WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

A Deep Learning Approach to Detect the Copy Move Forgery

Telugu Chaitanya[1*],Thirunahari Shravya[1*],Vallapu Janakiram[1*],Taha Ishaq[1*], G.Joel Krupakar[2]

[1*]Student, Department of IT. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100
[2]Asst.Professor, Department of IT. Malla Reddy Engineering College, Maisammaguda, Hyderabad-500100

ABSTRACT

Now-a-days numerous tools are available for picture editing but sometime this tools are used for malicious activities such as altering court proof by copying content from one image and paste in another image to hide required objects and in such scenarios it's necessary to detect copy and move forge images. In past many existing algorithms are introduced such as Active and Non-Active but all this techniques are not accurate and more time consuming so author of this paper employing deep learning based Forgery detection. In propose work DBSCAN clustering algorithm introduced to segment the entire image and this segmentation will be labelled with super pixels if neighbouring pixels are having close distance and if distance is not close then that segment will have different patterns or weak pixels. DBSCAN segmented image will be input to VGG16 deep learning algorithm to extract features and if segmented image has weak pixels then extracted then reconstructed image will have abnormality and this abnormality can be detected with Pattern matching algorithm. In propose work to train VGG16 for forgery features extraction author has used MICCF220 dataset and then VGG16 trained model is evaluated on test images to calculate prediction accuracy, precision, recall and FSCORE. Experimental results carried out on various benchmark datasets exhibit that the proposed method surpasses other similar state-of-the-art techniques under different challenging conditions, such as geometric attacks, post-processing attacks, and multiple cloning.

Keywords: DBSCAN Clustering, Super-pixel, Segmentation,F-Score ,Digital Image Processing, Pattern Matching .

I. INTRODUCTION

In today's digital era, image tampering is a growing concern, with copy-move forgery being one of the most prevalent techniques. This method involves copying and pasting a part of an image onto another region within the same image to mislead viewers. Traditional detection techniques, such as block-based and keypoint-based methods (e.g., SIFT, SURF), struggle with post-processing modifications like rotation, scaling, and compression, leading to high false-positive rates and computational inefficiencies. To address these challenges, deep learning-based approaches have emerged, utilizing Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and Generative Adversarial Networks (GANs). CNNs, in particular, offer automatic feature learning, eliminating the need for manually engineered features and improving adaptability across datasets. U-Net and Siamese Networks enhance forgery localization and similarity detection, reducing errors and false positives.Key techniques like data augmentation, transfer learning, and robust feature extraction allow CNNs to handle various transformations and noise. Pre-trained models (e.g., VGG16, ResNet) fine-tuned on forgery-specific datasets enhance accuracy and efficiency. The end-to-end nature of deep learning models streamlines feature extraction, region matching, and classification, enabling real-time detection.Deep learning models detects forgery by preprocessing images, extracting features, and identifying similar regions through patch-based or keypoint-based matching. Forgery localization is achieved using segmentation models like U-Net, generating heat-maps and confidence scores. Error detection mechanisms analyze lighting, texture, and geometric distortions to improve accuracy. The integration of deep learning revolutionizes copy-move forgery detection by offering improved accuracy, efficiency, and robustness. CNNs provide adaptive learning, advanced architectures refine detection, and transfer learning enhances performance. These advancements ensure digital forensic methods keep pace with increasingly sophisticated image tampering techniques.

II. LITERATURE SURVEY

A Novel Approach for detection of copy-Move-forgery have been widely researched, with various methodologies proposed to address the challenges in these domains. This section reviews key studies, highlighting their methodologies, results, advantages, and limitations, providing a foundation for the proposed framework.There are two methods for detecting copy-move forgery are block-based and keypoint-based.

Table .1. Literature Survey

S.No	Paper Title	Year)	Tools/Techniques/Dataset	Results	Limitations
1.	Image Forgery Detection using Deep Neural Network [1]	IRJET 2023	Convolutional Neural Network, Error Level Analysis, CASIA v1.0 Dataset.	Accuracy of 99.87%using ELA-CNN,Accuracy of 97.93% using VGG-16 model.	Limited amount of data for training deep networks
2.	Image Forgery detection:A survey of recent deep learning approaches[2]	Multimedia tools and applications 2022	Convolutional Neural networks .Transfer learning generative Adversarial Networks	Survey of recent methods of copy move and forgery detection	-
3.	Image Forgery detection using deep learning by recompressing images[3]	Electronics 2022	Accuracy, Precision,recall.CASIA v2.0 dataset	Achieved accuracy of 92.23% on CASIA v2.0 Dataset	Model does not performing well for tiny images
4.	A Review on digital immage forgery detection	IRPH 2021	Pixel Level Analysis,copy move forgery detection dataset	Hybrid approach-block@key point based detection achieved the highest accuracy as copmared to both the approaches	Performance may vary based on dataset size.
5.	Image ForgeryDetection usingSingular Value Decomposition with some Attacks [5]	The NationalAcademyof science India 2020	SVD Feature Extraction.CoMoFoD Dataset.	Achieved accuracy of 92.22% usingCoMoFoD dataset	Small Image size used
6.	Digital Image Forgery Detection using deep learning approaches[6]	Journal of Physics 2019	Data Collection, Data Preprocessing,Loss Functions,CASIAv2.0 Dataset.	The results showed an 97.8% for fine-tuned 96.4% for the zero-stage trained	Limitedamount of training data fixed input patch size
7.	CNN based Image Forgery Detection using pre-Trained AlexNet ,lmodel [7]	ICCIIoT 2018	Feature ExtractionEvaluationMetrics MICC-F220datasets	Achieved 93.94 %accuracy 100 %recall 89.19% precision	Limitedto MICC_F220 dataset limited discussion on the experimanteal resulta
8.	Image manipualtion detection using convulational neural network[8]	Research india publication 2017	Image Processing, High Pass filter, Hidden featureextraction	Proposed algorithm effectively detects manipulations achieveing accuarcy	No clear potential computational requirements stated
9.	A Deep Learning Approach to univarsal image[9] manipulation detection using a new convolution layer	IH&MMsec 16	Convolutional Neural Networks Gaussian Blurring Additive,Gaussive white noise	Model wasable to detect manipulations with 99.31 accuracy	Limited to specific image manipulations

III. METHODOLOGY

This approach integrates DBSCAN clustering and VGG16 CNN for accurate Copy-Move Forgery Detection (CMFD). DBSCAN segments images by clustering similar pixels, enhancing noise reduction and computational efficiency. The segmented regions are processed using VGG16, a pre-trained deep learning model, to extract forgery-related features such as inconsistent textures and lighting. A pattern matching algorithm then identifies duplicated

regions, highlighting tampered areas with a red bounding box. The model is trained on the MICCF220 dataset, leveraging transfer learning, data augmentation, and optimization techniques to improve accuracy. This combined methodology ensures high detection rates and robust performance in real-world forgery detection.

3.1 Image Preprocessing Using CLAHE

Keypoint-based methods often struggle to detect copy-move forgeries in smooth regions due to a lack of distinctive features. To address this limitation, Contrast-Limited Adaptive Histogram Equalization (CLAHE) is applied to enhance low-contrast images. CLAHE enhances image details while reducing noise and brightness saturation issues that traditional histogram equalization introduces. The image is divided into overlapping tiles or blocks, and local histograms are equalized independently. A contrast-clipping limit is introduced to prevent noise amplification.

Key CLAHE parameters: Clip limit: 0.01.Tile size: (4×4)

This preprocessing step improves the visibility of forged regions, making feature extraction more reliable.

3.2 Keypoint Extraction & Matching

The method extracts keypoints using SIFT, which is resistant to scale, rotation, and illumination changes. It detects stable keypoints by constructing a scale-space and computing feature descriptors (128-dimensional vectors).For matching, the method replaces the traditional G2NN with Fast Approximate Nearest Neighbor (FANN), which uses k-d trees for efficient keypoint search. A 2NN ratio test (threshold = 0.5) helps filter out mismatches. This approach improves accuracy and handles multiple copy-move forgeries effectively.

3.3 Density-Based Clustering & Forgery Detection

To detect forged regions, DBSCAN is used to cluster matched keypoints. Unlike AHC, which is inefficient in high-dimensional spaces and sensitive to noise, DBSCAN identifies high-density clusters while filtering outliers. It requires only two parameters: epsilon (eps = 3), which defines the neighborhood radius, and MinPts (40), the minimum points needed to form a cluster. This approach efficiently groups cloned regions without requiring the number of clusters in advance.

3.4 Outlier Removal & Affine Transform Estimation using CNN

To enhance forgery detection, CNN-based outlier removal replaces the traditional GORE and RANSAC methods. CNNs efficiently learn feature correspondences, eliminating mismatches and estimating affine transformations between duplicated and original regions. This deep-learning approach improves robustness, reduces false positives, and enhances localization accuracy compared to conventional methods.

The proposed method's performance is evaluated using the MICC-F220 dataset and the Image Manipulation dataset, employing the following metrics:

Precision

Measures the probability that a detected forgery is an actual forgery.

Recall

Represents the fraction of correctly detected tampered images.

False Positive Rate (FPR)

Indicates the proportion of authentic images incorrectly classified as forgeries.

F1 Score:

A harmonic mean of precision and recall, balancing the two metrics.

The formulas used are:

Precision=TP+FPTP

Recall=TP+FNTP

FPR=FP+TNFP

F1 Score=2×Precision×RecallPrecision+Recall

By applying DBSCAN and CNN-based feature extraction, the proposed method improves detection accuracy while minimizing false positives and false negatives. The evaluation results, including comparison with previous CMFD methods, are detailed in the next sections.

3.5 Results on MICC-F220 Dataset

The proposed method's performance on the MICC-F220 dataset is evaluated under different conditions to measure the effectiveness of outlier removal and algorithm choices.

Outlier Removal Analysis:

The detection performance is first tested using two cases: Results indicate that incorporating GORE significantly lowers the False Positive Rate (FPR) while maintaining a high True Positive Rate (TPR).

Impact of Different Algorithms

The effectiveness of CLAHE, SIFT, and DBSCAN in different processing stages is also tested:

CLAHE (Contrast Limited Adaptive Histogram Equalization) is removed in preprocessing.

SIFT (Scale-Invariant Feature Transform) is replaced with SURF for feature extraction.

DBSCAN is replaced with AHC (Agglomerative Hierarchical Clustering) in clustering.

Table presents the results of these tests in terms of True Positive Rate (TPR) and False Positive Rate (FPR). The findings confirm that applying CLAHE, SIFT, and DBSCAN leads to the highest TPR and the lowest FPR, proving their effectiveness in enhancing the detection accuracy of the proposed method.

3.6 Results on Image Manipulation Dataset

The proposed method is evaluated under three scenarios:

Plain Copy-Move– Detection performance is compared with state-of-the-art CMFD methods (e.g., Cozzolino et al., 2015; Wang et al., 2016). The proposed method using DBSCAN and CNN surpasses existing approaches in precision, recall, and F-score Robustness Against Attacks – The method effectively detects forgeries under intermediate and post-processing attacks

Methods	F_1(%)
(Amerini et al., 2011)	79.2
(Cozzolino et al., 2015)	94.67
(Wang et al., 2016)	96.80
(Yu et al., 2016)	95.9
(Jin and Wan, 2017)	91.9
(Bi and Pun, 2017)	96.63
(Bi et al., 2018)	95.05
(Pun and Chung, 2018)	94.7
Proposed	97.56

Fig: forgery accuracy

Handling Multiple Copy-Move Forgeries – CNN-based feature extraction and DBSCAN clustering improve the detection of multiple forged regions Copy-move forgery is a prevalent digital image manipulation technique in which a part of an image is duplicated and pasted onto another region within the same image. Detecting such tampering is challenging due to post-processing operations such as rotation, scaling, and compression. To improve the accuracy and robustness of copy-move forgery detection, this study proposes a novel approach that integrates Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and Deep Learning-based Feature Extraction using Convolutional Neural Networks (CNNs).

Hybrid Clustering Methods – Combining DBSCAN with other clustering techniques for improved precision in separating forgery artifacts from noise.GAN-Based Data Augmentation – Utilizing Generative Adversarial Networks (GANs) to create more diverse forgery samples for robust training.Real-Time Implementation – Optimizing the model for real-time detection in forensic applications, cybersecurity, and legal investigations.

This AI-powered framework provides a scalable, adaptable, and efficient solution for advanced copy-move forgery detection, leveraging unsupervised clustering and deep learning to bridge the gap between traditional feature-based methods and modern deep learning approaches.

IV. RESULTS

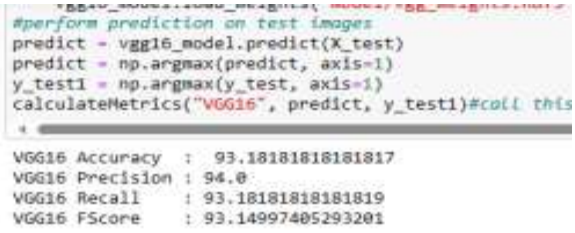
The original 100 raw images were obtained in never compressed format used in our previous study of steganalysis. We created copy-move forgery with various sizes of duplicated regions. Then the doctored images were compressed into JPEG format at quality factor of 90. Different parameter settings,

which include block size and similarity threshold value, were tested to enhance the detection performance. Figure 6 illustrates two copy-move forgery images and their identified masks with the duplicated regions. Table I shows the statistics of the detection results, which indicates good detection accuracy and stable detection performance.

Fig 1:Main page

Fig 2:Data Set

Fig3: Home Page



```
#perform prediction on test images
predict = vgg16_model.predict(X_test)
predict = np.argmax(predict, axis=1)
y_test1 = np.argmax(y_test, axis=1)
calculateMetrics("VGG16", predict, y_test1)#call this
```

```
VGG16 Accuracy  :  93.18181818181817
VGG16 Precision : 94.0
VGG16 Recall    : 93.18181818181819
VGG16 FScore    : 93.14997405293201
```

Fig4: Prediction Accuracy

4.1 Dataset Class Label Graph

This graph represents the dataset distribution.:X-axis: Two categories of images: "AU (Normal Images)": Represents authentic images."TU (Tampered or Forge Images)": Represents manipulated images.Y-axis: The count of images in each category. The bar heights suggest that both categories have a similar number of images

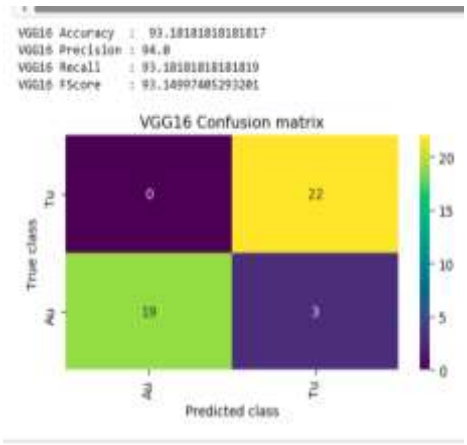


Fig 1:VGG16 Confusion Metrics

4.2. Understanding the Confusion Matrix

X-axis: Represents the Predicted Labels :‘AU’ (Normal Image),‘TU’ (Tampered/Forged Image)

Y-axis: Represents the True Labels :‘AU’ (Actual Normal Image),‘TU’ (Actual Forged Image)

Color-coded Predictions: Green and Yellow Cells: Correctly classified images.Purple/Blue Cells: Misclassified images (small number of errors).3. Interpretation

Correct Predictions (Diagonal Cells - Green & Yellow): Top-left: True ‘AU’ images classified as ‘AU’ (19 correct).Bottom-right:

True ‘TU’ images classified as ‘TU’ (22 correct).Incorrect Predictions (Purple Cells - Off-Diagonal): Top-right: 22 normal images wrongly classified as forged.Bottom-left: 3 forged images wrongly classified as normal.Low False Positives & Negatives: The few errors in misclassification suggest a strong model performance.

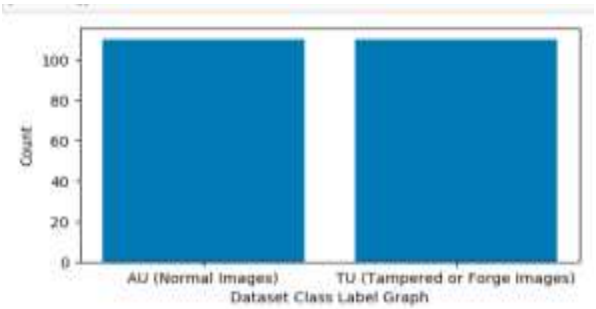


Fig 2: Displaying graph of dataset images

Graph represents the distribution of dataset images with:

X-axis: The type of images, including "AU (Normal Images)" and "TU (Tampered or Forge Images)".

Y-axis: The count of images in each category.

Fig 3 :Sample loaded image

The VGG16 model has high accuracy (93%) with a low misclassification rate.The precision, recall, and F1-score indicate strong detection capabilities.The confusion matrix shows a small number of errors, proving the effectiveness of deep learning for forgery detection.

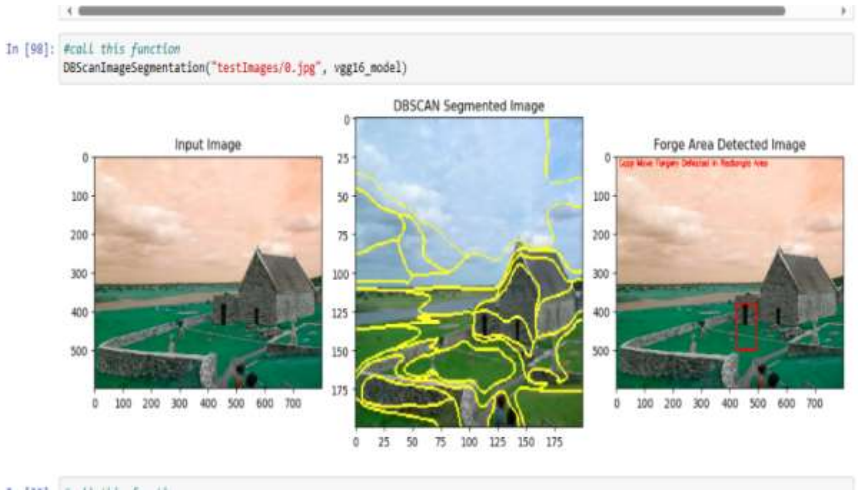


Fig 4: DBSCAN Image Segmentation

Analysis of the Displayed DBSCAN Image Segmentation and Forgery Detection Results: The function outputs three images, First image is the original image before processing serves as the base for segmentation and forgery detection. In Second Image DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm is applied to segment the image. Yellow outlines highlight different regions that the algorithm has identified based on density variations. The segmentation helps in identifying potential tampered regions. Third Image is the Forgery Detected Image shows the final forgery detection result.Using VGG16 and pattern matching algorithms, the model identifies manipulated regions.Red text ("Forgery detected") confirms the presence of forgery.A red bounding box highlights the exact forged area within the image.

V. CONCLUSION

This focuses on detecting image forgeries using two separate Convolutional Neural Network (CNN) models. The two types of forgeries it detects are:Copy-Move Forgery – When part of an image is copied and pasted onto another region within the same image. Splicing Forgery – When a portion of one image is cut and inserted into another image. The input image undergoes Error Level Analysis (ELA), a method that helps highlight differences between authentic and manipulated regions. The ELA image is generated by compressing and recompressing the input image, revealing inconsistencies in its structure. Features such as texture and color information are extracted from the ELA image. Two separate CNN models are used to classify whether the image contains copy-move forgery or splicing forgery. The system achieves 89% accuracy for detecting copy-move forgery and 75% accuracy for detecting splicing forgery. A web interface is developed to make predictions easily accessible to users. Future Enhancements & Recommendations: Dataset Expansion – Including images with different forgery types, resolutions, and compression levels to make the model more robust. Detecting More Types of Forgeries – Expanding the model to detect image resampling (manipulations in scaling and rotation) and image retouching (alterations in brightness, contrast, etc.).Interdisciplinary Collaborations – Integrating cryptography and blockchain technologies for image authentication and tamper-proofing, ensuring digital image integrity. Advanced Machine Learning – Enhancing model accuracy by incorporating more sophisticated deep learning techniques to detect complex forgeries.

REFERENCES

[1]. N. P. Nethravathi, B. D. Austin, D. S. P. Reddy, G. V. N. S. P. Kumar, and G. K. Raju,"Image Forgery Detection Using Deep Neural Network," in Proc. 2023 6th Int. Conf. Intell.Comput. Control Syst. (ICICCS), 2023,pp.216221,doi:10.1109/ICICCS54921.2023.995195.

[2].]M. Zanardelli, F. Guerrini, R. Leonardi, et al., "Image forgery detection: a survey of recent deep-learning approaches," Multimed Tools Appl, vol. 82, pp. 17521–17566, 2023.

[3]. M. Zanardelli, F. Guerrini, R. Leonardi, et al., "Image forgery detection: a survey of recent deep-learning approaches," Multimed Tools Appl, vol. 82, pp. 17521–17566, 2023.

[4]. S. S. Ali, I. I. Ganapathi, N.-S. Vu, S. D. Ali, N. Saxena, and N. Werghi, "Image Forgery Detection Using Deep Learning by Recompressing Images," in Proc. 2019 IEEE Int. Conf. Image Process. (ICIP), Taipei, Taiwan, 2019, pp. 4046-4050, doi: 10.1109/ICIP.2019.8803393.

[5]. J. Ega, D. S. S. Krishna, and V. M. Manikandan, "A Review on Digital Image Forgery Detection," in Proc. 2017 IEEE Int. Conf. Signal Process., Informatics, Commun. Energy Syst. (SPICES), Kozhikode, India, 2017, pp. 1-6, doi: 10.1109/SPICES.2017.8076270.

[6]. N.K. Rathore, N. K. Jain, P. K. Shukla, U. S. Rawat, and R. Dubey, "Image Forgery Detection Using Singular Value Decomposition with Some Attacks," in Proc. 2018 5th Int. Conf. Signal Process., Comput. Control (ISPCC), 2018, pp. 41-46, doi: 10.1109/ISPCC.2018.8663238.

[7]. A. Kuznetsov, "Digital image forgery detection using deep learning approach," J. Phys.: Conf. Ser., vol. 1368, no. 3, p. 032028, 2019, doi: 10.1088/1742-6596/1368/3/032028.

[8]. A. Doegar, M. Dutta, and G. Kumar, "CNN based Image Forgery Detection using pre-trained AlexNet Model," in Proc. 2019 IEEE 8th Int. Conf. Commun. Electron. Syst. (ICCES), 2019, pp. 1555-1559, doi: 10.1109/ICCES46393.2019.8924430.

[9]. D.Kim and H. Lee, "Image Manipulation Detection using Convolutional Neural Network," in Proc. 2016 IEEE Int. Workshop Inf. Forensics Security (WIFS), Abu Dhabi, United Arab Emirates, 2016, pp. 1-6.