strong mechanism to detect image forgeries. This kind of solution helps in improving digital image forensics as well. Machine learning and deep learning techniques, of late, are contributing to computer vision applications. In this paper, we used deep learning for automatic detection of image tampering. There are many existing approaches found in the literature for digital image forensics.

In [4] computer generated faces are used for the study of image tampering detection using deep learning. In [5] an attack known as face morphing attack and its related tampering was explored. In [14] CNN based methodology was proposed for tampering localization while wavelet dual-branch network is defined in [15] for face tampering detection. In [16] fusion approaches are studied to detect copy-move forgery in images. In [17], deep learning models were used to diagnose bearing faults in automotive industry. Identification of texture difference is the main method employed in [18] for image tampering detection. From the literature, it was observed that CNN based models were widely used for image tampering detection and computer vision applications. It was also found that pre-trained models could provide better performance in image processing. Based on these observations, in this paper, we proposed framework to exploit three deep learning models. Our contributions in this model are as follows.

1. We proposed a deep learning framework that exploits three deep learning models for automatic detection of image tampering detection towards digital forensics.
2. We proposed an algorithm known as Learning based Image Tampering Detection (LbITD).
3. We built an application to evaluate the framework and the underling algorithm that exploits three deep learning models.

The remainder of the paper is structured as follows. Section 2 reviews literature on different existing methods for image tampering detection. Section 3 presents the proposed framework. Section 4 presents results of our experiments. Section 5 concludes our work and provides future scope.

## 2. RELATED WORK

This section reviews literature on various existing methods. Different methods used for image tampering detection are explored in [1], [8] and [20]. As discussed in [19], the raise of machine learning has helped in making novel approaches to detection image tampering automatically. Later on many deep learning models were used by the researchers for image tampering detection. In [2], [3] and [6] copy-move forgery is detected using deep learning. In [4] computer generated faces are used for the study of image tampering detection using deep learning. In [5] an attack known as face morphing attack and its related tampering was explored. In [7] deep auto encoder is used to detect forgery. The model is actually combined with CNN extracted features. In [9] passive image forgery detection method is explored using deep learning.

In [10], both copy-move and splicing kind of forgeries are detected using ML techniques. In [11], CNN and fuzzy C means algorithms are combined to detect image forgery. Many learning based approaches towards image forensics are discussed in [12] and [13]. In [14] CNN based methodology was proposed for tampering localization while wavelet dual-branch network is defined in [15] for face tampering detection. In [16] fusion approaches are studied to detect copy-move forgery in images. In [17], deep learning models were used to diagnose bearing faults in automotive industry. Identification of texture difference is the main method employed in [18] for image tampering detection. From the literature, it was observed that CNN based models were widely used for image tampering detection and computer vision applications. It was also found that pre-trained models could provide better performance in image processing. Based on these observations, in this paper, we proposed framework to exploit three deep learning models.

## 3. PROPOSED FRAMEWORK

We proposed a deep learning based framework, as shown in Figure 1, for automatic detection of image tampering. The framework exploits pre-trained models such as ResNet50, VGG16 and Inception. These models are widely used in solving different real world problems. They are known for efficiency in image processing. A supervised learning approach is used to have training and testing phases. CASIA 2 [27] is the dataset used in the experiments. This dataset contains benchmark samples used for image forgery detection research. The proposed framework has mechanisms and algorithm to detect tampered images automatically.

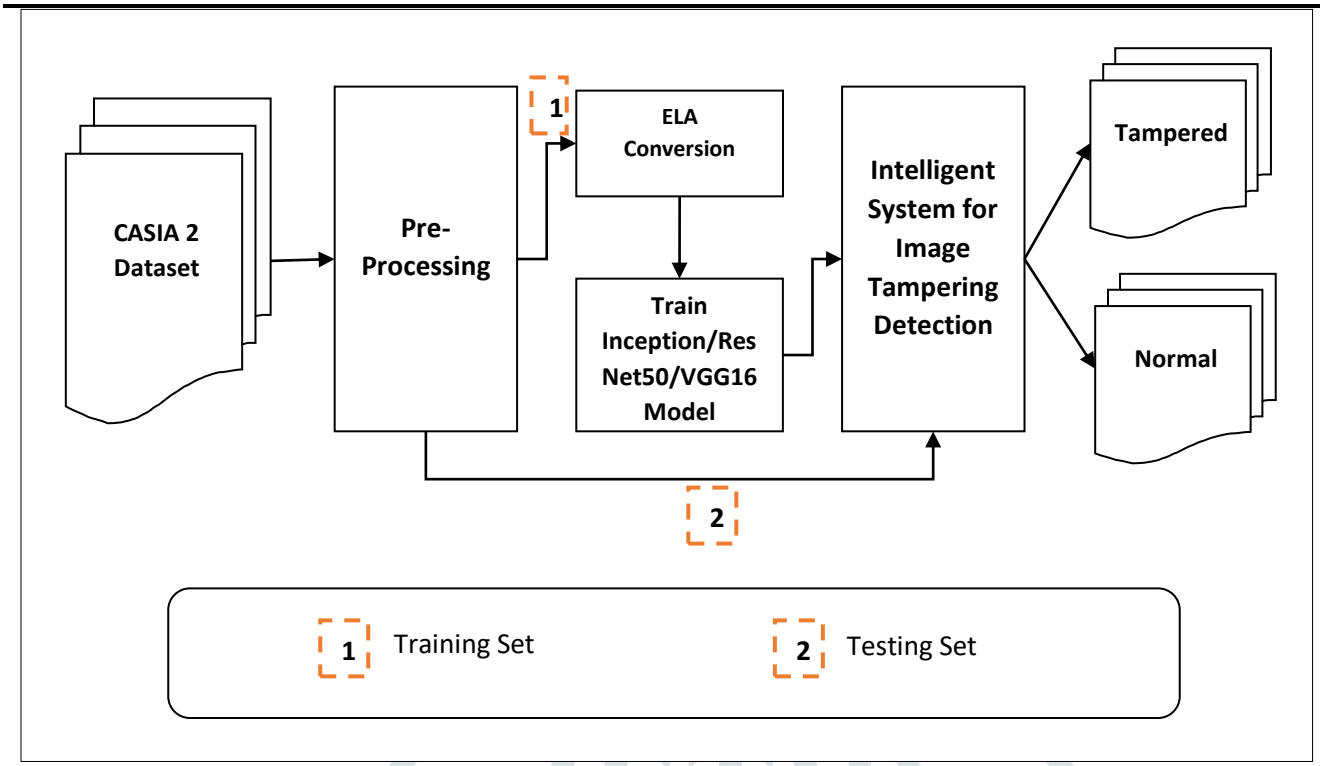


**Figure 1:** Proposed framework based on deep learning techniques for image tampering detection

The given dataset is divided into training and test data in order to perform supervised learning. Once data is split into two parts, the data is subjected to ELA (Error Level Analysis) process that enables ease of detection of tampering. Afterwards, the ELA form of training images is used to train different deep learning classifiers. They are known as Inception, ResNet50 and VGG1 models. Out of them we improved ResNet50 model with transfer learning for better performance. Once training of the models is completed, it results in an intelligent system for automatic tampering detection.

| Layer (type) | Output Shape | Param # |
|---|---|---|
| conv2d_9 (Conv2D) | (None, 126, 126, 32) | 896 |
| max_pooling2d_9 (MaxPooling2 | (None, 63, 63, 32) | 0 |
| conv2d_10 (Conv2D) | (None, 61, 61, 64) | 18496 |
| max_pooling2d_10 (MaxPooling | (None, 30, 30, 64) | 0 |
| conv2d_11 (Conv2D) | (None, 28, 28, 128) | 73856 |
| max_pooling2d_11 (MaxPooling | (None, 14, 14, 128) | 0 |
| conv2d_12 (Conv2D) | (None, 12, 12, 128) | 147584 |
| max_pooling2d_12 (MaxPooling | (None, 6, 6, 128) | 0 |
| dropout_3 (Dropout) | (None, 6, 6, 128) | 0 |
| flatten_3 (Flatten) | (None, 4608) | 0 |
| dense_5 (Dense) | (None, 512) | 2359808 |
| dense_6 (Dense) | (None, 2) | 1026 |

Total params: 2,601,666
Trainable params: 2,601,666
Non-trainable params: 0

**Figure 2:** Inception model used for image tampering detection

As presented in Figure 2, the inception model is used to learn from training samples. Its architecture contains number of convolutional layers and max pooling layers. They are used to extract features from image and optimizing the features respectively. It has dropout, flatten and dense layers configured appropriately for image tampering detection.

| Layer (type) | Output Shape | Param # |
|---|---|---|
| resnet50 (Model) | (None, 4, 4, 2048) | 23587712 |
| global_average_pooling2d_1 ( | (None, 2048) | 0 |
| dense_1 (Dense) | (None, 256) | 524544 |
| dropout_1 (Dropout) | (None, 256) | 0 |
| batch_normalization_1 (Batch | (None, 256) | 1024 |
| dense_2 (Dense) | (None, 2) | 514 |

Total params: 24,113,794
Trainable params: 578,690
Non-trainable params: 23,535,104

**Figure 3:** ResNet50 model enhanced with transfer learning

As presented in Figure 3, the ResNet50 base mode is improved with additional layers using transfer learning. The additional layers include global average pooling (for feature optimization), two dense layers, dropout layer and batch normalization layers. With the enhancement of ResNet50 model, it was observed that the detection accuracy could be improved.

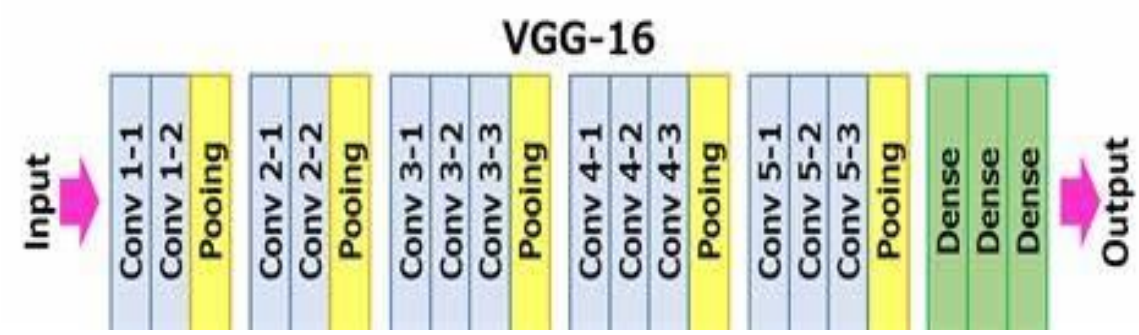


**Figure 4:** VGG 16 model used for image tampering detection

As presented in Figure 4, the VGG 16 model has its layers containing convolutional and pooling layers followed by dense layers to deal with computer vision applications. In this paper, these layers are meant for learning from train images and then classify test images into tampered and normal images.

**Algorithm:**Learning based Image Tampering Detection (LbITD)

**Input:** CASIO 2 dataset D

**Output:** Tampering detection results R, performance statistics P

1. Begin
2. (T1, T2)←Pre-process(D)
3. M←CreateDeepLearningModel() //Incpetion/VGG16/ResNet50
4. M←CompileModel()
5. M←TrainModel(T1)
6. Save model M
7. (R,P)←TestData(M, T2)
8. Display R
9. Display P
10. End

**Algorithm 1:**Learning based Image Tampering Detection (LbITD)

As presented in Algorithm 1, it takes CASIO 2 dataset as input, performs pre-processing to generate train and test datasets. Then the algorithm proceeds with deep learning model including model creation, model compilation and model training. Once the model is trained with T1, it is subjected to persisting to reuse in future. In testing phase, the model M is reused to test unlabelled data (T2). Then the algorithm computes result and also performance statistics and display the same. Based on confusion matrix, the evaluation of the proposed algorithm is compared with the state of the art. Table 1 shows different metrics used in the evaluation process.

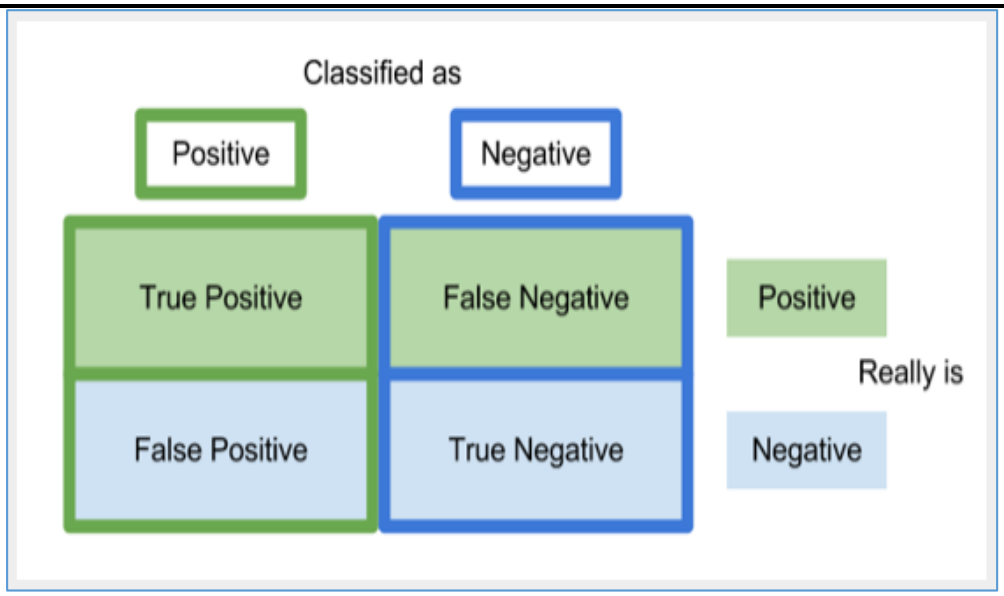


**Figure 5:** Confusion matrix

Based on the confusion matrix presented in Figure 5, the confusion matrix shows the measures like true positive (TP), false positive (FP), false negative (FN) and true negative (TN). These are determined by comparing result of ML algorithm when compared with the ground truth.

| **Metric** | **Formula** | **Value range** | **Best Value** |
|---|---|---|---|
| Accuracy | $\frac{TP+TN}{TP+TN+FP+FN}$ | [0; 1] | 1 |
| Precision (p) | $\frac{TP}{TP+FP}$ | [0; 1] | 1 |
| Recall (r) | $\frac{TP}{TP+FN}$ | [0; 1] | 1 |
| F1-Score | $2*\frac{(p*r)}{(p+r)}$ | [0; 1] | 1 |

**Table 1:** Performance metrics used for evaluation

Precision refers to positive predictive value while the recall refers to true positive rate. F1-score is the harmonic mean of both precision and recall which is used to have a measure without showing imbalance while accuracy measure may show imbalance.

## 4. RESULTS AND DISCUSSION

This section presents results of experiments. The results include the observations with each deep learning model. Each model showed different level of performance due to their modus operandi and internal functionality.

### 4.1 Results of Inception Model

This sub section presents experimental results of inception model which is used for image tampering detection.

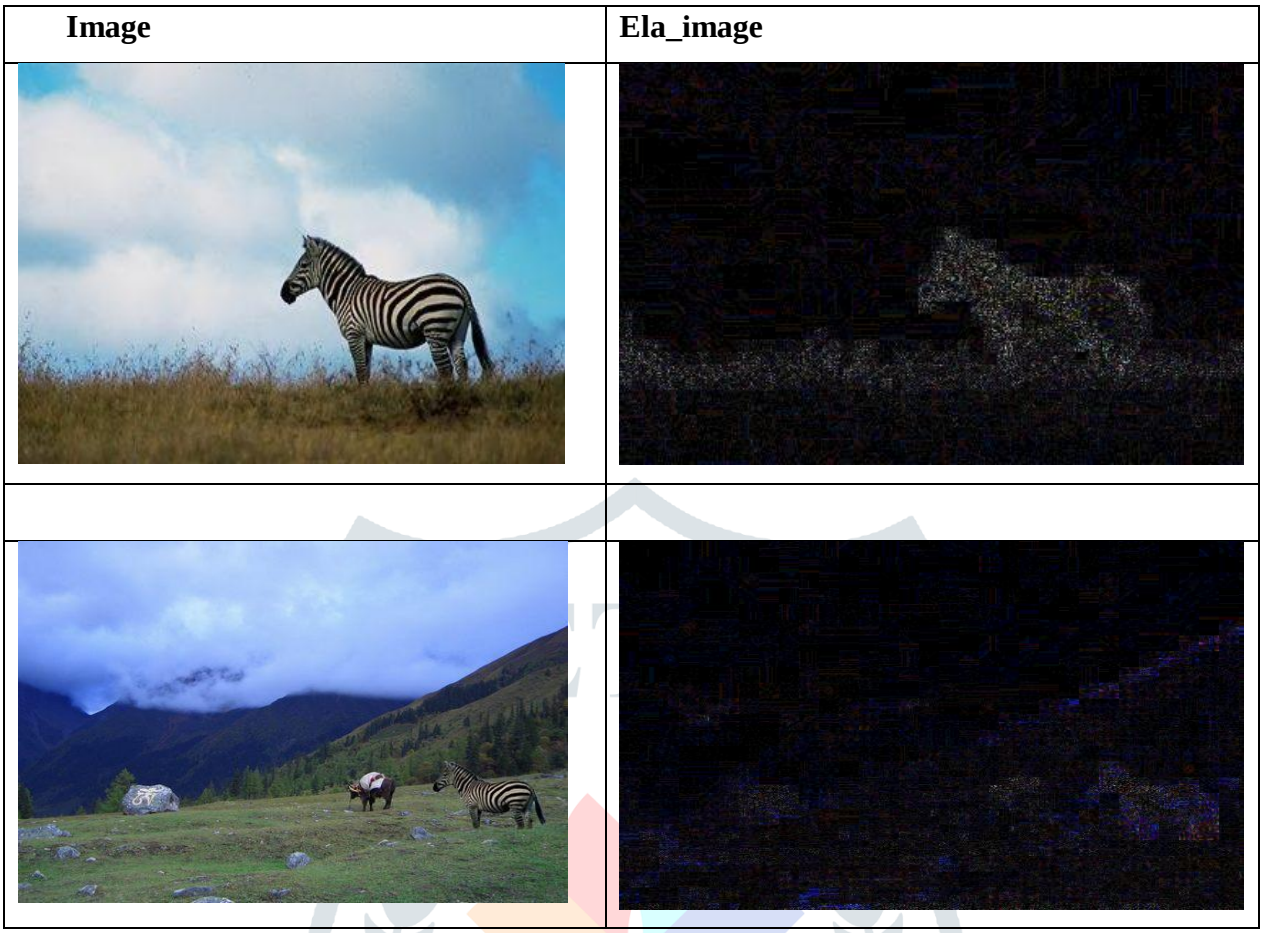


**Figure 6:** Results associated Inception model

As presented in Figure 6, the input images are subjected to ELA processing prior to training Inception model. The results of ELA are shown here.

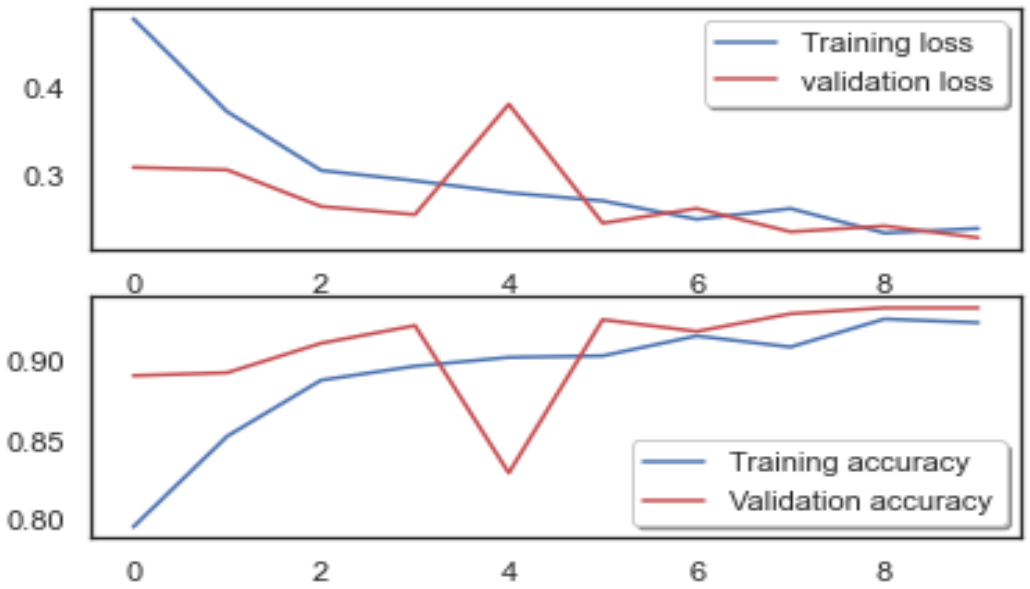


**Figure 7:** Shows results of Inception model

As presented in Figure 7, the loss and accuracy results of Inception model against different number of epochs are provided.

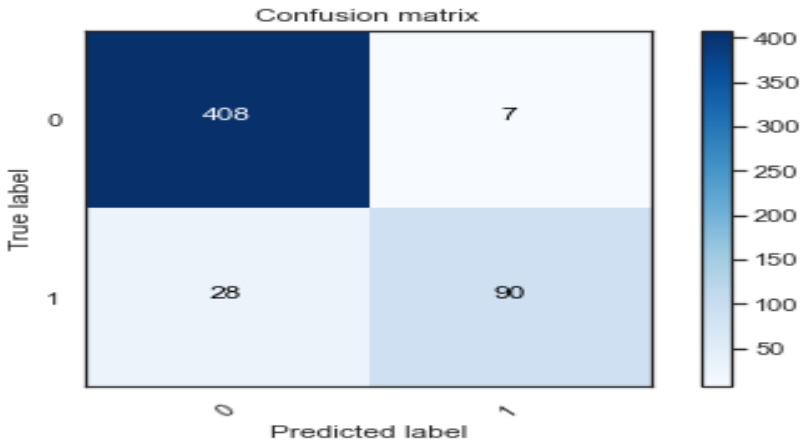


**Figure 8:** Shows confusion matrix reflecting results of Inception model

As presented in Figure 8, it shows confusion matrix reflecting detection performance of the Inception model.

## 4.2 Results of ResNet50 Model

This sub section presents experimental results of inception model which is used for image tampering detection.


**Figure 9:** Results associated ResNet50 model

As presented in Figure 9, the input images are subjected to ELA processing prior to training ResNet50 model. The results of ELA are shown here.

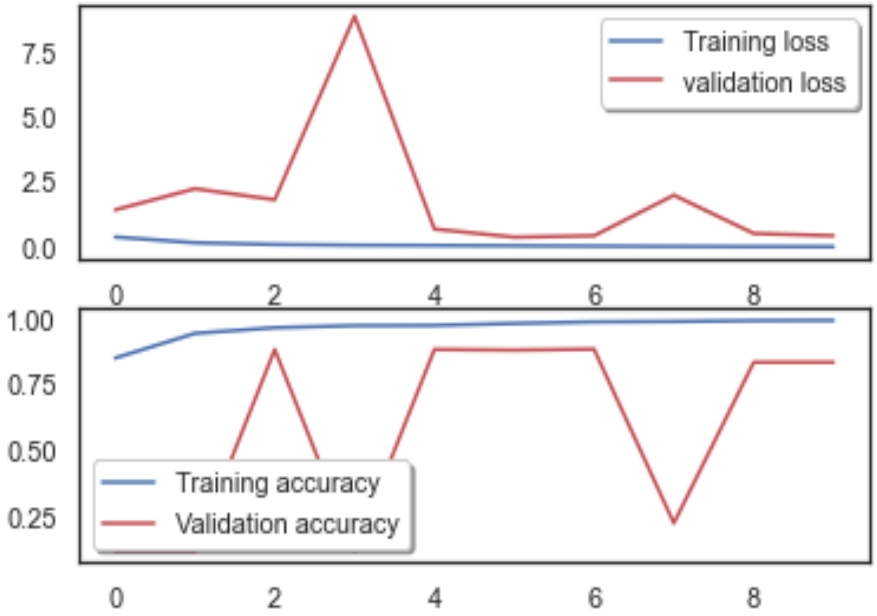


**Figure 10:** Shows results of ResNet50 model

As presented in Figure 10, the loss and accuracy results of ResNet50 model against different number of epochs are provided.

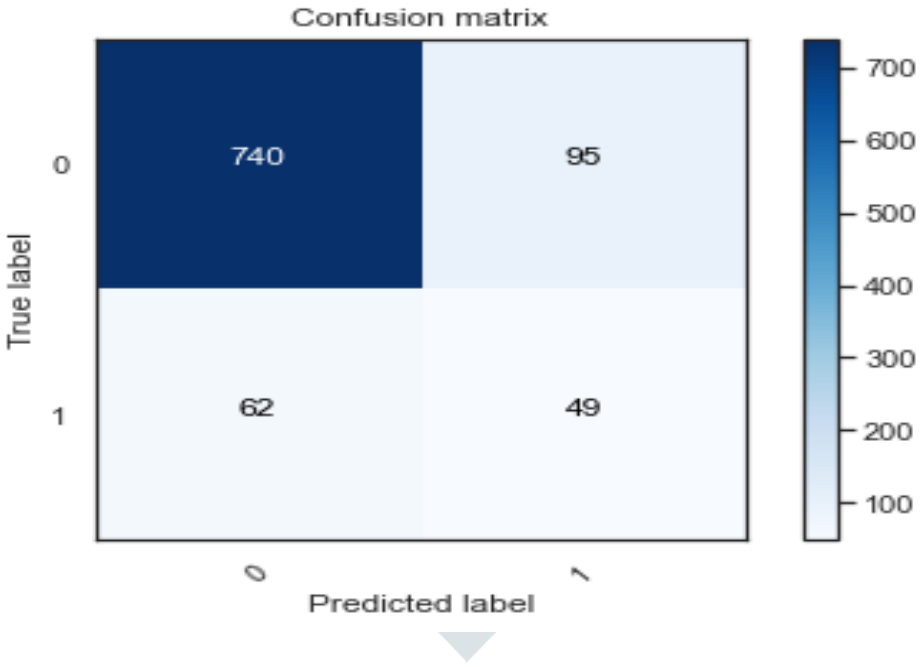


**Figure 11:** Shows confusion matrix reflecting results of ResNet50 model

As presented in Figure 11, it shows confusion matrix reflecting detection performance of the ResNet50 model.

### 4.3 Results of VGG16 Model

This sub section presents experimental results of inception model which is used for image tampering detection.


**Figure 12:** Results associated VGG16 model

As presented in Figure 12, the input images are subjected to ELA processing prior to training VGG16 model. The results of ELA are shown here.

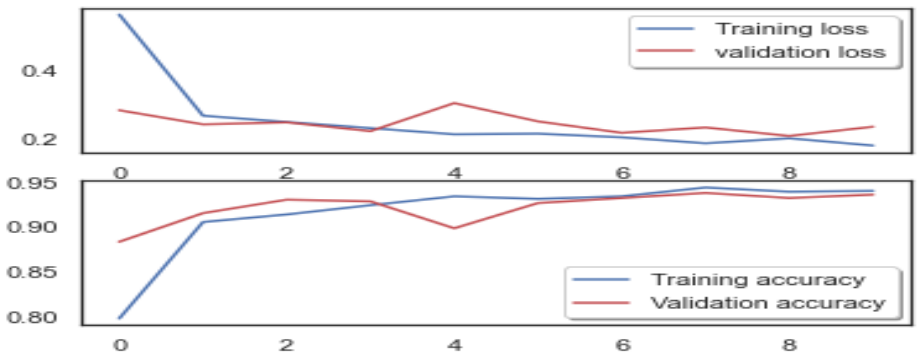


**Figure 13:** Shows results of VGG16 model

As presented in Figure 13, the loss and accuracy results of VGG16 model against different number of epochs are provided.

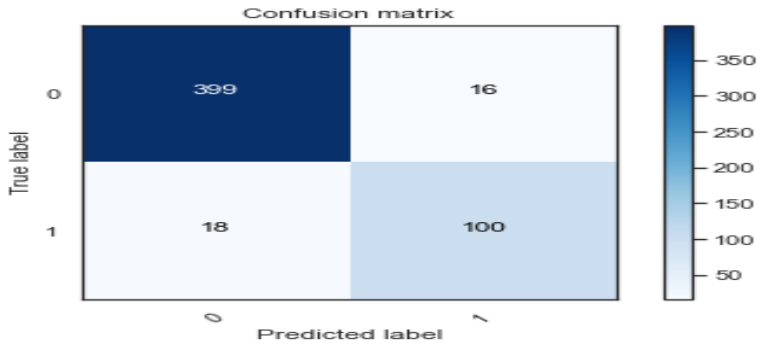


**Figure 14:** Shows confusion matrix reflecting results of VGG16 model

As presented in Figure 14, it shows confusion matrix reflecting detection performance of the VGG16 model.

## 4.4 Performance Comparison

This section presents performance comparison among the three models used in the empirical study meant for automatic detection of image tampering. Observations are made in terms of accuracy of the deep learning models.

| Image Tampering Detection Model | Accuracy (%) |
|---|---|
| Inception | 83.97 |
| Resnet | 92.45 |
| VGG16 | 86.96 |

**Table 2:** Shows performance of different models

As presented in Table 2, the performance of different models is compared in terms of accuracy in tampering detection.

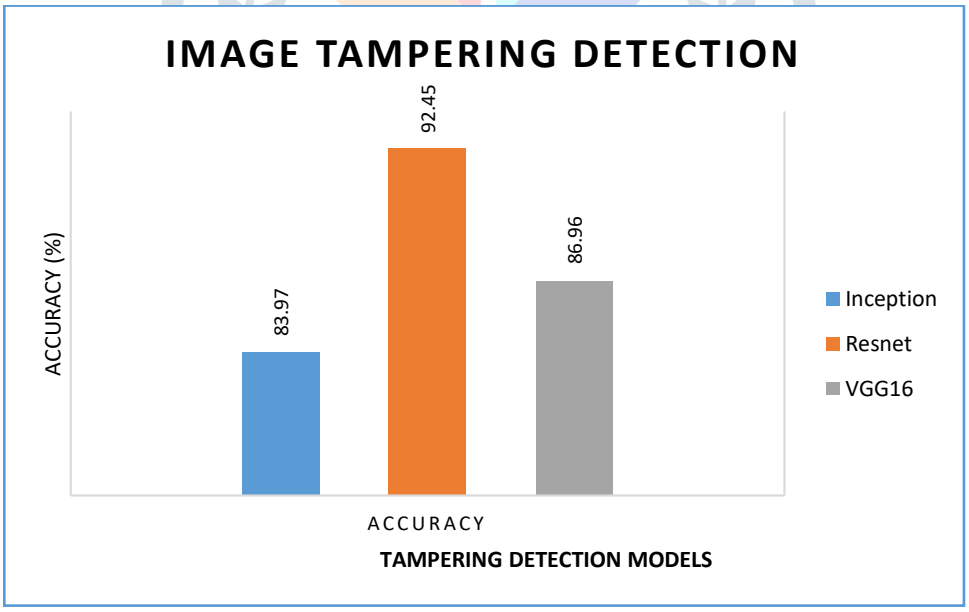


**Figure 15:** Performance comparison among models

As presented in Figure 15, three models used in the experiment are compared to know their performance for automatic image tampering detection. Each deep learning model was found to have different level of performance. In fact, the models differ in their architecture and also learning process. Therefore, there is difference in accuracy evident in the experimental results. Accuracy exhibited by Inception mode is 83.97%. VGG16 model achieved 86.96% accuracy while the ResNet50 model achieved highest performance with 92.45% accuracy.

## 5. CONCLUSION AND FUTURE WORK

In this paper, we proposed a deep learning framework that exploits three deep learning models for automatic detection of image tampering detection towards digital forensics. The framework exploits pre-trained models such as ResNet50, VGG16 and Inception. These models are widely used in solving different real world problems. They are known for efficiency in image processing. A supervised learning approach is used to have training and testing phases. Out of the three models we enhanced ResNet50 model with transfer learning. CASIA 2 is the dataset used in the experiments. This dataset contains benchmark samples used for image forgery detection research. The proposed framework has mechanisms and algorithm to detect tampered images automatically. To realize our framework, we proposed an algorithm known as Learning based Image Tampering Detection (LbITD).Accuracy exhibited by Inception mode is 83.97%. VGG16 model achieved 86.96% accuracy while the ResNet50 model achieved highest performance with 92.45% accuracy. In future we intend to improve the models further with model scaling for dealing with large volumes of data.

**References**

[1] da Costa, Kelton A.P.; Papa, JoÃ£o P.; Passos, Leandro A.; Colombo, Danilo; Ser, Javier Del; Muhammad, Khan and de Albuquerque, Victor Hugo C. (2020). A critical literature survey and prospects on tampering and anomaly detection in image data. Applied Soft Computing, 106727.

[2] Elaskily, Mohamed A.; Elnemr, Heba A.; Sedik, Ahmed; Dessouky, Mohamed M.; El Banby, Ghada M.; Elshakankiry, Osama A.; Khalaf, Ashraf A. M.; Aslan, Heba K.; Faragallah, Osama S. and Abd El-Samie, Fathi E. (2020). A novel deep learning framework for copy-moveforgery detection in images. Multimedia Tools and Applications,.

[3] YohannaRodriguez-Ortega;Dora M. Ballesteros and Diego Renza; (2021). Copy-Move Forgery Detection (CMFD) Using Deep Learning for Image and Video Forensics . Journal of Imaging,.

[4] Dang, L. M., Min, K., Lee, S., Han, D., & Moon, H. (2020). Tampered and Computer-Generated Face Images Identification Based on Deep Learning. Applied Sciences, 10(2), 505.

[5] Kraetzer, Christian; Shi, Yun-Qing; Dittmann, Jana and Kim, HyoungJoong (2017). [Lecture Notes in Computer Science] Digital Forensics and Watermarking Volume 10431 || Detection of Face Morphing Attacks by Deep Learning. , 10.1007/978-3-319-64185-0(Chapter 9), 107–120.

[6]Ritu Agarwal1 & Om Prakash Verma. (2019). An efficient copy move forgery detection using deep learning feature extraction and matching algorithm. *Springer.*, pp.1-25.

[7]Sumaira Bibi1, Almas Abbasi1, IjazUl Haq2, Sung Wook Baik2, and Amin Ullah. (2021). Digital Image Forgery Detection Using Deep Autoencoder and CNN Features. *Research*. 11, pp.1-19

[8] SurbhiGupta;Neeraj Mohan and Priyanka Kaushal; (2021). Passive image forensics using universal techniques: a review . Artificial Intelligence Review,.

[9]KalyaniDhananjayKadam, Dr. Swati AhirraoAhirrao and Dr.KetanKotecha. (2020). Bibliometric Analysis of Passive Image Forgery Detection and Bibliometric Analysis of Passive Image Forgery Detection an. *ResearchGate*. , pp.1-31.

[10] Asghar, Khurshid; Habib, Zulfiqar and Hussain, Muhammad (2016). Copy-move and splicing image forgery detection and localization techniques: a review. Australian Journal of Forensic Sciences, 1–27.

[11] ParameswaranNampoothir and Sugitha N. (2021). Analysis of Copy Move Forgery Detection Process Using Fuzzy C Means Based DeepLearning Algorithm in Digital Image. Research Square, pp.1-26.

[12] Ivan Castillo Camacho and Kai Wang; (2021). A Comprehensive Review of Deep-Learning-Based Methods for Image Forensics . Journal of Imaging,.

[13] Ferreira, William D.; Ferreira, Cristiane B.R.; da Cruz JÃºnior, Gelson and Soares, Fabrizzio (2020). A review of digital image forensics. Computers & Electrical Engineering, 85, 106685.

[14] PeiyuZhuang;HaodongLi;ShunquanTan;Bin Li and Jiwu Huang; (2021). Image Tampering Localization Using a Dense Fully Convolutional Network . IEEE Transactions on Information Forensics and Security.

[15] GengyunJia;MeisongZheng;ChuanruiHu;XinMa;YutingXu;LuoqiLiu;Yafeng Deng and Ran He; (2021). Inconsistency-Aware Wavelet Dual-Branch Network for Face Forgery Detection . IEEE Transactions on Biometrics, Behavior, and Identity Science.

[16] Ferreira, Anselmo; Felipussi, Siovani C.; Alfaro, Carlos; Fonseca, Pablo; Vargas-Munoz, John E.; dos Santos, Jefersson A. and Rocha, Anderson (2016). Behavior Knowledge Space-Based Fusion for Copy-Move Forgery Detection. IEEE Transactions on Image Processing, 1–1.

[17]Neupane, D., &Seok, J. (2020). Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset With Deep Learning Approaches: A Review. IEEE Access, 1–1.

[18] Yang, J., Xiao, S., Li, A., Lan, G., & Wang, H. (2021). Detecting fake images by identifying potential texture difference. Future Generation Computer Systems, 125, 127–135.

[19] Gibert, Daniel; Mateu, Carles and Planes, Jordi (2020). The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. Journal of Network and Computer Applications,102526.

[20] Jiao, Licheng and Zhao, Jin (2019). A Survey on the New Generation of Deep Learning in Image Processing. IEEE Access, 7, 172231–172263.

[21] AlkaLeekha;ArpanGupta;Amit Kumar and Tarun Chaudhary; (2021). Methods of Detecting Image forgery using convolutional neural network . Journal of Physics: Conference Series.

[22] Mehta, R., Aggarwal, K., Koundal, D., Alhudhaif, A., &Polat, K. (2021). Markov features based DTCWS algorithm for online image forgery detection using ensemble classifier in the pandemic. Expert Systems with Applications, 185, 115630.

[23] Kasban, H. and Nassar, Sabry (2020). An efficient approach for forgery detection in digital images using Hilbertâ□“Huangâ□‹ transform. Applied Soft Computing, 97, 106728–.

[24] Diallo, Boubacar; Urruty, Thierry; Bourdon, Pascal and Fernandez-Maloigne, Christine (2020). Robust forgery detection for compressed images using CNN supervision. Forensic Science International: Reports, 2, 100112–.

[25] Zhang, Kejun; Liang, Yu; Zhang, Jianyi; Wang, Zhiqiang and Li, Xinxin (2019). No One Can Escape: A General Approach to Detect Tampered and Generated Image. IEEE Access, 7, 129494–129503.

[26] ZeYang;ZhenyuYu;YuyingLiang;RuiGuo and Zhihua Xiang; (2020). Computer Generated Colorized Image Forgery Detection using VLAD Encoding and SVM . 2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC).

[27] CASIA 2 Dataset. Retrieved from https://www.kaggle.com/datasets/divg07/casia-20-image-tampering-detection-dataset