WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

IMAGE FORGERY DETECTION

Harini M[1], Dr Chithra K[2]

[1] UG Student, Department of Computer Science, Sri Krishna Adithya College of Arts and Science, Coimbatore.
[2] Head of Department, Department of Computer Science, Sri Krishna Adithya College of Arts and Science, Coimbatore.

ABSTRACT :

Image forgery presents a significant challenge with potentially grave implications across multiple sectors. The application of deep learning techniques, particularly convolutional neural networks (CNNs), has demonstrated encouraging effectiveness in identifying such forgeries. CNNs are inherently well-suited for image-related tasks due to their proficiency in extracting pertinent features from visual data. The proposed methodology utilizes a CNN to derive features based on residual noise from images, facilitating the detection of forgeries. This approach focuses on recognizing the noise patterns that result from the forgery process, enabling a clear distinction between genuine and altered images. A notable benefit of employing CNNs for image forgery detection is their capacity to manage previously unseen forgeries. As forgery techniques evolve in complexity, traditional detection methods may struggle to keep pace. In contrast, CNNs possess the ability to learn and identify patterns that are not explicitly programmed, thus enabling them to uncover novel and unrecognized forms of forgery. In summary, leveraging CNNs for the detection of image forgery holds significant promise in addressing the issue of image manipulation. With continued research and advancement, this technology could substantially improve the reliability and credibility of digital images across various fields, including medical documentation and criminal investigations.

Keywords: Image Forgery, Error Level Analysis, Convolutional Neural Networks, Deep Learning, Forgery Detection of image, Peak Signal to Noise Ratio.

1. Introduction :

It is true that image forgery has become a prevalent issue in today's society, and the increasing availability of image processing tools has made it easier for anyone to manipulate and share images online. This has led to a need for more sophisticated techniques for detecting manipulated images. Machine learning, and specifically convolutional neural networks, have shown promise in this area. The Error Level Analysis (ELA) method and Peak Signal to Noise Ratio (PSNR) are commonly usedmanipulation and measuring the quality of compressed or reconstructed images. These methods can be applied to images to identify any discrepancies or alterations, which can then be flagged for further analysis.In the recognition process, each character in the image can be segmented and identified using machine learning algorithms. This can be particularly useful in forensic investigations or biomedical research, where accurate identification of images and their components is crucial. Overall, the use of machine learning and image processing techniques can help to improve the efficiency and accuracy of image forgery detection, and enable better law enforcement and forensic investigations.

2.Literature Survey :

Image Forgery Detection Using Recompressing Images, carried out by Syed Sadaf Ali [1] The techniques used are adapted to the individual needs, interests, and preferences of the user or society. Image compression involves reducing the pixels, size, or colour components of images in order to reduce the file size for forgery detection. Advanced image optimization techniques can detect the more important image components and discard the less vital ones. Image Forgery Detection by using Support Vector Machine developed by J.Malathi [2] Forgery detection technique that uses illuminant color inconsistency and machine learning classifiers such as Support Vector Machine (SVM). SVM is a supervised classification algorithm that is used to differentiate between two separate categories by drawing a line between them. In this technique, the illuminant color of input images is estimated, and illuminant maps are created for each image. Furthermore, all faces present in one image and corresponding faces of other individual images are extracted for investigation. However, it seems that this technique has some drawbacks, such as requiring clear textural and inclination highlighting and affecting the acknowledged substance of the image entirely.It is worth noting that there are several other forgery detection techniques available that use different approaches, such as image forensics, watermarking, and deep learning-based methods. Each technique has its advantages and limitations, and the selection of an appropriate technique depends on various factors such as the type of forgery, the available data, and the required level of accuracy. A Full-Image Full-Resolution End-to-End-Trainable CNN Framework for Image Forgery Detection, [6] carried out by F. Marra It proposes a framework for detecting image forgery using a convolutional neural network (CNN).The framework includes a feature extraction module and a classification module, both using CNNs, and it operates on full-resolution images.The dataset used is authentic and forged images, including various types of forgeries, to train and test the framework. It also proposes a data augmentation method to improve the framework's robustness. Statistical Features

based Optimized Technique for Copy Move Forgery Detection, carried off by S. B. G. T. Babu and C. S. Rao [8] The techniquesuggests a novel method for identifying copy-move forgeries in digital photos.The approach uses statistical features to represent the image and employs an optimized technique based on iterative voting to detect the forgeries.The suggested approach is evaluated using different benchmark datasets, and the findings reveal that it detects copy-move forgeries with high accuracy. Digital Image Forgery Detection Based on the Expectation Maximization Algorithm, Executed by M. H. Alkawaz [9] It proposes a new approach for detecting digital image forgeries using an expectation-maximization (EM) algorithm. The approach models and the probability distribution of the forgery and the original image were used to estimate the parameters of the distribution and detect the forgery. Image Forgery Detection Using Image Similarity, carried out by S. al-Zahir and R. Hammad [10] The approach compares the similarities between different regions of an image and uses a clustering algorithm to identify forged regions. The suggested approach was evaluated using multiple benchmark datasets, and the findings reveal that it detects image forgeries with high accuracy. Copy-move forgery detection based on keypoint clustering and similar neighborhood search algorithm, executed by H. Chen, X. Yang and Y. Lyu [12] The algorithm uses a clustering technique to group similar keypoints based on scale and color, and then matches them to identify tampered regions. To locate the tampered regions accurately, a novel localization algorithm is employed, which compares the close neighborhoods of matching pairs using two similarity measures and marks the tampered regions in the pixels of the images iteratively. Overall, this algorithm seems to be designed to identify tampered regions in images with high accuracy and efficiency.

Table 1 Summary of related work on Image Forgery Detection

Literature	Image forgery detection			
	DNN	SVM	GBI	COPY MOVE
J.Malathi, et al. 2019 [2]	YES	YES		
F. Matern, et al. 2020 [3]			YES	
Anushka Singh and Jyotsna Singh, 2022 [5]	YES			
S. B. G. T. Babu and C. S. Rao, 2020 [8]				YES
H. Chen, et al. 2020 [12]	YES			YES

Collectively, the summary of various techniques used as per the recent literature for image forgery detection as shown in Table 2

PAPER	Technique
Syed Sadaf Ali, et al. 2022 [1]	Recompression of Images
J.Malathi, et al. 2019 [2]	SVM
F. Matern, et al. 2020 [3]	Gradient-Based Illumination
Anushka Singh and Jyotsna Singh, 2022 [5]	ResNet
F. Marra, et al. 2020 [6]	ResNet
S. B. G. T. Babu and C. S. Rao, 2020 [8]	Copy-move
S. alZahir and R. Hammad, et al. 2020 [10]	Expectation Maximization

3. SYSTEM METHODOLOGY :

Methodology

ELA (Error Level Analysis) is a technique used for detecting image forgery. It involves compressing an image to a low quality, then re-saving it at a higher quality, and then calculating the difference between the two versions of the image. The resulting image is known as an ELA image, and it highlights the parts of the image that have been manipulated or edited. Convolutional neural networks (CNNs) are a popular choice for image forgery detection because they can learn to recognize patterns and features in images. The CNN is trained on a dataset of real and manipulated images to learn the characteristics of forged images. Once the CNN is trained, it can be used to classify new images as either real or manipulated. To implement an image forgery detection system using ELA and CNN, the following steps can be taken: Convert the input image into an ELA image. Preprocess the ELA image and prepare it for input into the CNN. Use the CNN to classify the ELA image as either real or manipulated. If the image is classified as manipulated, further analysis can be performed to determine the type of forgery that was used. It is important to note that while ELA can be a useful technique for detecting image forgery, it is not foolproof and can produce false positives or false negatives. Therefore, it is important to combine ELA with other techniques and methods for a more accurate and robust detection system.

B. System Architecture

The proposed system architecture for image fraud detection consists of several steps, starting with dataset preparation. The open image dataset's annotations are converted into a format accessible by the model during the training process. The testing process involves converting the image into an ELA image format, calculating the noise and signal ratio, denoising the image, and converting it to a black-and-white format. The model is split into two datasets using the train/test method, with 80% used for training and 20% used for testing the model. The CNN model is applied to high-scoring regions within the image that is considered forgeries. A confusion matrix technique is used to summarize the performance of the classification algorithm. A table is plotted of all the predicted and actual values of the classifier, and a confidence score is calculated as an evaluation standard. The confidence score represents the probability of the image being detected correctly by the algorithm and is given as a percentage. If the confidence score is not above a sufficient threshold (i.e., 0.9), it may be prudent to hold back from making decisions. By making fewer predictions, the model's accuracy can be significantly improved. Each label is assigned a numerical value called Confidence, while Predict is evaluating an Issue. Overall, the proposed system architecture appears to be a comprehensive approach to detecting image fraud, with multiple steps to prepare and test the model's accuracy. The use of a confusion matrix technique and confidence scores adds an extra layer of evaluation, ensuring that the algorithm'spredictions are reliable before making any decisions.

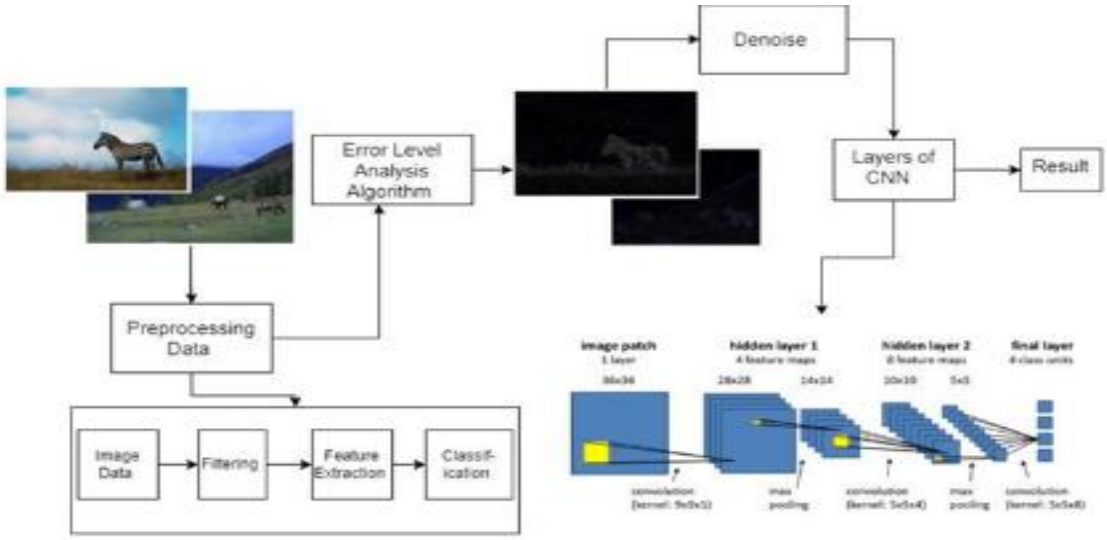


Fig.1 System architecture for Image Forgery Detection

C. Convolutional Neural Network

Convolutional Neural Networks (CNNs) have indeed become a popular tool for detecting forgery images. CNNs are a type of deep learning algorithm that can be trained to extract features from images and classify them into different categories. They are inspired by the human visual system and consist of multiple layers of interconnected neurons that perform convolution operations on the input image to extract features. One of the advantages of using CNNs for image forensics is their ability to detect subtle artifacts that may not be visible to the naked eye. For example, when an image is manipulated, such as by copy-pasting a fragment from one image to another, there may be slight variations in the pixel values or texture that are indicative of the manipulation. CNNs can learn to detect these differences and classify the image as either genuine or fake. Overall, CNNs have shown great promise in a variety of computer vision and image processing applications, including image forensics. With the increasing prevalence of digital manipulation in today's world, the ability to detect forged images has become more important than ever, and CNNs provide a powerful tool for this purpose. The input

layer in a convolutional neural network (CNN) is where the images from the dataset are fed. The images are usually in the form of 3-dimensional arrays, with the first two dimensions representing the height and width of the image (the number of pixels), and the third dimension representing the red, green, and blue (RGB) colors present in each pixel. In the feature-extraction part of the CNN architecture, the input image is passed through a series of convolutional layers, which apply a set of learnable filters to extract features from the image. Each filter produces a feature map that highlights a particular pattern or feature in the input image. These feature maps are then passed through activation functions like ReLU to introduce non-linearity and avoid the vanishing gradient problem. After the feature extraction process, the output of the last convolutional layer is flattened into a 1-dimensional vector and fed into a series of fully connected layers for classification. The fully connected layers use the extracted features to make predictions about the class of the input image. The final output layer usually employs the softmax function to generate a probability distribution over the classes, indicating the most likely class for the input image.

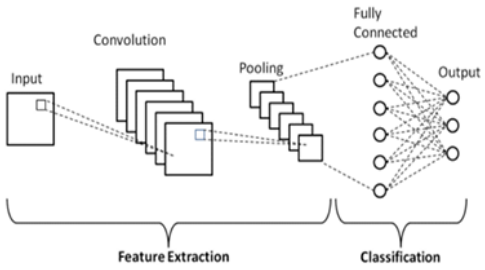


Fig. 2 CNN Architecture

That's a good summary of the main functions of each layer in a convolutional neural network. Here's a bit more detail on each layer: Convolutional layer: This layer applies a set of filters to the input image or feature map, generating a set of output feature maps. Each filter looks for a specific pattern or feature in the input, and the output feature maps highlight where those features are present in the input. By stacking multiple convolutional layers, the network can learn increasingly complex and abstract features. Pooling layer: This layer downsamples the output feature maps from the convolutional layer, reducing their spatial size and number of parameters. The most common type of pooling is max pooling, which takes the maximum value within a small region of the feature map. This helps to capture the most salient features while discarding redundant information, and also makes the network more robust to variations in input position and scale.

Fully-connected layer: This layer takes the flattened output of the previous layer and applies a set of weights to produce a vector of class probabilities. The weights are learned during training using backpropagation and gradient descent. The softmax activation function ensures that the output probabilities sum to 1, allowing the network to make a single prediction for the input image. The number of neurons in the fully-connected layer corresponds to the number of output classes.

D. Image Forgery Detection

The detection of fake images is done by convolutional neural networks. CNN helps in recognising tempered images, and it is mainly used to find tampered images accuracy. Image Processing: The image obtained undergoes the following steps: ELA conversion, grayscale conversion, thresholding, and calculation of confidence.

- ● The Error Level Analysis (ELA) method is one way to identify the areas of the image that have been modified. It works by creating a difference map of the image by compressing and decompressing it with a low-quality JPEG algorithm. The parts of the image that have been modified will have a different compression rate and will appear as bright spots in the difference map
- . ● Grayscale conversion is often used to simplify the image and reduce its complexity. It involves converting the image into a black-and-white or gray-scale image, where each pixel's value represents its intensity
- . ● Thresholding is a technique used to convert the grayscale image into a binary image, where each pixel is either black or white. This process helps eliminate noise in the image and can improve the accuracy of the detection algorithm.
- ● Confidence: A confidence score is a number between 0 and 1.The confidence value can be calculated for only one input, which gives the meaning of the algorithm confidence for this class

Fig.3 A CNN which acts a backbone for the model

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 124, 124, 32)      2432

 conv2d_1 (Conv2D)           (None, 120, 120, 32)      25632

 max_pooling2d (MaxPooling2D  (None, 60, 60, 32)       0
 )

 dropout (Dropout)           (None, 60, 60, 32)        0

 flatten (Flatten)           (None, 115200)            0

 dense (Dense)               (None, 256)               29491456

 dropout_1 (Dropout)         (None, 256)               0

 dense_1 (Dense)             (None, 2)                 514

=================================================================
Total params: 29,520,034
Trainable params: 29,520,034
Non-trainable params: 0
_________________________________________________________________
```

4. SYSTEM IMPLEMENTATION :

A. Software and Hardware

The system requirements for running an image forgery detection system using a CNN model implemented in Python 3.7.X (IDLE) and the CASIA dataset.It is recommended that the computer system has at least:

- ● RAM:8GBormore
- ● HardDiskDrive (HDD): 80GB or more
- ● Processor: i5 or higher

These system requirements are necessary to handle large amounts of data and the constant nature of the environment. It is important to note that the specific hardware and software requirements may vary depending on the size of the dataset and the complexity of the CNN model being used. Therefore, it is always a good idea to check the specific requirements of the software and datasets being used before implementing an image forgery detection system. Additionally, it is recommended to have sufficient cooling and power supply to ensure that the system can run smoothly and avoid any unexpected shutdowns or errors.

Dataset

The CASIA v2.0 database contains a total of 10,000 images, divided into two subsets: a training set of 5,000 images and a testing set of 5,000 images. Each subset includes eight categories of images: animal, architecture, article, character, nature, plant, scene, and texture. The images are in JPEG format and have a size of either 256 x 384 or 384 x 256 pixels.CASIA V2.0 dataset is used for image forgery detection.Two classes make up this dataset: actual photos and tampering detection. There are 7354 images, which are classified into real images and altered images in JPG format.

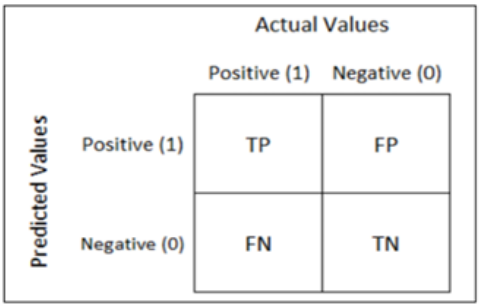
Dataset	Size	Categories	Format
CASIA V2.0	5 GB	8 categories of images	JPEG

Table 3 Details of CASIA Dataset

Fig.5. Images from the CASIA V2.0 Dataset

C. Confusion Metrics

A confusion matrix is a table that is commonly used to evaluate the performance of a classification algorithm by comparing the predicted labels to the true labels of a set of test data. The matrix displays the number of true positive, false positive, true negative, and false negative predictions made by the algorithm



The above table is a confusion matrix that summarizes the performance of a binary classification model, and it includes four possible outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). True positives occur when the model correctly predicts apositive outcome, and true negatives occur when the model correctly predicts a negative outcome. False positives occur when the model predicts a positive outcome, but the actual outcome is negative, and false negatives occur when the model predicts a negative outcome, but the actual outcome is positive.

5. RESULTS ANDANALYSIS :

The original image and its ELA-converted counterpart are shown in Fig. 7 and Fig. 8 of the dataset, respectively. And the fake image and its corresponding ELA-converted image are shown in Figs. 9 and 10, respectively. In Fig. 11, the red line represents the model's training loss and training accuracy, while the blue line represents the model's validation loss and validation accuracy. The model is iteratively trained and has an accuracy of 78.08%

Fig.7 Original image from dataset

Fig.8 ELA conversion of original image

Fig.9 Fake image from the dataset

Fig 10 ELA conversion of fake image

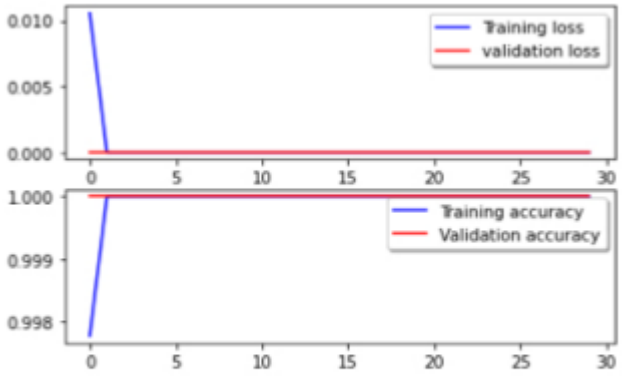


Fig. 11: Evaluation between training loss w.r.t. validation loss and training accuracy w.r.t. validation accuracy

5. CONCLUSION :

Image Forgery Detection System is developed and implemented using Convolutional Neural Networks for detecting manipulation in the images using the model over the CASIA V2.0 dataset. These images are converted into black-and-white format using the ELA method, then PSNR is applied to calculate noise and denoise the images, which are then passed to the detection system where recognition of manipulated images takes place. Once the forged images are recognized, they are displayed as output. Aconfusion matrix is used to evaluate performance, and the findings are displayed in a table that takes into account all of a classifier's anticipated and actual values. The confidence score is then computed as an evaluation standard. The model's accuracy after iterative training is 78.08%.

REFERENCES :

[1] Ali, S.S.; Ganapathi, I.I.; Vu,N.-S.; Ali, S.D.; Saxena, N.; Werghi, N., “Image Forgery Detection Using Deep Learning by Recompre-ssing Images,” Electronics 2022, 11, 403.

[2] J.Malathi, B.Narasimha Swamy, Ramgopal Musunuri, "Image Forgery Detection by using Machine Learning, International Journal of Innovative Technology and Exploring Engine-ering (IJITEE)ISSN: 2278-3075, Volume-8, Issue- 6S4, April 2019.

[3] F. Matern, C. Riess and M. Stamminger, "Gradient-Based Illumination Description for Image Forgery Detection," in IEEE Transac-tions on Information Forensics and Security, vol. 15, pp. 1303-1317, 2020, doi:10.1109/TIF S.2019.2935913.

[4] Z. J. Barad and M. M. Goswami, "Image Forgery Detection using Deep Learning: A Survey," 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS), 2020, pp. 571-576, doi: 10.1109/ICACCS48705.2020.9074408.

[5] Anushka Singh and Jyotsna Singh,“ Image forgery detection using Deep Neural Network,” Conference: 2021 8th InternationalConference on Signal Processing and Integrated Networks (SPIN)At: New Delhi, January 2022 DOI:10.1109 SPIN525336.2021 .9565953.

[6] F. Marra, D. Gragnaniello, L. Verdoliva and G. Poggi,"A Full-Image Full-Resolution End-to-End-Trainable CNN Framework for Image Forgery Detection," in IEEE Access, vol. 8, pp. 133488-133502, 2020, doi:10.1109/ACCE SS.2020.3009877.

[7] R. Agarwal, D. Khudaniya, A. Gupta and K. Grover, "Image Forgery Detection and Deep Learning Techniques: A Review," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 1096-1100, doi: 10.1109/ICICCS48265.2020.9121083.

[8] S. B. G. T. Babu and C. S. Rao, "Statistical Features based Optimized Technique for Copy Move Forgery Detection," 2020 11th Inter national Conference on Computing, Communi-cation and Networking Technology (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225426.

[9] M. H. Alkawaz, M. T. Veeran and R. Bachok, "Digital Image Forgery Detection based on Expectation Maximization Algorithm," 2020 16th IEEE International Colloquium on Signal Processing and Its Applications (CSPA), Langkawi, Malaysia, 2020, pp. 102-105, doi: 10.1109/CSPA48992.2020.9068731.

[10] alZahir, S., Hammad, R. Image forgery detection using image similarity. Multimed Tools Appl 79, 28643–28659 (2020).

[11] K. M. Hosny, A. M. Mortda, M. M. Fouda and N. A. Lashin, "An Efficient CNN Model to Detect Copy-Move Image Forgery," in IEEE Access, vol. 10, pp. 48622-48632, 2022, doi: 10.1109/ACCESS.2022.3172273.

[12] H. Chen, X. Yang and Y. Lyu, "Copy-Move Forgery Detection Based on Keypoint Clustering and Similar Neighborhood Search Algorithm," in IEEE Access, vol. 8, pp. 36863-36875,2020,doi:10.1109/ACCESS.202 0.2974804.