2024 International Conference on Computer, Electronics, Electrical Engineering & their Applications (IC2E3) | 979-8-3503-8853-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/IC2E362166.2024.10827254

Deep Learning Strategies for Effective Image Forgery Detection and Localization

1nd Nandini Kashyap
Computer Science Engineering
Delhi Technological University
Delhi, India
nandinikashyap101@gmail.com

2rd Prince Yadav
Computer Science Engineering
Delhi Technological University
Delhi, India
2001prince03yadav@gmail.com

3th Nikita
Computer Science Engineering
Delhi Technological University
Delhi, India
nikitameena769@gmail.com

4st Ms. Anukriti Kaushal
Computer Science Engineering
Delhi Technological University
Delhi, India
anukritikaushal30@gmail.com

***Abstract*—The field of digital forensics assumes a crucial role in the investigation of digital crimes, with image forgery emerging as a critical area of concern. Delving into the intricacies of image forgery, the paper specifically addresses the challenges posed by splicing and copy/move techniques, underscoring the necessity for advanced methods in accurate localization and segmentation. Despite the existence of various techniques, persistent gaps remain in achieving precise forgery localization and segmentation, highlighting the imperative for robust models capable of handling diverse forgery scenarios. In this context, we propose two distinctive models where VGG19 serves as the backbone for U-Net, and ResNet acts as the backbone for LinkNet. Leveraging the inherent strengths of both convolutional architectures, the depth of VGG19 seamlessly integrates with U-Net, while ResNet's residual learning capabilities effectively meld with LinkNet, presenting a holistic approach to forgery detection. Both models are deployed to localize and segment forged regions within images. Comprehensive experimental evaluations confirm the efficiency of the suggested models when contrasted with existing techniques. Demonstrating superior performance, both models achieve notable accuracy in forgery localization and segmentation tasks. The outcomes highlight the potential impact of these models in advancing the field of digital forensics, particularly in mitigating challenges associated with splicing and copy/move attacks.**

***Index Terms*—Copy/Pasting attacks; VGG19; U-Net; ResNet; LinkNet; Forgery Localization;**

I. Introduction

Lately, there has been a significant increase in both the availability and advancement of user-friendly image editing software and techniques. This has considerably simplified and made image manipulation more affordable. The ease of manipulating images raises concerns, as forgers can exploit these tools to create deceptive photos, spreading misleading information, rumors, or providing false testimony, thereby posing significant societal risks. The emerging field of image forensics, with its focus on distinguishing manipulated photographs, has garnered considerable attention in both scientific and industrial realms. Its objective is to thwart forgers from employing such manipulated images for unethical purposes, be it in business or politics.

In recent times, the identification of altered regions relies on various forensic clues, such as copy/pasting, removal, and splicing. Splicing happens when segments of authentic images are replicated and incorporated into different images, whereas copy/pasting forgery entails duplicating and positioning portions within the confines of a single image. Despite thorough examination, pinpointing these manipulated areas remains a daunting challenge.

Recently, conventional feature extraction-based approaches in detecting forgery are classified into four primary groups: those centered on imaging device properties [1] [2] [3] [4], image compression features [5] [6] [7], essence attributes [8] [9] [10], and hash methods [11] [12] [13]. While classified into distinct categories, these traditional approaches have inherent limitations. i) Techniques centered around image essence may fall short if concealed procedures accompany splicing forgery. ii) Detection methods reliant on imaging device attributes may struggle when device noise levels are low. iii) Recognition based on image compression traits is restricted to identifying images stored in JPEG format. iv) Hashing techniques depend on the hash of the unchanged original image, rendering reliance on a single forgery detection method unfeasible.

In response to the limitations of traditional forgery detection methods, there is a growing need for advanced techniques that can address the evolving landscape of image manipulation. With the rapid progression of digital technologies, novel

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on March 19,2025 at 03:14:20 UTC from IEEE Xplore. Restrictions apply.

approaches are essential to enhance the accuracy and reliability of forensic analyses. This paper delves into the exploration of cutting-edge methodologies to overcome the drawbacks of traditional techniques and provide a more robust framework for image forensics.

This research focuses on the specific challenges posed by splicing and copy/pasting attacks, acknowledging their nuanced nature and potential for evading conventional detection methods. The ability of forgers to exploit hidden processes after splicing forgery highlights the inadequacy of relying solely on image essence methods. Similarly, the vulnerability of detection methods based on imaging device attributes to weak noise intensity necessitates a shift towards more adaptive and resilient approaches.

As we delve into the intricacies of image compression, it becomes evident that exclusive reliance on detecting JPEG format limitations hampers the versatility of forgery detection. Therefore, a comprehensive methodology must encompass a broader spectrum of image formats to ensure the identification of manipulated images, regardless of the compression method employed.

Lately, hash techniques, despite their fundamental nature, encounter challenges when the original, unaltered image cannot be accessed, thereby reducing their usefulness in real-world forensic scenarios. Acknowledging these limitations, there is an urgent need to investigate novel methodologies convolutional neural networks (CNNs) to improve the precision and efficiency of forgery detection.

In the subsequent sections of this paper, we present an advanced forgery detection framework that integrates state-of-the-art convolutional architectures, namely VGG19, U-Net, ResNet, and LinkNet. Leveraging the strengths of these architectures, our proposed models aim to provide a more encompassing solution to the challenges posed by splicing and copy/pasting attacks. Through extensive experimental evaluations, we validate the efficacy of our models, demonstrating their superior performance compared to traditional techniques. The findings of this research add to the ongoing discussion in image forensics and provide a promising direction for addressing the risks linked with image manipulation across different societal sectors.

II. Methodology

To improve image forgery detection and localization, we propose two novel image localization models.

A. Proposed Resnet34+Linknet model

n the proposed framework, ResNet34 serves as the foundational backbone model utilized for feature extraction, which is crucial for subsequent segmentation or localization tasks. Renowned for its deep residual connections, ResNet34 excels in capturing intricate details such as edges, textures, and object components, essential for precise localization and segmentation. The inclusion of these residual connections is vital to ensure a consistent flow of gradients during training, effectively mitigating the vanishing gradient issue commonly faced. ResNet34 is typically pre-trained on extensive datasets such as ImageNet, utilizing the broad knowledge acquired to improve its versatility across different localization or segmentation tasks.

In parallel, the proposed architecture integrates LinkNet because of its computational efficiency while maintaining segmentation accuracy. Operating on an encoder-decoder architecture, LinkNet's [14] encoder extracts pertinent features, while the decoder generates segmentation masks. Leveraging residual connections and skip connections, LinkNet adeptly processes feature maps, preserving crucial spatial information during the upsampling process.

The fusion of ResNet34 with LinkNet is orchestrated through the amalgamation of their respective output feature maps. These feature maps, enriched with comprehensive representations acquired by ResNet34, serve as the input for the decoder component of LinkNet. This fusion capitalizes on the high-quality features extracted by ResNet34, further refined and processed by LinkNet for precise segmentation or localization. The integration of skip connections allows LinkNet's decoder to access feature maps from earlier stages of ResNet34, enabling the model to effectively utilize both low-level and high-level features for segmentation.

During the model's training phase, meticulous optimization procedures are employed. Input images, standardized to a predetermined shape, are fed into the model. Careful adjustment of the learning rate is pivotal to ensure stable convergence, typically favoring a lower learning rate to prevent overshooting during optimization. The optimization objective employs binary cross-entropy loss to train the model in distinguishing between genuine and altered segments within images.

Efficient management of training data, comprising images and corresponding masks, is facilitated using TensorFlow's Dataset API, enabling seamless batch processing and efficient learning from extensive datasets. Furthermore, callbacks such as ModelCheckpoint and TensorBoard are integrated to monitor and enhance the training process. ModelCheckpoint ensures periodic saving of the best-performing model weights, while TensorBoard provides visualizations and metrics for comprehensive performance tracking.

In essence, the fusion of ResNet34 with LinkNet harnesses the strengths of both architectures, resulting in a segmentation model capable of accurately localizing objects or regions of interest within images.

B. Proposed VGG19+UNet model

The architectural fusion of VGG19 and UNet offers a novel approach to the task of image forgery detection, prioritizing intricate feature extraction and precise localization. VGG19, renowned for its depth and comprehensive convolutional layers, serves as the foundational feature extractor in this hybrid model. Having been pre-trained on ImageNet, VGG19 inherently possesses a robust capability to capture nuanced details within images, a vital aspect in discerning subtle manipulations indicative of forgery.

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on March 19,2025 at 03:14:20 UTC from IEEE Xplore. Restrictions apply.

Simultaneously, the proposed architecture incorporates UNet's design [15], enhancing VGG19's feature extraction capabilities. UNet's encoder-decoder structure, along with skip connections, allows the model to effectively handle the extracted features. The encoder section accurately identifies pertinent features from the input image, while the decoder part, employing upsampling techniques, reconstructs these features into a comprehensive segmentation mask. The deliberate inclusion of skip connections enables the merging of high and low-level features, a crucial factor that enhances the precise identification of manipulated regions in the image.

The fusion of VGG19 with UNet takes place by combining their individual output feature maps. The feature maps extracted by VGG19 are integrated with the decoder segment of UNet, which is responsible for upsampling and generating the segmentation mask. This fusion allows the model to effectively merge deep contextual information from VGG19 with high-resolution feature maps from UNet, leading to improved object localization and boundary delineation accuracy.

Methodologically, the training protocol for the VGG19+UNet model mirrors that of its counterpart, the ResNet+LinkNet model. Input images are resized and subsequently fed into the model for processing. Consistency is upheld in terms of the learning rate, loss function, and optimization algorithm, ensuring a standardized approach to training. The utilization of TensorFlow's Dataset API streamlines the handling of training data, while the integration of callbacks such as ModelCheckpoint and TensorBoard facilitates monitoring and optimization of the training process.

Despite the inherent computational demands attributed to VGG19's depth and UNet's expansive feature maps, the resultant VGG19+UNet architecture emerges as a paradigm of unparalleled performance. Its proficiency in capturing intricate details and accurately localizing image forgeries signifies a remarkable advancement in the domain of image forgery detection.

III. EXPERIMENTAL RESULTS

A. Dataset

We have considered the dataset from the IEEE IFS-TC Image Forensics Challenge [16], which comprises a collection of both authentic and manipulated images. The dataset encompasses both 'pristine' images, representing unaltered originals, and 'forged' images, which have undergone modifications. These images originate from diverse digital cameras capturing various indoor and outdoor scenes. Each forged image is accompanied by a corresponding map image indicating the specific regions that have been altered.

The dataset is segmented into distinct training sets, testing sets, and validation sets, comprising a total of 1,273 test images, 3,893 train images, and 1,000 validation images. Within the training set, pristine images are denoted by a white mask, where all pixels are set to 255, indicating the absence of any tampering. Conversely, regions of manipulation within the images are marked by black regions in the mask, with pixels set to 0.

The ResNet34 + LinkNet and VGG19 + U-Net models underwent thorough evaluation using a dataset containing manipulated images. The dataset encompassed diverse forms of digital image manipulation, such as copy/pasting and splicing. Model training employed an Adam optimizer with a fixed learning rate of 1e-4, and the training duration was capped at a maximum of 10 epochs.

B. Evaluation Metrices

1 . F1 Score : Incorrectly detecting manipulated pixels (FP), precisely identifying manipulated pixels (TP), and erroneously identifying unaltered pixels (FN) are pivotal measures for evaluating the detection of image forgery localization. Recall, Precision, and F1 Score are utilized as metrics for evaluation to gauge the effectiveness of the suggested algorithms in detecting manipulation in the conducted tests.

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \tag{1}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{2}$$

$$\text{F1 Score} = 2 \cdot \frac{\text{Recall} \cdot \text{Precision}}{\text{Recall + Precision}} \tag{3}$$

In this context, Recall, as described in Eq. (1), indicates the probability of accurately detecting manipulated areas in the ground truth image. Precision, formulated in Eq. (2), represents the likelihood that the identified regions in the ground truth image truly correspond to manipulated areas. The F1 Score, articulated in Eq. (3), integrates Recall and Precision into a unified metric to comprehensively evaluate the overall localization performance.

2. Log Loss : Log Loss, commonly known as binary entropy loss, measures the accuracy of predicted probabilities compared to true labels. Log Loss Formula:

$$\text{Log Loss} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)]$$

where $N =$ number of samples,
$y_i =$ true label of the i-th sample,
$p_i =$ predicted probability of the i-th sample.

C. Inferences

The findings derived from our results suggest a high degree of accuracy in identifying the forged regions' boundaries. This is apparent from the performance observed in the output. In the provided data, the first column pertains to the forged images, the central column displays the predicted mask of the forged regions, and the last column shows the ground-truth mask.

The assessment of our results on the IEEE IFS-TC Image Forensics Challenge dataset is presented in a comprehensive

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on March 19,2025 at 03:14:20 UTC from IEEE Xplore. Restrictions apply.

TABLE I
EVALUATION OF PROPOSED MODEL LOCALIZATION

Model	**Evaluation Metrics**	
Architecture	***F1 score***	***logloss***
VGG19 + Unet	0.8702	0.5801
ResNet + linknet	0.9557	0.3166

TABLE II
ASSESSMENT OF LOCALIZATION PERFORMANCE OF RECENT MODELS ON CASIA

Model	**F1 Score**
Rich Feature [18]	0.408
Local Descriptor [19]	0.58
Ours (VGG19+Unet)	0.6579
Ours (ResNet34+Linknet)	0.7147

table, showcasing the F1 score and log loss metrics for both of our proposed models. Subsequent to this, we present comparisons between our experimental results and several other techniques that have been previously employed on the CASIA dataset [17].

In evaluating the performance of our proposed models for image forgery detection, we compared them against existing techniques using two key metrics: F1 score and logloss. The results, presented in Table I, demonstrate the effectiveness of our approach. Both models, employing VGG19 with U-Net and ResNet with LinkNet architectures, showed significant improvements over previous methods. Specifically, the ResNet-based model achieved an impressive F1 score of 0.9557 and a logloss of 0.3166, surpassing the performance of the VGG19-based model, which still produced commendable results with an F1 score of 0.8702 and a logloss of 0.5801.

Furthermore, when comparing our models' performance on the CASIA dataset against recent approaches, as detailed in Table II, it is evident that our ResNet-based model outperforms existing methods with an F1 score of 0.7147. The VGG19-based model also demonstrates competitive performance with an F1 score of 0.6579. These findings underscore the efficacy of our proposed models in accurately localizing and segmenting forged regions within images, thereby advancing the field of digital forensics and addressing challenges associated with image forgery, particularly splicing and copy/move attacks.

D. Conclusion

In summary, our study offers compelling evidence of the effectiveness of the proposed ResNet + LinkNet and VGG19 + UNet models for detecting and localizing image forgery. The thorough evaluation conducted on the IEEE IFS-TC Image Forensics Challenge dataset illustrates their superior performance, delivering precise and dependable results. The visualizations, metrics, and comparative analyses presented in this study highlight the potential applications of our models in real-world scenarios. Ongoing research is essential for continuous refinement and adaptation to emerging challenges in image manipulation. Our proposed models contribute significantly to the discourse on image forensics, paving the way for future innovations in forgery detection and localization.

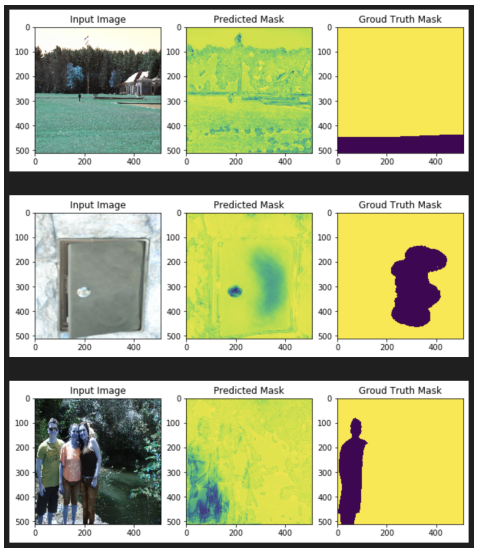

Fig. 1. Visual Comparison: Input Image, Predicted Mask, and Ground Truth Mask

REFERENCES

[1] Y.-f. Hsu and S.-f. Chang, "Detecting image splicing using geometry invariants and camera characteristics consistency," pp. 549–552, 2006.

[2] M. K. Johnson and H. Farid, "Exposing digital forgeries through specular highlights on the eye," pp. 311–325, 2007.

[3] B. Mahdian and S. Saic, *Detection of Resampling Supplemented with Noise Inconsistencies Analysis for Image Forensics*, 2008.

[4] H. Gou, A. Swaminathan, and M. Wu, "Noise features for image tampering detection and steganalysis," in *2007 IEEE International Conference on Image Processing*, 2007, vol. 6, pp. VI – 97–VI – 100.

[5] M. K. Johnson and H. Farid, *Exposing Digital Forgeries in Complex Lighting Environments*, 2007, vol. 2, no. 3.

[6] Z. Lin, J. He, X. Tang, and C.-K. Tang, "Fast, automatic and fine-grained tampered jpeg image detection via dct coefficient analysis," vol. 42, no. 11, 2009, pp. 2492–2501. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031320309001198

[7] S. Ye, Q. Sun, and E.-C. Chang, "Detecting digital image forgeries by measuring inconsistencies of blocking artifact," pp. 12–15, 2007.

[8] W. Chen, Y. Shi, and W. Su, "Image splicing detection using 2-d phase congruency and statistical moments of characteristic function - art. no. 65050r," 02 2007.

[9] W. Wang, J. Dong, and T. Tan, "Effective image splicing detection based on image chroma," pp. 1257–1260, 2009.

[10] X. Zhao, J. Li, S. Li, and S. Wang, "Detecting digital image splicing in chroma spaces," pp. 12–22, 2011.

[11] Z. Tang, X. Zhang, X. Li, and S. Zhang, "Robust image hashing with ring partition and invariant vector distance," pp. 200–214, 2016.

[12] X. Wang, K. Pang, X. Zhou, Y. Zhou, L. Li, and J. Xue, "A visual model-based perceptual image hash for content authentication," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 7, pp. 1336–1349, 2015.

[13] C.-P. Yan, C.-M. Pun, and X.-C. Yuan, "Quaternion-based image hashing for adaptive tampering localization," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 12, pp. 2664–2677, 2016.

[14] A. Chaurasia and E. Culurciello, "Linknet: Exploiting encoder representations for efficient semantic segmentation," pp. 1–4, 2017.

[15] M. M. Qureshi and M. G. Qureshi, "Image forgery detection & localization using regularized u-net," pp. 434–442, 2021.

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on March 19,2025 at 03:14:20 UTC from IEEE Xplore. Restrictions apply.

[16] "IEEE IFS-TC Image Forensics Challenge dataset." [Online]. Available: https://web.archive.org/web/20171013200331/http://ifc.recod.ic.unicamp.br/fc.website/index.py?sec=5
[17] "IEEE IFS-TC Image Forensics Challenge dataset." [Online]. Available: https://paperswithcode.com/dataset/casia-v1
[18] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Learning rich features for image manipulation detection," pp. 1053–1061, 2018.
[19] Y. Rao, J. Ni, and H. Zhao, "Deep learning local descriptor for image splicing detection and localization," *IEEE Access*, vol. 8, pp. 25 611–25 625, 2020.

Authorized licensed use limited to: BEIHANG UNIVERSITY. Downloaded on March 19,2025 at 03:14:20 UTC from IEEE Xplore. Restrictions apply.