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

IMAGE FORGERY DETECTION AND LOCALIZATION USING DEEP LEARNING

[1]Prof. D.D. Pukale,

[1]H.O.D. Department of Computer Engineering, BVCOEW, Pune

[2]Prof. V.D. Kulkarni,

[2]Assistant Professor, Department of Computer Engineering, BVCOEW, Pune

[3]Julekha Bagwan, [4]Pranali Jagadale, [5]Sanjivani More, [6]Renuka Sarmokdam

BE students at BVCOEW, Pune,

Department of Computer Engineering,

Bharati Vidyapeeth's College of Engineering for Women, Pune, India.

Abstract: The proliferation of digital images across various platforms accentuates the pressing need for robust techniques in detecting image forgeries, especially passive forgery methods such as Copy-move and Splicing. This study presents a novel approach employing deep learning methodologies for accurate identification of manipulated regions within images, specifically targeting these passive forgery techniques. Our proposed framework harnesses the power of Mobile Net and ResNet as feature extraction networks, extracting rich representations from image data to capture intricate patterns indicative of potential tampering, including Copy-move and Splicing manipulations. Additionally, we utilize Mask R-CNN, a state-of-the-art instance segmentation model, for precise localization and quantification of forged regions, enabling the calculation of the percentage of altered content within images.

Index Terms **- Copy-Move Forgery, Splicing Forgery, ResNet, Mobile Net V2, Mask R-CNN, Percentage calculation, feature extraction, Localization, Digital Forensics.**

I. INTRODUCTION

The convergence of Deep Learning has indeed propelled advancements in the realms of digital forensics and cybersecurity. Notably, with the proliferation of various image editing applications and manipulation tools, tampering with images has become an effortless task. Detecting the disparity between a tampered image and its original counterpart is exceedingly challenging for the unaided human eye. Within the spectrum of image forgery and manipulation techniques, two primary categories emerge: Active and Passive methods. Active methods encompass techniques such as Watermarks and Digital Signatures, where concealed information assumes the form of a digital signature. On the other hand, Passive methods encompass techniques like Copy-Move Forgery, Splicing, and Retouching. In our paper, we focus primarily on Copy-Move Forgery and Splicing Techniques, aiming to delve deeper into these areas.

Our proposed technique places emphasis on localization, aiming to precisely identify forged regions within an image. Furthermore, our research endeavors to provide a forged percentage score for specific regions within an image, thereby enhancing the precision and granularity of forgery detection. In the landscape of image manipulation, Copy-Move Forgery involves duplicating a section of an image and pasting it elsewhere within the same image, while Splicing Technique entails merging different sections from multiple images to create a fabricated composite. Detecting such manipulations demands sophisticated algorithms and techniques capable of discerning subtle inconsistencies within the image data.

Our proposed solution consists of the usage of ResNet (Residual Network) and MobileNet V2 for feature extraction of the images. ResNet and MobileNet V2 can be utilized to extract relevant features from images, such as textures, edges, and structural elements, which can then be fed into subsequent stages of the forgery detection pipeline.

Mask R-CNN is a state-of-the-art deep learning model used for instance segmentation and object detection. Building upon the Faster R-CNN framework, Mask R-CNN extends it by adding a branch for predicting segmentation masks on each Region of Interest (RoI). This model not only identifies objects but also provides pixel-level segmentation masks, enabling precise localization of object boundaries within the image. In the context of Forgery Detection Mask R-CNN, with its ability to provide pixel-level segmentation masks, is valuable for precisely localizing regions within an image that have been tampered with or forged. It can identify the boundaries and extent of manipulated regions, aiding in the forensic analysis of tampered images.

Leveraging ResNet and MobileNet V2 for feature extraction combined with Mask R-CNN for accurate localization provides a robust framework for detecting and analyzing forged areas within images, enhancing the capabilities of image forensics, and

ensuring the integrity of visual content. The percentage calculated serves as an estimation of the extent of forgery within the image and aids in quantifying the level of tampering. Within the experimental domain, meticulous attention will be given to data pre-processing, model training, and the rigorous evaluation of results. Performance metrics, encompassing sensitivity, specificity, and overall accuracy, will be scrutinized to quantify the efficiency of our proposed approach.

II. Related Work

This literature review aims to delve deeper into the contemporary methodologies, their performance metrics, and the ongoing research trends in image forgery detection, with a focus to contribute to the understanding and advancement of reliable forgery detection mechanisms.

A hybrid method based on color illumination, DCNN, and semantic segmentation was developed by Abhishek and Jindal [1] to detect and localize copy-move and splicing forgery in images. The proposed algorithm consists of three steps. The proposed solution is to use a hybrid technique of color illumination, deep convolution neural network, and semantic segmentation to detect and localize image forgeries. The future work of this paper includes improving the accuracy and robustness of the proposed method by using more advanced deep learning models and techniques, such as attention mechanisms, adversarial learning, and self-supervised learning. Extending the proposed method to other types of image forgeries, such as face swapping, object removal, and inpainting, and evaluate its performance on larger and more diverse datasets. Exploring the applications of the proposed method in other domains, such as video forgery detection, digital forensics, and multimedia security.

Another system proposed by Jwaid and Baraskar consists the use of a novel technique for image splicing forgery detection based on local binary pattern (LBP) and discrete wavelet transform (DWT) [2]. The technique first converts the input image into YCbCr color channel, and then divides the chrominance component into non-overlapping blocks. Then, LBP operator is applied to each block, and wavelet transform is used to extract the features. Finally, principal component analysis (PCA) is used to reduce the dimensionality of the features, and support vector machine (SVM) is used to classify the blocks as authentic or forged. The paper evaluates the performance of the proposed technique on two public datasets: CASIA and Columbia, and compares it with other existing methods. The paper suggests some directions for future research, such as improving the processing time and complexity of the proposed technique, extending it to detect other types of image forgery, such as inpainting and retouching, and incorporating other features and classifiers to enhance the accuracy and robustness of the technique.

The system proposed by [3] presents a lightweight deep learning model based on Mask R-CNN with MobileNet V1 to detect and identify copy move and image splicing forgeries in digital images. The model also provides a forged percentage score for a region in an image. The model is evaluated on seven standard datasets and compared with ResNet-1014. The paper suggests some directions for future work, such as extending the model to handle other types of image forgeries, such as inpainting and removal, improving the accuracy and robustness of the model against various attacks and noise, and developing a web or mobile application for image forgery detection and identification.

The research proposed in [4] introduces a new image forgery detection method based on Discrete Cosine Transformation (DCT) and Local Binary Pattern (LBP) and a new feature extraction method using the mean operator1. The method is robust against low availability of forged training samples, rotation, scaling, and translation of images. The method is also applicable to both grayscale and color images, and outperforms existing methods on four benchmark datasets and a newly created IoT dataset. The paper presents the following steps for the proposed algorithm: Convert color images into grayscale and YCbCr color space images. Divide the images into non-overlapping fixed size blocks. Apply 2D-DCT on each block to obtain the DCT coefficients. Apply LBP on the magnitude of the DCT coefficients to enhance the forgery artifacts. Divide the LBP array into the same size of blocks as before. Calculate the mean value of each cell across all LBP blocks to obtain the feature vector. Use SVM classifier to distinguish authentic and forged images. The paper suggests some possible directions for future work, such as: Extending the proposed method to detect other types of image forgery, such as inpainting and splicing with different backgrounds. Developing a more comprehensive IoT image forgery dataset with different types of sensors and scenarios. Exploring other feature extraction and classification methods to improve the detection performance and efficiency. Investigating the impact of different block sizes and LBP parameters on the detection accuracy.

The paper [5] proposes a method based on convolutional neural network with global average pooling for splicing and copy-move tampering detection. The proposed method outperformed some state-of-the-art methods in experiments on three public image tampering datasets. Future work could focus on improving the performance of the proposed method on more complex datasets and exploring the application of the proposed method in other areas of image forensics.

The paper reviews the existing methods for forgery detection and their limitations, and motivates the use of a deep learning approach, specifically a Convolutional Neural Network (CNN) model, to detect both types of forgeries without knowing their types beforehand [6]. The paper proposes a CNN model that consists of three main phases: image pre-processing, feature extraction, and classification. The image pre-processing phase involves resizing the images and converting them into Error Level Analysis images. The feature extraction phase consists of convolution layers, pooling layers, and Rectified Linear Units (ReLU) layers that learn to extract features from the images. The classification phase consists of fully connected layers that map the extracted features to the final output, which indicates whether the image is original or forged. The paper explains the details of each phase and the hyperparameters used for the CNN model. Some future work of the paper includes evaluating the performance of the CNN model under different attacks such as compression, noise, filtering, scaling, and rotation. The CNN model could be enhanced to be more resilient to these attacks and preserve the accuracy of forgery detection.

The paper [7] reviews the state-of-the-art techniques of deep learning for copy-move image forgery detection (CMFD), which is a common type of image manipulation that involves copying and pasting a part of an image to another location in the same image. The paper discusses the challenges and limitations of the conventional methods for CMFD, such as block-based and keypoint-based

approaches, and highlights the advantages and potential of deep learning methods, such as convolutional neural networks (CNNs), for CMFD. The paper proposes [7] a novel deep learning architecture for CMFD, which consists of three main components: i) a pre-processing module that converts the input image into grayscale and applies data augmentation techniques; ii) a feature extraction module that uses a CNN to learn hierarchical features from the image patches; and iii) a classification module that uses a support vector machine (SVM) to classify the image patches as original or forged. The paper also introduces a filter layer in the CNN to suppress the image content and enhance the tampering traces. The paper suggests some possible directions for future research on CMFD using deep learning methods, such as: i) exploring other deep learning models, such as recurrent neural networks (RNNs) and generative adversarial networks (GANs), for CMFD; ii) developing more robust and efficient methods for dealing with complex tampering operations, such as rotation, scaling, and blurring; iii) creating more realistic and diverse datasets for CMFD; and iv) integrating CMFD with other image forensics tasks, such as source identification and tampering localization.

The proposed method [8] uses multiscale to ceck if there is any counterfeit in the image. By applying one-level Discrete Wavelet Transform, the sharped edges, which are traces of cut-paste manipulation, are high frequencies and detected from LH, HL and HH sub-bands. A threshold is proposed to filter the suspicious edges and the morphological operation is applied to reconstruct the boundaries of forged regions. If there is no shape produced by dilation or no highlight sharped edges, the image is not faked. In case of forgery image, if a region at the other position is like the defined region in the image, a copy-move is confirmed. If not, a splicing is detected. The suspicious region is extracted the feature using Run Difference Method (RDM) and a feature vector is created. Searching regions having the same feature vector is called detection phase.The proposed architecture of the algorithm is simulated in Matlab with high efficiency not only in the copy-move or spliced images but also the image with both copy-move and splicing. The authors have also suggested that the proposed method can be extended to detect forgery in videos.

The paper by Yuan Rao and Jiangqun Ni [9] presents a new image forgery detection method based on deep learning technique. The proposed method utilizes a convolutional neural network (CNN) to automatically learn hierarchical representations from the input RGB color images. The proposed CNN is specifically designed for image splicing and copy-move detection applications. The weights at the first layer of the network are initialized with the basic high-pass filter set used in calculation of residual maps in spatial rich model (SRM), which serves as a regularizer to efficiently suppress the effect of image contents and capture the subtle artifacts introduced by the tampering operations. The pre-trained CNN is used as patch descriptor to extract dense features from the test images, and a feature fusion technique is then explored to obtain the final discriminative features for SVM classification. The experimental results on several public datasets show that the proposed CNN based model outperforms some state-of-the-art methods.The proposed method could be extended to detect other types of image forgeries such as image retouching and image splicing with different resolutions. The proposed method could also be applied to other types of media such as videos and audio files. Additionally, the proposed method could be improved by using more advanced deep learning techniques such as Generative Adversarial Networks (GANs) and Recurrent Neural Networks (RNNs). These techniques could help improve the accuracy of the proposed method and make it more robust to different types of image forgeries.

The paper [10] focuses on detecting image manipulation and tampering and the localization of tampered areas. The authors use the core idea of end-to-end training of u-net network to further optimize the performance of image tampering detection tasks in combination with residual network. The proposed architecture in the paper is based on the end-to-end training of u-net network to optimize the performance of image tampering detection tasks in combination with residual network. The authors use a deep learning framework to detect image manipulation and tampering and the localization of tampered areas. The method involves training the network on a dataset of authentic and tampered images, and then using the trained network to classify new images as authentic or tampered. The authors also propose a new loss function to improve the performance of the network. In terms of future work, the authors suggest that the proposed method can be extended to detect other types of image manipulations, such as copy-move forgery, splicing, and retouching. They also suggest that the method can be applied to other domains, such as video forgery detection and document forgery detection.

III. PROPOSED METHOD

Our approach for Image Forgery Detection aims to harness the capabilities of deep learning algorithms to accurately identify and localize forged regions within images. The methodology involves a sequence of essential steps, commencing with data preprocessing and culminating in the evaluation of model performance. Figure 2 depicts the proposed system architecture.

Image Preprocessing:

In the realm of Image Forgery Detection, meticulous preprocessing of the input dataset stands as a critical initial stride in fostering robust and accurate model performance. This preprocessing regimen consists of several imperative steps tailored to optimize the images for subsequent forgery detection and localization tasks.

Image Resizing serves as the inaugural step in this process. The rationale behind resizing is two-fold: to standardize images to a uniform dimension and to facilitate computational efficiency within the detection model. The consistent sizing of images harmonizes the dataset, enabling the model to effectively learn patterns and features across all inputs. This uniformity in dimensions is instrumental in minimizing potential biases that might arise from disparate image sizes. Moreover, computationally standardized dimensions ease the burden on subsequent processing steps and model architectures, fostering a more streamlined and efficient workflow.

Following resizing, Normalization becomes pivotal. The normalization process normalizes pixel values, typically scaling them within a predefined range, such as 0 to 1 or -1 to 1. This crucial transformation mitigates the influence of lighting variations or disparities in pixel intensity across images. By standardizing the pixel values, the model training process is expedited, leading to enhanced convergence, and reducing the model's susceptibility to being skewed by variations in pixel intensity due to illumination discrepancies. Gaussian Filtering assumes a crucial role in refining the dataset for forgery detection. The application of Gaussian blur aids in reducing high-frequency noise and smaller details within the images. This noise reduction serves a twofold purpose: it

smoothens the images while concurrently diminishing the impact of minor artifacts or distortions that might obscure the genuine features. The resultant images are more robust and conducive to effective feature extraction, crucial in accurately discerning forged regions.

Simultaneously, Grayscale Conversion streamlines the dataset by converting images from RGB to grayscale. This conversion not only reduces computational complexity but also retains pertinent structural information necessary for forgery detection. Grayscale representations preserve key image attributes while alleviating the computational burden associated with color-based image analysis. The orchestration of these preprocessing steps, meticulously calibrated for image forgery detection, serves as the bedrock for subsequent feature extraction, model training, and accurate localization of forged regions. It is this initial preparatory groundwork that profoundly influences the efficacy, robustness, and accuracy of the forgery detection system.

Feature Extraction:

In the pursuit of robust Image Forgery Detection, the integration of both ResNet and MobileNet v2.0 architectures for feature extraction embodies a pivotal advancement in the arsenal of detection systems. Leveraging these state-of-the-art convolutional neural network architectures enables a comprehensive exploration of distinct feature extraction methodologies, each with its unique advantages in discerning forged regions within images.

A deep convolutional neural network (CNN) called **ResNet**, or Residual Network introduced by [11], addresses the vanishing gradient problem that hinders training of deeper networks. It achieves this through the introduction of residual blocks, each containing a skip connection that allows information to bypass a few layers. This enables the network to construct deeper architectures without compromising learning capabilities.

Preprocessed images are fed as input to the ResNet architecture. Convolutional layers extract features hierarchically, starting from low-level edges and textures to higher-level semantic information. ResNet's secret weapon lies in its residual blocks, which bypass the limitations of deep networks and allow them to extract intricate features from images. A quintessential ResNet block comprises two components, Identity mapping: A direct skip connection that bypasses several layers, allowing information flow without being affected by nonlinearities or transformations. Transformation branch: Consists of stacked convolutional layers followed by non-linear activation functions (e.g., ReLU) that apply transformations to the input.

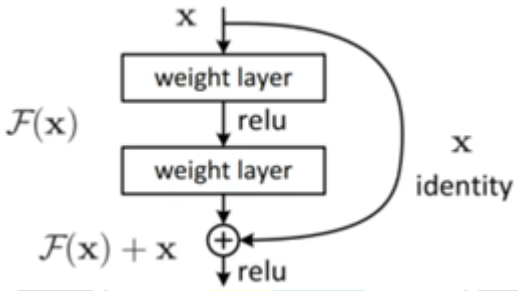

Figure 1: Residual learning: a building block. [11]

The output of the transformation branch is added to the identity mapping, resulting in the final block output: Y = X + F(X) where: Y: Output of the ResNet block, X: Input to the ResNet block, F(X): Output of the transformation branch. This simple addition bypasses the vanishing gradient issue, facilitating information flow even in networks with hundreds of layers. The skip connections directly inject lower-level information into the output, preserving crucial details that might be lost through nonlinearities. This becomes particularly valuable in forgery detection, where subtle manipulations can manifest at various image scales.

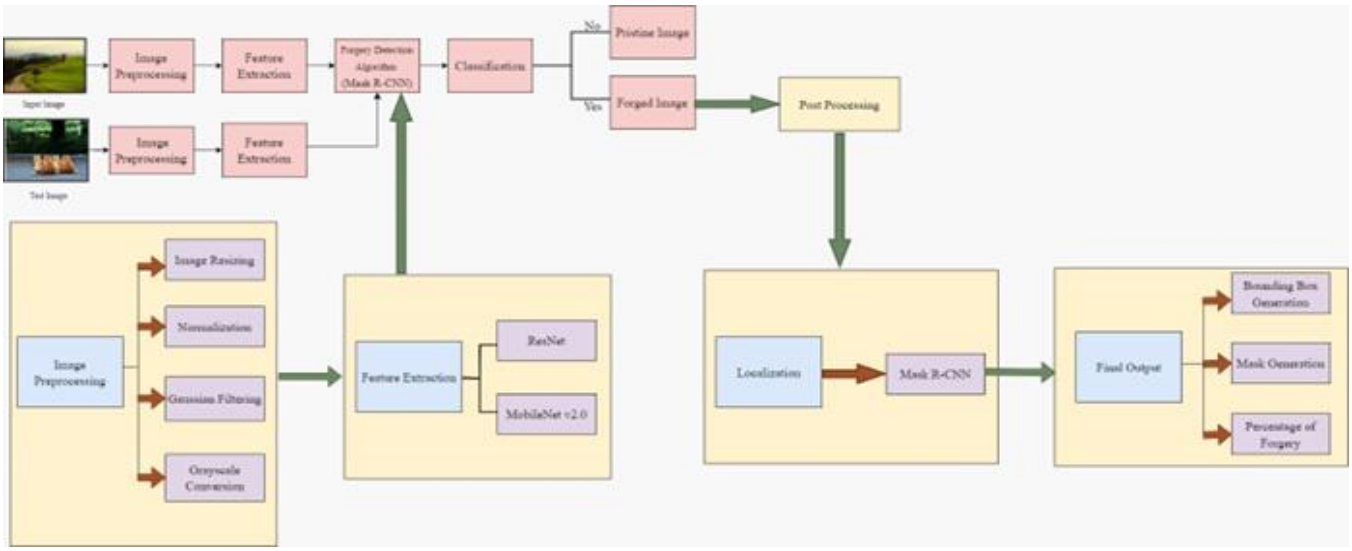

Figure 2: Proposed System Architecture

The input image enters the network, traversing one ResNet block after another. Each block extracts increasingly complex features, Early layers: Detect basic elements like edges and textures. Mid-layers: Combine these elements to form higher-level features like object parts or specific textures. Later layers: Global average pooling aggregates spatial information into a condensed representation concepts like object classes or, in our case, signatures of manipulation for classification. The global average pooling operation computes the average value across each feature map:

$$y_i = \frac{1}{H \times W} \sum_{j=1}^{H} \sum_{k=1}^{W} x_{ijk}$$

Here, y_i represents the i-th feature in the output, H and W denote the height and width of the feature map, and x_{ijk} denotes the i-th feature at position (j,k) in the feature map. The final feature maps act as a fingerprint of the image, encoding details about its content and potential tampering.

Mobile Net v2.0 [12], a lightweight convolutional neural network architecture, emerges as a powerful tool for feature extraction in image recognition and detection tasks. Its strength lies in its compact design, ideal for situations with limited computational resources or real-time processing needs. Unlike traditional CNNs, Mobile Net v2.0 employs several innovative techniques to achieve efficiency while preserving feature quality. Depth wise separable convolutions form the cornerstone of the architecture. This clever approach splits the convolution process into two stages: depth wise and pointwise. Depth wise convolution as shown in figure 4 offers a compelling alternative to traditional convolutional approaches in feature extraction tasks. Its key strength lies in its channel-wise specialization, addressing the redundancy inherent in applying a single filter to all image channels. This redundancy becomes particularly problematic in networks with numerous channels, leading to increased computational cost and potentially obscuring channel-specific features. In contrast, depth wise convolution takes a divide-and-conquer approach. An image is conceptually divided into individual channels, akin to separate stalls in a bustling marketplace. Each channel, representing distinct information like textures or edges, is then assigned a dedicated "detective" filter. This targeted analysis allows each filter to specialize in extracting features relevant to its assigned channel, analogous to an expert focusing on their specific domain within the marketplace.

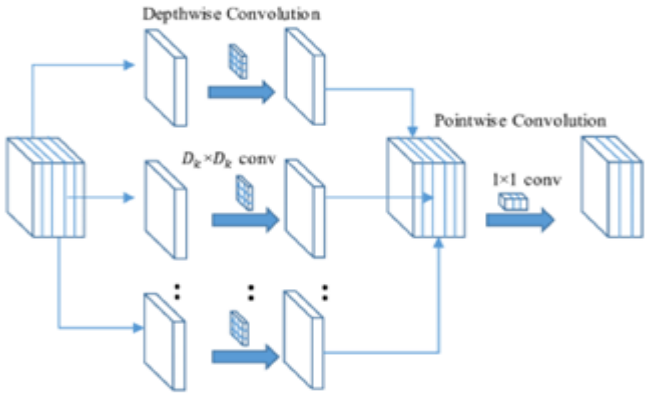

Figure 3: Depth wise separable convolution [14]

Unlike traditional convolutions, pointwise convolution operates through compact 1x1 filters, fostering cross-channel collaboration via element-wise fusion and dimensionality reduction. This technique unlocks hidden relationships between channels, enriching feature maps with complex interdependencies. Consequently, pointwise convolution delivers not just channel-specific details but a richer, more comprehensive understanding of the underlying data, empowering networks with deeper insights for tasks like classification, detection, and segmentation, all while maintaining computational efficiency. This makes it a crucial component of modern CNNs, enabling them to tackle complex tasks with greater accuracy and understanding.

MobileNetV2's architecture relies on two types of blocks, Residual Blocks (Stride 1): These blocks preserve spatial resolution and enable smooth information flow. They consist of three layers: An initial layer that expands channels using 1x1 convolutions with ReLU6 activation, A layer that extracts spatial features efficiently using depth wise convolutions, A final layer that projects feature back to a lower channel count using 1x1 convolutions without non-linearity. Downsizing Blocks (Stride 2): These blocks reduce spatial resolution, allowing the network to capture higher-level features. They share a similar structure to residual blocks, but with a stride of 2 in one layer to achieve down sampling. ReLU6 is used in the first layer, but subsequent layers avoid non-linearities to maintain feature complexity and prevent model limitations. An expansion factor (t=6 in most experiments) strategically increases channel depth within blocks, leading to richer feature representations. For example, an input with 64 channels would result in an internal output with 384 channels as shown in figure 5.

The extracted features from ResNet and MobileNet v2.0 were concatenated along the feature dimension, resulting in a combined representation that leverages the complementary strengths of both models. This straightforward early fusion approach efficiently integrates information from both sources, potentially enhancing performance in downstream tasks.

Forgery Detection Algorithm (Mask R-CNN):

Mask R-CNN [15], an extension of the Faster R-CNN architecture, offers a compelling solution for post-processing tasks in forgery detection. Through its ability to precisely localize objects and generate pixel-level masks, Mask R-CNN becomes an invaluable tool for identifying and delineating forged regions within digital images.

During training, Mask R-CNN learns to identify and delineate the characteristics unique to each category. In this context, forged regions within images are treated as objects, allowing the model to precisely segment and outline these manipulated areas using its pixel-level masks. Simultaneously, pristine images contribute to the model's understanding of unaltered regions and their distinctive features.

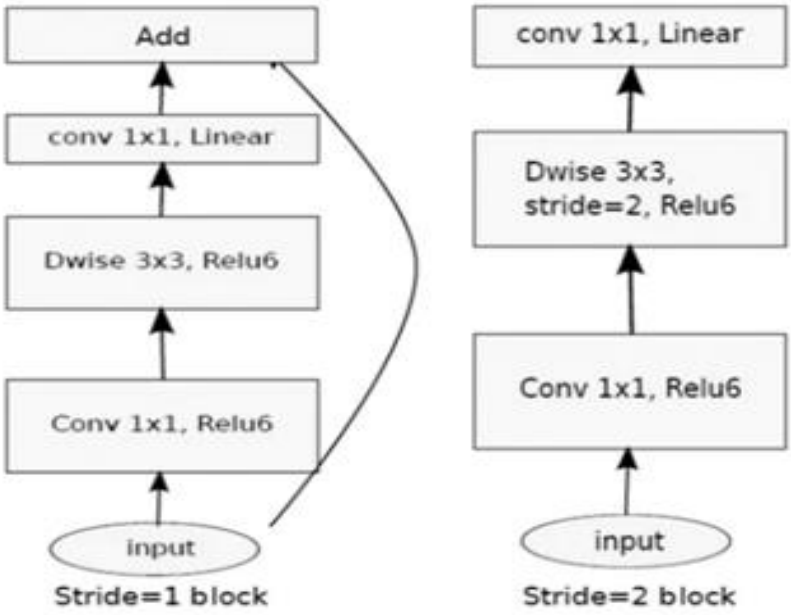

Figure 4: MobileNetV2 Convolutional Blocks [13]

Post-processing:

Image segmentation, a fundamental task in computer vision, entails partitioning digital images into distinct segments or image objects, thereby facilitating the identification and localization of objects and boundaries within them. Mask R-CNN, a versatile deep learning model, adeptly employs two primary segmentation approaches: semantic and instance segmentation.

Semantic segmentation categorizes each pixel within an image into predefined classes, effectively classifying similar objects as a collective group at the pixel level. This approach focuses on the identification and classification of objects without differentiating between individual instances, offering a global view of the image's content.

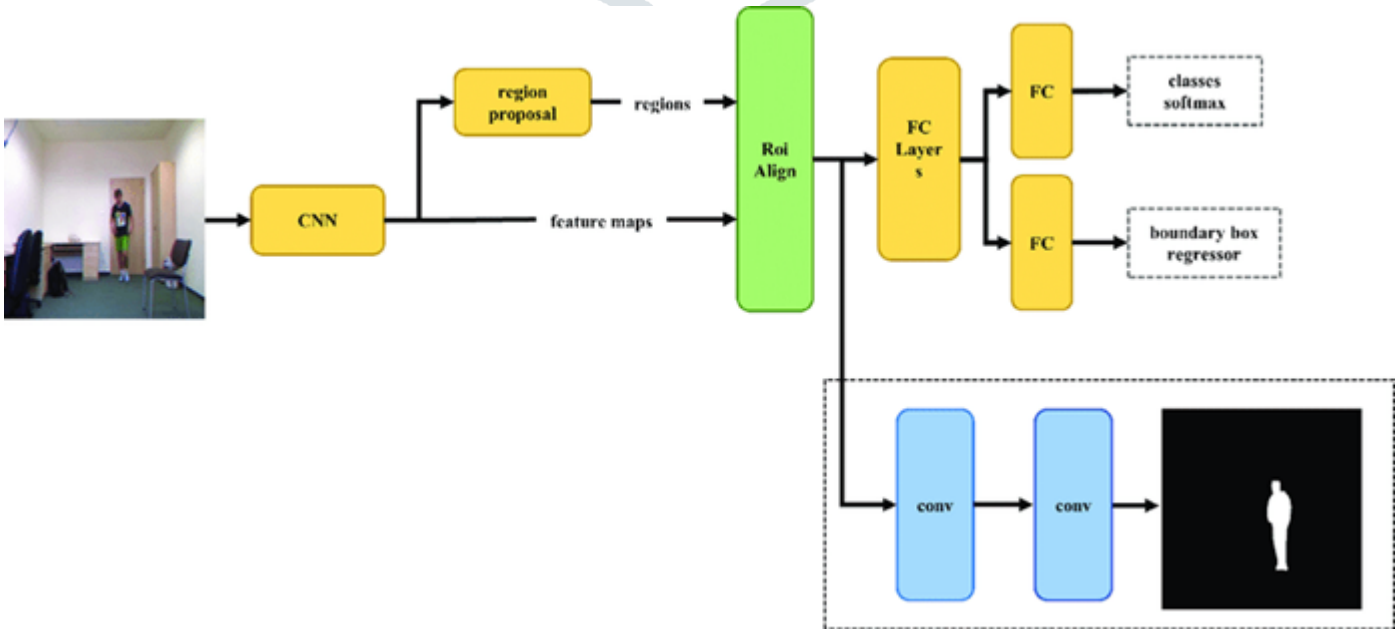

Figure 5: Architecture of Mask R-CNN [16]

In contrast, instance segmentation delves deeper, seamlessly integrating object detection, localization, and classification. It meticulously delineates the exact boundaries of every individual image object, thereby facilitating the identification and localization of objects and boundaries within them. Mask R-CNN, a versatile deep learning model, adeptly employs two primary segmentation approaches: semantic and instance segmentation. Semantic segmentation categorizes each pixel within an image into predefined classes, effectively classifying similar objects as a collective group at the pixel level. This approach focuses on the identification and classification of objects without differentiating between individual instances, offering a global view of the image's content. In contrast, instance segmentation delves deeper, seamlessly integrating object detection, localization, and classification. It meticulously delineates the exact boundaries of each individual object within an image, providing a granular understanding of the scene's composition.

Localization and Mask Generation:

Mask R-CNN built upon Faster R-CNN by incorporating an additional mask prediction branch alongside the existing region proposal network (RPN) and classification/regression branches. The backbone network (e.g., ResNet and MobileNet v2.0) processes the input image and extracts features. These features serve as the basis for subsequent branches in the Mask R-CNN.

Region Proposal Network (RPN):

The RPN acts as an initial step to identify potential forged regions by generating candidate bounding boxes within the image. Leveraging extracted features from the backbone network, it proposes regions of interest (RoIs) along with objectness scores and bounding box coordinates. This process efficiently identifies areas suspected of containing forged content, providing essential hypotheses for further scrutiny during forgery detection and mask generation.

RoI Align:

Post-region proposals, RoI Align plays a crucial role in ensuring accurate feature extraction within proposed regions. Unlike conventional pooling methods, RoI Align preserves spatial fidelity by employing bilinear interpolation, enabling precise feature alignment. This meticulous alignment aids in extracting detailed spatial information crucial for generating masks that accurately delineate manipulated regions within the image.

Fully Connected Layers:

After RoI Align, fully connected layers or intermediate convolutional layers further refine features within detected regions. These layers facilitate additional feature extraction and refinement, enhancing the model's capability to discern intricate details of suspected forged areas, contributing to more precise mask generation and forgery localization.

Mask Head:

The mask head branch specializes in predicting pixel-wise masks for each identified object or suspected forged region. Leveraging features extracted through RoI Align, it employs convolutional layers to predict masks, typically binary representations indicating the presence or absence of object pixels. This branch plays a pivotal role in generating masks that precisely outline forged areas, facilitating accurate localization and analysis of manipulated regions within the image.

Calculating Forged Percentage:

Beyond simply declaring an image forged, our architecture delves deeper, quantifying the manipulation's extent through a precise "forged percentage." The formula for calculating the percentage of forged region can be :

$$\text{Percentage of Forged Region} = \frac{[X-Y]}{\text{Dimension of image}} \times 100$$

were,

X = number of pixels of the entire image,

Y = number of pixels of the forged region

This begins with isolating tampered regions using bounding boxes and pixel-level masks, each a unique fingerprint of suspected forgery. These masks transform the image into a binary format, with white pixels representing the manipulated areas and black pixels the untouched regions. By meticulously counting the white pixels within each mask, or measuring the black background and subtracting it from the image's total pixel count, we get the percentage of the image compromised by tampering. Thus, the final percentage of the forged area is derived using the following formula:

$$\text{Forged Percentage} = \frac{\text{White Pixel Count}}{\text{Total Pixel Count}} \times 100$$

This forged percentage offers a powerful metric, not only allowing us to compare the severity of manipulation across images and datasets but also providing valuable insights into potential impacts on interpretation and analysis. In the realm of forensic investigations, this precise quantification serves as a potent weapon, adding a crucial layer of detail to the evaluation of manipulated evidence. Ultimately, calculating the forged percentage transforms our understanding of digital manipulation from a binary "yes or no" to a nuanced spectrum, empowering us to combat forgery with greater precision and insight.

IV. DATASETS

The effectiveness and robustness of any image forgery detection system heavily rely on the quality and diversity of the datasets used for training, validation, and evaluation. In this study, multiple benchmark datasets renowned within the field of digital image forensics were employed to comprehensively evaluate the proposed image forgery detection framework based on Mask R-CNN.

The selected datasets were chosen for their varied manipulation types, diverse image content, and established ground truth annotations, providing a suitable foundation for assessing the framework's performance across different forgery scenarios.The CASIA dataset [17] comprises two versions, CASIA 1.0 and CASIA 2.0, featuring manipulated images primarily created using Adobe Photoshop. CASIA 1.0 contains 1725 JPEG images, including 800 original non-manipulated images across diverse categories like animals, architecture, scenes, and more, along with 925 tampered images resulting from splicing operations on the originals. CASIA 2.0 boasts 12614 images, featuring uncompressed TIFF, BMP, and JPEG files with varying dimensions from 320×240 to 800×600 pixels. This collection comprises 7491 authentic images in nine categories and 5123 tampered images exhibiting copy-move and splicing manipulations. These datasets offer a diverse array of authentic and manipulated images, showcasing natural and artificial subjects. CASIA 1.0's spliced images stem from alterations made to genuine photos, while CASIA 2.0 encompasses various manipulation types like copy-move and splicing, serving as vital resources for assessing and benchmarking image forgery detection algorithms in digital image forensics.

V. CONCLUSION AND FUTURE WORK

In conclusion, this study presents a comprehensive review and proposed framework for image forgery detection leveraging advanced deep learning architecture, specifically focusing on Mask R-CNN. Through an extensive examination of various components within Mask R-CNN, including the Region Proposal Network (RPN), RoI Align, Fully Connected Layers, and the Mask Head, this research has elucidated the critical role of these components in the context of image forgery localization and mask generation. The in-depth exploration of Mask R-CNN's components showcased their significance in accurately localizing forged regions within images and generating precise masks outlining the manipulated areas. Leveraging the proposed architecture's capabilities, future implementations hold the potential to revolutionize the field of image forensics, enabling more robust and accurate detection of digitally manipulated content.

A significant prospective advancement for the proposed image forgery detection framework involves extending its capability beyond binary classification to include the identification and classification of distinct forgery classes. Augmenting the framework to differentiate between specific manipulation techniques like splicing, copy-move, or retouching within images represents a crucial avenue for future research. This expansion would require the integration of fine-grained classification methodologies within the existing architecture of Mask R-CNN, enabling the system to not only detect but also classify diverse forgery types, offering invaluable insights for forensic analysis. Moreover, to achieve this objective, future research endeavors should focus on developing sophisticated models and training strategies that encompass a broad spectrum of forgery classes. Robust experimentation involving diverse datasets covering various forgery complexities will be essential to validate the model's efficacy in accurately identifying and categorizing different manipulation types. Enhancing the interpretability of the model's decision-making processes and understanding the discriminative features for diverse forgery classes will also be pivotal in advancing the framework's forensic analysis capabilities.

REFERENCES

[1] Abhishek, Jindal, N. Copy move and splicing forgery detection using deep convolution neural network, and semantic segmentation. Multimed Tools Appl 80, 3571–3599 (2021).

[2] M. F. Jwaid and T. N. Baraskar, "Study and analysis of copy-move & splicing image forgery detection techniques," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2017, pp. 697-702, doi: 10.1109/I-SMAC.2017.8058268.I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.

[3] K. D. Kadam, S. Ahirrao, and K. Kotecha, “Efficient Approach towards Detection and Identification of Copy Move and Image Splicing Forgeries Using Mask R-CNN with MobileNet V1,” Computational Intelligence and Neuroscience, vol. 2022, Article ID 6845326, 21 pages, 2022.25R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[4] Islam, Mohammad Manzurul, Gour Karmakar, Joarder Kamruzzaman, and Manzur Murshed. 2020. "A Robust Forgery Detection Method for Copy–Move and Splicing Attacks in Images" Electronics 9, no. 9: 1500. https://doi.org/10.3390/electronics9091500M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[5] Zhang, Q., Sang, J., Wu, W., Cai, B., Wu, Z., Hu, H. (2019). An Image Splicing and Copy-Move Detection Method Based on Convolutional Neural Networks with Global Average Pooling. In: Zhao, Y., Barnes, N., Chen, B., Westermann, R., Kong, X., Lin, C. (eds) Image and Graphics. ICIG 2019. Lecture Notes in Computer Science(), vol 11903. Springer, Cham. https://doi.org/10.1007/978-3-030-34113-8_22

[6] Abdalla, Younis & Iqbal, M. Tariq & Shehata, Mohamed. (2019). Convolutional Neural Network for Copy-Move Forgery Detection. Symmetry. 11. 1280. 10.3390/sym11101280.

[7] A. B. Z. Abidin, H. B. A. Majid, A. B. A. Samah and H. B. Hashim, "Copy-Move Image Forgery Detection Using Deep Learning Methods: A Review," 2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS), Johor Bahru, Malaysia, 2019, pp. 1-6, doi: 10.1109/ICRIIS48246.2019.9073569.

[8] Tu Huynh-Kha, Thuong Le-Tien, Synh Ha-Viet-Uyen, Khoa Huynh-Van and Marie Luong, “A Robust Algorithm of Forgery Detection in Copy-Move and Spliced Images” International Journal of Advanced Computer Science and Applications(IJACSA), 7(3), 2016. http://dx.doi.org/10.14569/IJACSA.2016.070301

[9] Y. Rao and 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), Abu Dhabi, United Arab Emirates, 2016, pp. 1-6, doi: 10.1109/WIFS.2016.7823911.

[10] Y. Deng, "Image Forgery Detection Using Deep Learning Framework," 2022 IEEE 5th International Conference on Information Systems and Computer Aided Education (ICISCAE), Dalian, China, 2022, pp. 105-108, doi: 10.1109/ICISCAE55891.2022.9927668.
[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. arXiv:1512.03385. DOI: 10.48550/arXiv.1512.03385
[12] Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). MobileNetV2: Inverted Residuals and Linear Bottlenecks. arXiv:1801.04381 DOI: 10.48550/arXiv.1801.04381
[13] https://towardsdatascience.com/review-mobilenetv2-light-weight-model-image-classification-8febb490e61c
[14] Chollet, F. (Year). Xception: Deep Learning with Depthwise Separable Convolutions. *arXiv preprint arXiv:1610.02357. https://arxiv.org/abs/1610.02357
[15] K. He, G. Gkioxari, P. Dollár and R. Girshick, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 2980-2988, doi: 10.1109/ICCV.2017.322.
[16] https://www.researchgate.net/figure/Mask-R-CNN-for-background-subtraction_fig2_346894576
[17] J. Dong, W. Wang, and T. Tan, “CASIA image tampering detection evaluation database,” in Proceedings of the 2013 IEEE China Summit and International Conference on Signal and Information Processing, pp. 422–426, Beijing, China, July 2013.