Image Forgery Detection Using Deep Learning

Prof. D. D. Pukale¹, Prof. V. D. Kulkarni², Julekha Bagwan³, Pranali Jagadale⁴, Sanjivani More⁵, Renuka Sarmokdam⁶

¹HoD, Department of Computer Engineering, Bharati Vidyapeeth’s College of Engineering for Women’s, Dhanawadi, India.
²Assistant Professor, Bharati Vidyapeeth’s College of Engineering for Women’s, Dhanawadi, India.
³, ⁴, ⁵, ⁶B.E. Computer Engineering Students of Bharati Vidyapeeth’s College of Engineering for Women’s, Dhanawadi, India.

Emails: dhondiram.pukale@bharatividyapeeth.edu¹, vinaya.kulkarni@bharatividyapeeth.edu², julekhabagwan5402@gmail.com³, pranujagadale11@gmail.com⁴, sanjivani.s.more@gmail.com⁵, renuka9688@gmail.com⁶

Abstract
Image forgery is a big problem in digital media, making it important to have strong detection methods to fight misinformation and keep trust in visual content. In this project, we introduce an advanced image forgery detection system using VGG16, a powerful convolutional neural network, and Error Level Analysis (ELA) algorithms. Our goal is to create an efficient and accurate system that can identify real images from fake ones, especially focusing on detecting splicing and copy-move forgeries. By examining pixel intensities and patterns, our system can accurately find tampered areas, improving the integrity and trustworthiness of digital images. We use a diverse dataset of real and fake images from different sources to train and test the VGG16-ELA model. We aim to find the percentage of forgery, highlighting the forged areas and generating the mask of forged area. Through this effort, we aim to increase trust in visual content in fields like forensics, journalism, and social media, helping to ensure the reliability of digital information.

Keywords: Convolutional Neural Network; Copy Move Forgery; Deep Learning; Error Level Analysis (ELA); Forensic Analysis; Image Forgery Detection; Splicing Forgery; VGG16.

1. Introduction
The advent of deep learning methodologies has ushered in a new era of innovation in image forgery detection, with VGG16 and Error Level Analysis (ELA) algorithms emerging as pivotal tools in this field. Image forgery, characterized by the deliberate alteration or manipulation of visual content, poses significant challenges in various domains, including forensics, media authentication, and digital content integrity. Detecting forged images and localizing tampered areas are paramount for preserving the credibility and authenticity of digital media in today's information-driven society. In recent years, the deep learning community has witnessed remarkable progress, fueled by advances in convolutional neural networks (CNNs) such as VGG16. VGG16, a deep convolutional neural network architecture, has demonstrated exceptional performance in image classification tasks by virtue of its deep hierarchical structure and large receptive fields. With its ability to learn rich hierarchical representations directly from raw pixel data, VGG16 is poised to revolutionize image forgery detection by extracting intricate features indicative of manipulation. Complementing the prowess of VGG16, Error Level Analysis (ELA) algorithms offer invaluable insights into image authenticity. ELA exploits variations in error levels introduced during image compression to identify regions potentially subjected to alterations. By analyzing discrepancies in error levels across different areas of an image, ELA can pinpoint suspicious regions and anomalies, providing critical clues for detecting image forgeries[1-7]. The objective of this project is to leverage the synergies between VGG16 and ELA algorithms to develop an
advanced image forgery detection system. Our system aims to detect a wide spectrum of forgery techniques, including splicing and copy-move, by integrating cutting-edge deep learning methodologies with ELA-based analysis. By enhancing the accuracy, sensitivity, and specificity of forgery detection, we empower users to identify and mitigate the proliferation of manipulated images effectively. Moreover, a central focus of this project involves the development of a user-friendly graphical user interface (GUI) that simplifies the process of uploading, analyzing, and interpreting image data. The intuitive interface streamlines interaction with the forgery detection system, making it accessible to users with diverse levels of technical proficiency. Through user-centric design principles, we endeavor to democratize forgery detection and foster widespread adoption of reliable tools for combating digital misinformation.

2. Method

This section details the methodology employed in developing our robust image forgery detection system. The process is outlined chronologically, including the research design, research procedure and data acquisition.

2.1. System Architecture

The system architecture depicts the overall comprehensive process for detecting forged images using a combination of image preprocessing techniques, convolutional neural networks (CNN), error level analysis (ELA), and post-processing methods. Following Figure 1 shows a simplified system architecture diagram for image forgery detection using CNN [8-12]. The process starts with utilizing an image dataset alongside a test image. The images within the dataset are subjected to various preprocessing steps such as resizing, normalization, and conversion to grayscale.

After this preprocessing, the images undergo ELA (Error Level Analysis) to detect potential tampering or manipulation by analyzing the compression levels within the images. The results of the ELA, which indicate error levels, are then used as input features for training a CNN model to classify images as either unaltered or forged. Similarly, the test image undergoes a comparable preprocessing pipeline involving resizing, normalization, and conversion to grayscale. Subsequently, the preprocessed test image is analyzed using ELA to produce a heatmap that highlights regions with potential forgery or manipulation. The test image’s ELA result undergoes further analysis using post-processing techniques, which involves identifying the forged area within the image, determining the percentage of forgery present, and creating an ELA mask that visually displays the potentially manipulated regions. The comprehensive assessment of whether the test image has been forged or not is derived from combining the classification...
result from the CNN (pristine or forged) and the post-processing analysis. Moreover, visual aids such as the highlighted forged area, the percentage of forgery, and the ELA mask are provided to facilitate forgery detection and localization. This process leverages the strengths of both deep learning (CNN) and traditional image forensics (ELA) to accurately identify forged images and provide detailed information about the nature and extent of the forgery.

2.2. Proposed Methodology

The identification of forgeries involves a comprehensive dataset comprising images from CASIA 1.0 and CASIA 2.0, encompassing a variety of forgeries such as copy-move and spliced images. Initially, the dataset goes through pre-processing steps to standardize the images, which involve resizing them to a consistent size, normalizing pixel values for uniformity, and converting them to grayscale for simplified analysis. After pre-processing, the dataset images undergo Error Level Analysis (ELA) [13-17]. ELA is a critical step in identifying potential tampering by examining the variations in compression levels across different areas of the images. This analysis provides valuable insights into areas where forgeries may have occurred based on differences in compression artifacts. The preprocessed dataset is used to train a VGG16 deep learning model concurrently. The VGG16 model aims to differentiate between authentic and manipulated images by utilizing the features extracted during ELA and other pre-processing stages. Throughout the training process, the model's parameters are fine-tuned to effectively classify images as genuine or manipulated, considering the intricacies inherent in copy-move and spliced forgeries. Once the VGG16 model is trained, it is ready for inference on new, unseen images, including the test image in question. Similar preprocessing steps are applied to the test image, preparing it for ELA analysis. The ELA results provide a heatmap highlighting suspicious areas that may indicate forgery or manipulation within the test image. The preprocessed test image and its corresponding ELA heatmap are input for the VGG16 model, which then conducts classification to identify if the image is authentic or altered. If the classification indicates the image is altered, further post-processing steps are executed (Figure 2a & Figure 2b). These steps include recognizing and emphasizing the altered areas within the image, calculating the proportion of forgery present based on the detected manipulation, and producing an ELA mask to display the potentially modified areas. The final output of the forgery detection process includes the classification label (pristine or forged), if forged then highlighting the forged areas, the percentage of forgery detected, and the ELA mask illustrating the suspicious regions.

Figure 2 (a) Image

Figure 2 (b) Generated ELA Image

This comprehensive methodology integrates deep learning techniques with traditional image forensics, enabling the model to detect various types of
forgeries with accuracy and provide detailed insights into the nature and extent of the detected manipulations. **ELA:** Error Level Analysis (ELA) is a method in digital image forensics that is utilized to identify possible instances of manipulation or forgery in images. It operates by examining the changes in compression levels that happen when an image is saved or edited repeatedly. ELA points out areas with notable differences in error levels, indicating potential areas where modifications like copy-move or splicing may have taken place [18-21]. This approach is valuable for pinpointing questionable regions within images and is frequently employed as a preliminary stage in forgery detection algorithms, offering insights into potentially altered areas for further examination.

### 2.3. Pre-Processing
In order to prepare the dataset for training the deep learning model for image forgery detection, it is crucial to have a thorough pre-processing pipeline. The dataset contains images with genuine and altered samples, and the pre-processing steps are aimed at improving image quality and extracting relevant features to accurately detect and localize forged areas. The main objective of pre-processing is to improve image quality by eliminating or minimizing irrelevant and extra elements in the background of the images. The application of filters helps to eliminate noise and high-frequency components that could impact detection accuracy. Furthermore, improving image contrast, brightness, and sharpness highlights important features and enhances visibility. Techniques like histogram equalization or adaptive histogram equalization can be utilized to enhance image contrast.

#### 2.3.1. Image-Resizing
Resizing an image involves altering its width and height to achieve consistency, particularly when working with datasets containing images of different sizes. The aim is usually to scale down larger images to match the dimensions of smaller ones, thus conserving computational resources and processing time. In our case, the CASIA v1.0 and CASIA v2.0 datasets feature images of various sizes, so we are standardizing the images to a constant size of 128 X 128. Various interpolation methods or techniques are applied, enabling the estimation of new pixel values based on surrounding known pixels, resulting in a smooth and coherent appearance of the modified image.

### 2.3.2. Normalization
Image normalization is a common preprocessing step in many computer vision and image processing tasks. It involves transforming the pixel values of an image to a common scale or range, typically between 0 and 1 or -1 and 1. The main purpose of normalization is to ensure that the pixel values are within a consistent range, which can improve the performance and convergence of various algorithms that operate on the image data. **Min-Max Normalization:** This technique linearly scales the pixel values to a specified range, usually [0, 1] or [-1, 1]. The formula for min-max normalization is - x_norm = (x - min(x)) / (max(x) - min(x)) Where x is the original pixel value, min(x) and max(x) are the minimum and maximum pixel values in the image, respectively, and x_norm is the normalized pixel value.

### 2.3.3. Greyscale Conversion
Grayscale conversion is a fundamental operation in image processing, where a color image is converted into a grayscale (or monochrome) image. This process involves reducing the color information of each pixel to a single intensity value, representing the brightness or luminance of that pixel. Grayscale images are often used in computer vision and image analysis tasks, as they simplify the data and can enhance certain features or patterns in the image.

### 2.4. Dataset
In this project, we leveraged the renowned CASIA 1.0 and CASIA 2.0 datasets for image forgery detection. The CASIA datasets are widely regarded in the field of image forensics, offering a rich collection of authentic and manipulated images essential for training and evaluating forgery detection algorithms. CASIA 1.0 provided a solid foundation with its diverse forgery types such as copy-move and splicing, serving as a benchmark for our initial model development. As we progressed, incorporating CASIA 2.0 enriched our dataset with more complex forgery scenarios, enabling us to enhance the robustness and accuracy of our forgery detection system. CASIA 1.0: - CASIA 1.0 contains a diverse
collection of digital images, including both authentic (pristine) images and images that have been manipulated or forged. The dataset covers various types of forgeries, such as copy-move, splicing, and other digital manipulations (Figure 3a & Figure 3b). CASIA 1.0 was created to facilitate research and development in image forgery detection and digital forensics. It serves as a benchmark dataset for evaluating the performance of forgery detection algorithms. The images in CASIA 1.0 exhibit different levels of complexity in terms of forgery techniques, making it suitable for testing the robustness and accuracy of forgery detection models. CASIA 2.0: CASIA 2.0 is an extension and enhancement of the CASIA 1.0 dataset, offering a larger and more diverse collection of images. Compared to CASIA 1.0, CASIA 2.0 includes a broader range of forgery types and scenarios, including more sophisticated manipulations and realistic forgery challenges. Researchers and practitioners often use CASIA 2.0 to validate and improve forgery detection algorithms, as it provides a more comprehensive and challenging dataset for testing algorithm performance. Similar to CASIA 1.0, CASIA 2.0 serves as a benchmark dataset in the field of image forensics, enabling comparative studies and advancements in forgery detection techniques.

3. Results and Discussion

In this section, we present the results obtained from training three deep learning models CNN and VGG16 on a combination of CASIA1 and CASIA2 datasets for the task of image forgery detection. We provide a comprehensive analysis of the training and validation performance of each model, including accuracy metrics, loss curves, and confusion matrices. Additionally, we discuss the implications of these results and draw insights into the effectiveness of each model for detecting forged images.

3.1. Model Training and Validation

We first trained the CNN and VGG16 models using a dataset comprising authentic and forged images extracted from the CASIA1 and CASIA2 datasets. The training process involved optimizing the models' parameters to minimize the binary cross-entropy loss function while maximizing classification accuracy. The dataset was split into training and validation sets, with a portion of the data reserved for model evaluation. The training metrics, including loss and accuracy, for each model are visualized below. These plots provide insights into the training progress and performance of each model throughout the training process. The training and validation accuracy curves for VGG16 are depicted in Figure 4 & Figure 5. The model exhibits a steady increase in both training and validation accuracy, reaching a peak accuracy of 98.29% on the validation set. So, with a learning rate
of 0.0001 and trained over 100 epochs, the confusion matrix reveals a training accuracy of 97.14% for VGG16. Figure 6 illustrates the training and validation loss curves for the CNN model. Similar to VGG16, the model shows a decreasing trend in both training and validation loss with slight fluctuations. Also the threshold for both models was kept as 50%, i.e. if model predicts accuracy of authenticity as greater than 50%, then it will be labelled as forged else authentic.

The training and validation accuracy curves for CNN are presented in Figure 7. Despite some variations, the model achieves a notable accuracy of 88.16% on the validation set. 

### 3.2. Confusion Matrix

The confusion matrix below offers a detailed insight into the model's performance, showing the counts of true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), Figure 8. In the context of image forgery detection, TP represents the number of correctly identified forged images, FP indicates the count of authentic images misclassified as forged, TN reflects the accurate identification of authentic images, and FN signifies the forged images incorrectly labeled as authentic [22-25]. These metrics illustrate the accuracy of identifying authentic and forged images, along with any misclassifications.
Precision, a measure of the model's exactness in predicting positive instances, is calculated as the ratio of TP to the sum of TP and FP. In our study, precision quantifies the accuracy of the VGG16 model in correctly identifying forged images among all images classified as forged. With an overall model accuracy of 97.14%, the precision of the model is 97.50%. For the CNN, model has achieved accuracy of 88.16% and 85.94% as precision, Figure 9. Recall, also known as sensitivity, measures the model's ability to identify all positive instances, calculated as the ratio of TP to the sum of TP and FN. For our VGG16 model, recall indicates its effectiveness in capturing all forged images from the entire dataset of forged images. The recall for the VGG16 model is 96.69% and for the CNN is 90.91%. The F1 score, a harmonic mean of precision and recall, provides a balanced assessment of the model's performance. It considers both false positives and false negatives and is calculated as \(2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}\). In our analysis of the VGG16 model, the F1 score synthesizes precision and recall into a single metric, offering insight into the overall effectiveness of the model in detecting forged images while minimizing misclassifications, Figure 10 & Figure 11. The F1 score for the VGG16 model is 97.09% and for the CNN is 88.35%, and the outcomes are shown in Figure 12, Figure 13 & Figure 14.

### Table 1: Comparative Analysis of Accuracy Metrics of Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
<th>Precision</th>
<th>Recall</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Base Paper [25]</td>
<td>71.6%</td>
<td>75.69%</td>
<td>72.59%</td>
<td>74.09%</td>
</tr>
<tr>
<td>2. CNN</td>
<td>88.16%</td>
<td>85.94%</td>
<td>90.91%</td>
<td>88.35%</td>
</tr>
<tr>
<td>3. VGG16</td>
<td>97.14%</td>
<td>97.50%</td>
<td>96.69%</td>
<td>97.09%</td>
</tr>
</tbody>
</table>
3.4. Outcome Analysis

Figure 12 Original Image (without any forgery)

Figure 13 The Image with Forgery Detecting as a Forged Image

Figure 14 Highlighting the Forged Area Along with the Percentage of Forgery

3.5. Applications

1. Device Detection: Identifying forged images that have been altered to appear as if they were taken from a specific device.
2. Copy-Moving Detection: Detecting images where parts have been moved from one location to another within the same image.
3. Fake Colorization Detection: Identifying images that have been artificially colorized.
4. JPEG Post-Processing Detection: Detecting images that have undergone generic contrast adjustments after JPEG compression.
5. Contrast Enhancement Detection: Identifying images where contrast has been artificially enhanced.
6. Image Manipulation Detection: Detecting images that have been manipulated at the pixel or image level.
7. Industrial Object Detection: Identifying forged images in industrial settings, such as those created through physics-based rendering.

3.6. End Users

1. Forensic Scientists: Utilize deep learning for detecting forgeries in legal cases and investigations.
2. Security Companies: Implement deep learning-based systems to ensure the integrity of digital images and prevent misinformation.
3. Copyright Holders: Protect their intellectual property by detecting unauthorized use or alteration of their images.
4. Social Media Platforms: Use deep learning to monitor and remove fake images from their platforms.

3.7. Advantages

1. High Accuracy: Deep learning models can achieve high accuracy in detecting forgeries, especially when trained on large datasets.
2. Automation: Reduces the need for manual intervention, saving time and resources.
3. Adaptability: Can adapt to different types of image forgeries and processing methods.
4. Scalability: Capable of processing large volumes of images efficiently.
3.8. Disadvantages

1. Dependence on Dataset Selection: The effectiveness of deep learning models heavily depends on the selection and utilization of datasets.

2. Cumbersome Algorithms and Wrong Classifiers: Many models suffer from cumbersome algorithms and incorrect classifiers, leading to poor performance.

3. Extra Time Consumption and Expensiveness: The development and training of deep learning models can be time-consuming and costly.

3.9. Future Scope

1. Uncovering Deepfakes and AI-Generated Content: With the rise of deepfake technology and AI-generated content, there is a growing need to develop specialized models capable of distinguishing between authentic and manipulated media, including videos, audio, and images.

2. Privacy Protection: Integrating forgery detection mechanisms that respect and preserve individual’s privacy by ensuring sensitive information is not exposed during the analysis of images.

3. Deployment in Various Industries: Expanding the adoption of forgery detection systems in industries beyond traditional media, such as healthcare (medical imaging authenticity), banking (fraud detection), and e-commerce (product authentication).

Conclusion

In conclusion, deep learning has shown remarkable promise in detecting image forgeries, including DeepFake content, by learning complex patterns from large datasets. Image forgery detection using deep learning has advanced significantly, improving the integrity of visual content. Our project, combining Error Level Analysis (ELA) with CNN and VGG16 models, has surpassed previous results. The base paper [25] achieved 71.6% accuracy and 75.69% precision. In contrast, our CNN model achieved 88.16% accuracy and 85.94% precision, while the VGG16 model reached 97.14% accuracy and 97.50% precision. This robust system effectively identifies, calculates the percentage of forged area and highlights the forged region, demonstrating superior performance metrics for both copy-move and splicing forgery.

Acknowledgements

We would like to express our sincere gratitude to our Principal Prof. Dr. P.V. Jadhav for their constant support and guidance, our project guides, Prof. D. D. Pukale and Prof. V. D. Kulkarni, for their invaluable support and guidance throughout the development of our Image Forgery Detection using Deep Learning project. Their expertise and encouragement have been instrumental in our progress. We also thank the faculty of Bharati Vidyapeeth’s College of Engineering for Women, Pune for providing the necessary resources and a supportive environment. Special thanks to our peers for their constructive feedback. Thank you all for your contributions, which have been crucial to the success of our project.

References


