Applied Soft Computing 131 (2022) 109778

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Copy move forgery detection and segmentation using improved mask region-based convolution network (RCNN)

Tahira Nazir [a], Marriam Nawaz [b,c], Momina Masood [b], Ali Javed [c,*]

[a] Faculty of Computing, Department of Computer Science, Riphah International University Gulberg Green Campus, Islamabad, Pakistan
[b] Department of Computer Science, University of Engineering and Technology Taxila, 47050, Pakistan
[c] Department of Software Engineering, University of Engineering and Technology Taxila, 47050, Pakistan

ARTICLE INFO

Article history:
Received 2 July 2022
Received in revised form 24 October 2022
Accepted 31 October 2022
Available online 7 November 2022

Keywords:
Copy-move forgery
Deep learning
DenseNet
Manipulation detection
Mask-RCNN

ABSTRACT

Copy-move forgery (CMF) is a common image manipulation approach that uses the information from the same sample to manipulate it with the intent of hiding the required content. Several approaches have been designed for the timely detection of CMF; however, accurate identification of manipulated samples is a complicated job due to the similar capturing conditions of the copied content as the patch is taken from the same image. Moreover, the occurrence of several post-processing attacks i.e., noise, blurring, brightness variations, etc. further enhances the difficulties of the detection approaches. In this work, we attempted to cover the limitations of existing methods by proposing a deep learning (DL)-based approach for the accurate detection of CMF. A custom Mask-RCNN model with the DenseNet-41 as the base network is presented which is capable of nominating a better set of image features and presents the complex image transformation effectively. More descriptively, the DenseNet-41 model is used as the base network for deep keypoints extraction which is then localized, segmented, and categorized by the Mask-RCNN model to locate the manipulated area. We have tested the proposed model on three standard databases namely the CoMoFoD, MICC-F2000, and CASIA-v2 databases, and attained a precision of 98.12%, 99.02%, and 83.41%, respectively. We have reported the results for numerous image post-processing attacks and confirmed that the presented work is robust to detect the CMF in the presence of translation, scale variations, rotation, color changes, noise, compression, and blurring in images. We have confirmed through extensive quantitative and qualitative evaluation that the DenseNet-41-based Mask-RCNN model is robust to CMF detection and can assist forensic analyzers to detect forensic manipulations accurately.

1. Introduction

In this IT era, the economic prices of gadgets (i.e., cellphones, cameras, laptops, etc.) and the easier availability of the internet and social sites have resulted in exponential growth of images in cyberspace. These digital images are one of the main sources of information in multimedia content. Simultaneously, these digital samples are employed as evidence for investigating the legal claims [1,2]. At the same time, numerous easy-to-use apps and software (i.e. Photoshop, PhotoScape, Corel, etc.) have been introduced which can easily change the content of images [3,4]. The introduction of the latest AI-based tool has increased the realism of forged content to such a level that it has become difficult for humans to detect it with their naked eye. Because of such manipulations, it has now become difficult for people to trust digital information. Therefore, it is a need of the hour to introduce reliable forensic methods to assess the integrity of images before employing them in the processing of any legal work. In history, the techniques used for digital image analysis are dived into two types named active and passive techniques. Active approaches are workable for cases for which the prior data about the source sample is available i.e., watermarks and the presence of digital signatures [5,6]. Such methods are not proficient when prior details are missing. For that reason, typically passive methods are heavily explored by scientists which can be used to locate two kinds of digital forgeries named CMF and splicing [7]. To generate the spliced image forgery, the information from two or more images is combined while in the case of CMF, the data from the same sample is utilized to forge the image. It has been found that CMF is more complex than image spicing as the overall characteristics of the image remain the same which complicated the forgery detection process.

The existing approaches used for detecting the CMF either employed machine learning (ML) frameworks or deep learning (DL) architectures. The conventional ML-based methods used for

* Corresponding author.
E-mail address: ali.javed@uettaxila.edu.pk (A. Javed).

https://doi.org/10.1016/j.asoc.2022.109778
1568-4946/

CMF detection (CMFD) are further categorized into two types namely block-based and keypoints-based forgery detection (FD) methods [8]. For block-based FD methodologies, the sample under investigation is categorized into various overlapping or non-overlapping blocks and FD methods are employed for all blocks. The block-based FD methods are exponentially expensive due to performing the FD techniques on each block. While in the case of keypoints-based FD methods, the features are computed from the entire sample [9]. These approaches are computationally effective, however, they are unable to present the whole texture-based information of the sample, which result in a misclassification [10]. Recently, the advancement of DL frameworks has attracted the focus of the research community to utilize them in the field of image forensic investigation [11–13]. The DL-based methods present the image information with more detail and have been found more robust towards the CMFD. As producing the CMF does not require any significant human expertise, therefore, it is very easy for a layman to forge the digital content and alter the conveyed information. Such alterations are imposing a serious threat to the integrity of digital data. Furthermore, several image post-processing attacks are introduced in the forged samples to fool the detectors and complicate the detection process. Such attacks include the incidence of noise, compression, blurring, color, brightness, angle, position, rotation, and scale variations. These attacks hinder the detection approaches to accurately localize the forged area which is a basic need of forensic analysis approaches, as the digital samples can be employed in processing a legal claim or investigating a criminal case. The legitimate utilization of digital samples requires forensic analysis approaches to better explain the reason why a sample is designated as a forge. Therefore, an employed approach for CMF must be capable of effectively locating the manipulated content by exactly showing the copied regions.

Although, extensive work has been presented in the literature to unveil image manipulations, however, still there is a performance gap both in terms of performance efficiency and effectiveness. Moreover, the existing methods are unable to perform well under the presence of post-processing attacks i.e., color and light changes, blurring, scaling alterations, and various noise attacks. In the presented approach, we have attempted to tackle the existing challenges by proposing a custom Mask-RCNN model with the DenseNet-41 base network for deep features calculation of input samples and to locate and classify the samples as original or manipulated. The presented technique is effective to tackle the post-processing operations i.e., the presence of blurring, noise, scale, angle, compression, and color variations in the input samples. The extensive experiments showing both the numeric and visual results have comprehended that the presented method exhibits vigorous performance. The main contributions of the introduced framework are as follows:

i. Presented a custom Mask-RCNN model with a lightweight backbone which improved the accuracy to detect the manipulated content while minimizing both the training and testing time complexity.
ii. Employed DenseNet CNN-based architecture, which uses features collected by very early layers directly in deeper layers and thus allows the accurate identification of forged content due to the robustness of the extracted features.
iii. Proposed an explainable approach that is capable of accurately locating the manipulated region by generating a semantic mask along with the classification score.
iv. Proficient localization, segmentation, and classification performance of the forged region in manipulated samples due to the ability of Mask-RCNN to tackle multiple CMF cases and the over-fitted training data.
v. The presented approach is capable of tackling the post-processing attacks like noise, blurring, compression, scale, angle, position, and brightness variations as well due to the empowerment of the DenseNet-41-based Mask-RCNN model.
vi. Performed extensive experimentation using CoMoFoD, MICC-F2000, and CASIA-v2 databases to show the reliability of the presented method in identifying single and multiple CMF cases from digital images.

The remaining manuscript is divided as follows: Section 2 contains related work while the detailed technical description of the proposed methodology is given in Section 3. The performance evaluation parameters and experimental results of the proposed method are discussed in Section 4 and finally, the conclusion is presented in Section 5.

2. Related work

In this section, we have reviewed the existing work performed to detect the CMF from digital images. Tinnathi et al. [14] presented an approach to detect the CMF from input samples. In the first step, the given image was divided into non-overlapped regions by utilizing an adaptive watershed segmentation technique. In the next step, an approach namely Hybrid Wavelet Hadamard Transform (HWHT) was used over the segments to compute the features. Then, the similarity was measured by using the adaptive thresholding method, and outliers were eliminated by applying the Random Sample Consensus (RANSAC) algorithm. At last, the Forgery Region Extraction Algorithm (FREA) was applied to identify the manipulated region from the samples. The approach in [14] shows better performance for CMFD, however, it suffers from high computational time complexity. Lyu et al. [15] presented a solution for CMFD by utilizing the concept of double matching. Initially, the Local Intensity Order Pattern (LIOP) was used to compute the keypoints of an input image. Then, Delaunay Triangulation (DT) was applied over the calculated features and for every Delaunay triangle, the mean vector was computed by using the LIOP descriptors of three vertices. In the next step, the g2NN algorithm was used to accomplish the triangles matching with a static threshold. Then in the second matching phase, the g2NN with a looser threshold was applied to perform the features matching to locate the exact region of interest. Finally, RANSAC was employed for eliminating the wrong matches. The technique [15] is robust to CMFD, however, it was unable to perform well in the presence of intense color variations in the altered sample. Another CMFD technique was presented in [16], where the author proposed an optimized SIFT algorithm to compute the features from the digital image. The coarse and fine-grained methods were applied to compute the content matching. The method [16] exhibits better CMFD results, however, unable to tackle the feature's high dimensional space. Agarwal et al. [17] presented a method to locate forensic manipulations from digital samples. Initially, the fuzzy c-means (FCM) clustering approach was applied to divide the image into small blocks. Then, the Gabor filter was employed to compute the keypoints from each block. Finally, a matching process was performed to show the altered content. The technique [17] works well for CMFD with several post-processing attacks, however, detection accuracy needs further improvement. Another CMFD framework was presented in [18], where SIFT features were gathered over CLAHE applied sub-images. Then, after combining the matched features, the RANSAC method was applied to remove the false matches. The model in [18] exhibits better CMFD accuracy over the scaling, compression, and noise post-processing attacks. However, the detection accuracy degrades for images with strong light variations.

Nawaz et al. [19] proposed a framework to perceive the changes developed within digital images. After pre-processing, the stationary wavelet transform (SWT) technique was used over the input sample to get a set of shift-invariance local keypoints. Then, speeded-up robust features (SURF) algorithm was applied to extract the keypoints on which scaled density-based spatial clustering of applications with noise (sDBSCAN) clustering method was applied to group the similar objects. Finally, Euclidean distance was estimated to compute the resemblance of clusters. The technique in [19] performs well for CMFD under the presence of post-processing attacks, however, unable to perform well when manipulations are made within flat regions of input images. Lin et al. [20] introduced a methodology to detect image manipulations by using hybrid features. First, the features from the suspected sample were computed by employing LIOP and SIFT descriptors. Then transitive matching was applied to enhance the matching relation of computed features. After this, a filtering technique based on an image segmentation procedure was used to detect incorrect matches. Finally, affine transformations among detected keypoints were computed to identify the forged content. The approach in [20] exhibits better forgery detection performance under various transformation operations like scaling, blurring, etc., however, this method is computationally more complex. Chen et al. [21] presented an approach for the CMFD. Initially, image features were calculated by using the SIFT descriptor. Then Hu's invariant moments were employed to identify the initial matching blocks. In the next step, the alignments of detected similar points were adjusted for having similar orientations. Finally, a region growing technique based on Hu's invariant moment estimations was applied to grow the manipulated areas. The approach in [21] shows better CMFD results, however, for tiny and thin copied areas, the detection performance degrades. Moreover, the SIFT approach is unable to perform well for flat forged regions. Roy et al. [22] introduced a method to locate digital image forgeries. In the first step, the SURF descriptor was applied to identify the image keypoints, on which the Rotated Local Binary Pattern (RLBP) is applied through the circular neighborhood to calculate the feature vector. Then g2NN keypoints similarity measurement approach was applied to compute the similarity. Finally, hierarchical agglomerative clustering was utilized to show the manipulated content. This technique shows better CMFD performance, however, the performance degrades in those scenarios where alterations are made within smooth areas. Meena et al. [23] introduced a method to detect image forgeries. Initially, the suspected sample was divided into overlapped blocks. Next, the Tetrolet transform was employed to locate the four low-pass and twelve high-pass coefficients from all small blocks. In the next step, the computed keypoints were arranged lexicographically and a threshold-based similarity method was applied to detect the copied regions. The approach in [23] shows better CMFD performance, however, its detection accuracy degrades over the noisy and distorted samples.

Nowadays, as DL-based approaches are getting the intense attention of researchers, therefore, some of the research work employing DL techniques has been reported in the field of CMFD as well. One such framework was reported in [24] to locate the copy-move forgery attacks in digital images. After pre-processing step, a CNN-based model was utilized to calculate the keypoints vector of an input image which were then categorized as original or manipulated. The approach [24] works well for CMFD, however, not effective for images with resizing and rescaling post-processing attacks. Similarly, Wang et al. [25] introduced a DL technique namely Mask-RCNN to locate and segment the manipulated regions of a suspected sample. The work in [25] shows better CMFD accuracy, however, unable to perform well under the occurrence of post-processing attacks due to limited feature representation capacity. Zhong et al. [26] presented a DL approach named Dense-InceptionNET for identifying the CMF. First, Pyramid Feature Extractor (PFE) component was utilized to learn the deep keypoints from the input image. Then Feature Correlation Matching (FCM) component was used to measure the correlation of calculated features to identify the possible manipulated image patches. Finally, a Hierarchical Post-Processing (HPP) component was applied to eliminate the wrong matches. The work in [26] improves the CMFD performance, however, it is a computationally costly approach. Zhu et al. [27] presented a DL technique for identifying digital image manipulations. In this technique, a fully connected neural network was introduced, comprised of Adaptive attention and Residual refinement Network (AR-Net). Initially, an adaptive attention module was employed to learn the deep features. In the next step, deep matching was utilized to calculate the self-correlation among keypoints maps. Then Atrous Spatial Pyramid Pooling (ASPP) fused the computed maps to produce the coarse mask. Finally, the residual refinement component was applied to maintain the shape of the boundaries of objects. This approach improves the CMFD performance under the presence of various transformation operations, however, it is computationally complex. Another DL-based approach was presented in [28], where a CNN framework namely Faster-RCNN was applied to detect the forensic changes in images. The technique [28] is robust in locating the digital manipulations, however, only a few results are reported. Tahaoglu et al. [29] proposed a method for detecting CMF from digital images. Initially, the SIFT algorithm was applied to acquire the textual information of the input images. After this, a keypoint matching approach was used to expose the manipulations made in the suspected images. Finally, the Ciratef approach was used to locate the manipulated areas. The work in [29] shows better CMF recognition ability, however, it is unable to perform well for images with intense brightness changes. Yue et al. [30] presented a framework for locating the CMF from the input images. The feature vector of the input sample was computed using the SIFT approach. After this, a feature matching approach was applied to measure the similarity among features to locate the copied areas. In the next step, the AdaLAM algorithm was used to segment the detected forged area. The work demonstrated in [30] works well for identifying the CMF, however, unable to show better results for samples with huge-angle alterations. Another work for CMF detection was presented in [31] where a custom SIFT approach was used to measure the feature vector of the samples. After this, a new Feature Label Matching method was used to minimize the feature space. Finally, the Hierarchical Segmentation Filtering technique was used to filter the outliers. This work [31] is effective for CMF detection, however, suffers from a high computational cost. In [32], the authors presented a CNN model for the identification of CMFD. The presented CNN comprises six convolutional layers to learn hierarchical keypoint representation from the input samples and categorizes them into authentic and forged groups. This approach shows improved performance, however, it lacks generalization because of the dependence on training data. In [33], the authors presented a custom CNN based on the VGG-16 model and employed transfer learning to improve the generalization performance for CMFD. This approach showed improved generalization, however, it is computationally costly and has a long inference time.

From the conducted investigation of existing studies, it can be concluded that although extensive work has been presented to perform the forensic analysis of digital images, however, still there is a requirement for performance enhancement. The generation of realistic datasets and the easier availability of user-friendly editing tools are increasing the complexities of the detection process. Moreover, the incidence of several image post-processing attacks like blurring, noise, compression, and variations in the size, position, angle, and brightness of suspected samples are further increasing the difficulties of existing approaches

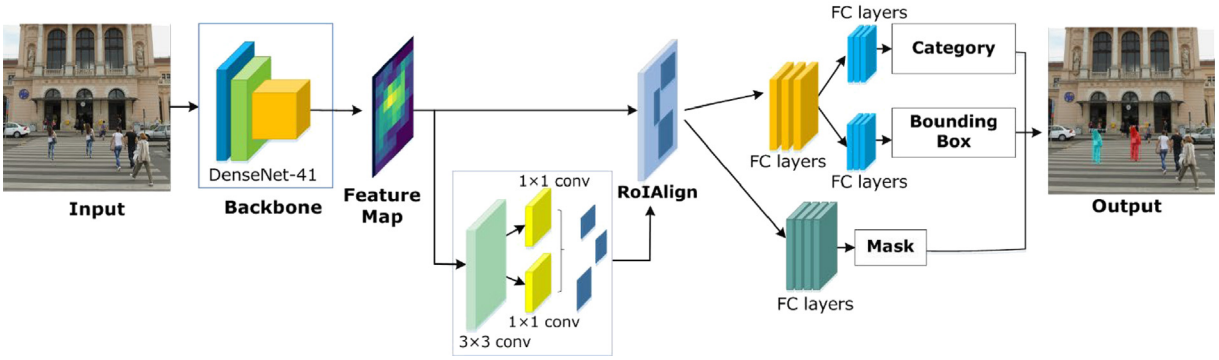

Fig. 1. Visual depiction of the proposed Model.

to accurately specify the forged regions in digital images. So, there is a need to develop a more reliable CMF detection framework.

3. Proposed method

In this part, we have described the structure of the employed approach for CMFD. The entire pipeline of our method is shown in Fig. 1. More specifically, we have used the Mask-RCNN approach with DenseNet-41 as the backbone architecture to detect and segment the forged content from the digital samples. For a given suspected sample, our objective is to automatically locate the manipulated portion without the need of performing any manual inspection. The proposed method comprises the following steps: (i) ground truth generation, (ii) feature extraction with DenseNet-41, (iii) forged Region of interest (RoI) generation, (iv) RoI localization along with the bounding boxes (bbox) creation, and (v) segmentation mask attainment. Initially, the ground truths are generated which are later used for network training. Then, for a given test sample, the custom Mask-RCNN model comprising DenseNet-41 as a keypoints calculator is used for computing the deep features. Then, the acquired deep keypoints are passed to the region proposal network (RPN) to produce RoIs by plotting all features on the feature map into the actual sample. Then, the RoIAlign module is utilized to align keypoints from the feature map related to the RoIs attained from the RPN module and passed to the Fully Connected Network (FCN) layers to identify and segment the CMF. Finally, the trained model is tested on two datasets with post-processing attacks to show the robustness of the proposed approach. We have reported both the numeric and visual results to explain the CMFD ability of our approach for single and multiple CMF attacks. Algorithm 1 specifies the details of steps followed by the improved Mask-RCNN model.

3.1. Problem formulation

The multitude of activities involved in the area of forensic analysis may potentially be made visible by focusing on the process of problem formulation in real-world applied settings, although this has not yet been the subject of significant empirical research. Several attempts have been made by researchers for the timely detection of CMF attacks, however, the increased realism of fake content is increasing the complexities of accurate recognition of such forgeries in digital images. Moreover, the existence of post-processing attacks like color and light changes, blurring, scaling alterations, and various noise attacks further complicate the detection process. Furthermore, techniques lack to identify the multiple CMF attacks in the digital samples. In this work, we have tried to fill this gap. How and why are specific problems posed? What challenges arise and how are they resolved in everyday practice? Answers to these questions, we argue, can help us to better understand image forensic analysis as a practice, but also the origin of the manipulation techniques that raise normative concerns. As researchers work to unpack the forgery detection process, we have proposed a deep learning model for CMFD lending to make visible the work of problem formulation in applied contexts. In so doing, we show how to trace the ethical implications of these systems back to the everyday challenges. To solve this problem, we have selected three challenging datasets which contain diverse samples. We have divided the employed datasets into two sets namely the train and test sets with a ratio of 70:30 respectively. The test samples are unseen and do not appear during the model training phase.

To detect the CMFD, we have given the images belonging to the CMF. The detection system generates the deep features of all images I_i in which $i = 1,2,3,\ldots n$. Each image contains two types of areas i.e., forged as 1 and real as 0. We have labeled and generated the bounding boxes according to the 0 and 1 areas. This repository of images is divided into train set *Tr* and test set *Ts*. The input images I_i and the bounding boxes are passed to the next phase for ground truth generation.

3.2. Ground truth generation

To perform accurate model training, it is mandatory to develop the Ground Truth (GTr) mask for all training samples to specify the forged regions. For this purpose, the input *Ii*, bounding boxes, *Tr*, and *Ts*. from the previous module are utilized for ground truth generation according to our proposed model. We have used the VGG Image Annotator (VIA) [34] to generate the annotations of the forged images by creating their respective polygon masks. A few samples of generated GTr are demonstrated in Fig. 2. A JSON file is created to save the readings of the VIA which contains the values of the polygon points for the forged portion and the corresponding region attribute value as 0 or 1. More specifically, all the pixel values found under the area of the bounding polygon being the part of forged content are set to 1. Whereas the remaining area is nominated as background (bg) and set to 0. The created JSON file is employed to develop a mask sample related to all forged training images which are later used to train the proposed framework.

3.3. CMFD detection and segmentation with custom Mask-RCNN

The Mask-RCNN is the latest DL-based approach that is investigated for object recognition and accomplishing pixel-level segmentation by the research community [35]. The Mask-RCNN [36] approach is an extended form of the Faster RCNN model [37], which is capable of identifying, localizing, and segmenting the RoI. The Mask-RCNN approach has less time and structural complexity than the faster-RCNN model as it introduces a minimum

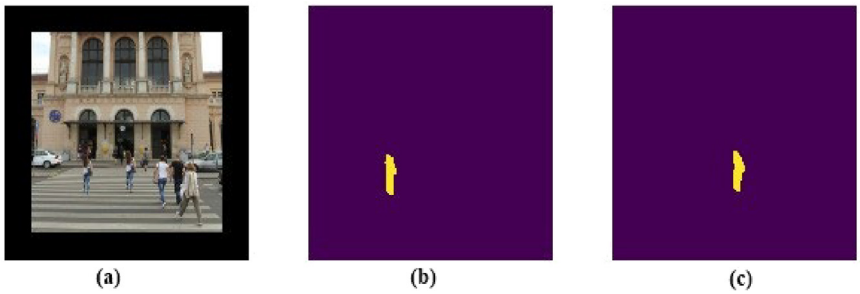

Fig. 2. Example of the input image (a) and related ground truth masks where the first mask (b) is related to the forged area and later (c) is related to the original.

model overhead, i.e., an FCN unit to accomplish the segmentation task. In this work, we present a modification of the Mask-RCNN by using the DenseNet-41 as its base network. DenseNet [38,39] is a well-established CNN-based approach that works by using the data from all proceeding layers. The DenseNet model consists of several dense blocks (DBs), where all DBs are consecutively interweaved via using additional convolutional and pooling layers. DenseNet framework has the potential to demonstrate the complicated transformations of the image effectively which in turn enhances the performance. The densely connected DBs better tackle the issue of the absence of information about the target's locality for the top-level keypoints to some degree. Further, DenseNet utilizes a small set of model parameters that also provides a computational benefit as well. Furthermore, the DenseNet approach contributes to an easier flow of image features and inspires their reuse, which presents them more effective for CMFD identification and segmentation. A detailed explanation of the entire technique is discussed in the succeeding sections.

3.3.1. Feature extraction

In this phase, *Ii, Tr, Ts, GTs,* and Bounding boxes are essential to detect the deep features of CMFD. For this purpose, these inputs are fed to the feature extractor to train the model for CMFD detection. In all object detection methods, a backbone network is used to learn the reliable set of keypoints vectors [40]. The base framework is any CNN approach envisioned for sample examination, i.e., VGG-16, VGG-19, ResNet-101, and DenseNet. A feature computation approach must be extremely effective to learn the important features of a sample and efficient to reduce the computational cost. The conventional Mask-RCNN approach usually employs the different variants of VGG or ResNet as a base network [41]. However, the complex and deep network structures of these base frameworks often cause to increase the computational burden of these models and complicate the model optimization process. Such complex feature extraction models result in the vanishing gradients problem. Moreover, the conventional feature extractor like ResNet uses skip connections which fail to learn all aspects of image features. Therefore, to deal with the issues of existing feature extraction techniques, we have presented the DenseNet-41 as the base backbone for the Mask-RCNN. The DenseNet model uses dense links which assists it to extract the reliable set of sample keypoints. The introduced DenseNet-41 framework is more robust than the traditional DenseNet because of the following reasons: (i) DenseNet-41 is a lightweight model in comparison to conventional DenseNet as the DenseNet-41 utilizes 24 channels in place of 64 on the first convolution layer along with the kernel window of 3×3 instead of 7×7; and (ii) the number of layers within each dense block in DenseNet-41 is reduced to deal with the computational complexity. A pictorial illustration of DB is shown in Fig. 3 while layer details are presented in Table 1. The DenseNet-41 network architecture comprises multiple convolutional layers, four dense blocks, and three transition layers. The dense block contains convolution layers with kernel window of size 3×3 and 1×1 that are connected with each other. Between dense blocks, transition layers are added to decrease the dimension of features.

Table 1
The architecture details of DenseNet CNN.

Blocks *(i)*	Layer type	Kernal size and number of layers	Stride value
1	Convolutional layer	7×7	2
2	Pooling layer	3×3 avgpool	2
3	Dense-Block 1	$\binom{1\times1 \text{ conv}}{3\times3 \text{ conv}} \times 3$	1
4	Transition Layer 1	1×1 conv 2×2 avg_pool	1 2
5	Dense-Block 2	$\binom{1\times1 \text{ conv}}{3\times3 \text{ conv}} \times 6$	1
6	Transition Layer 2	1×1 conv 2×2 avgpool	1 2
7	Dense-Block 3	$\binom{1\times1 \text{ conv}}{3\times3 \text{ conv}} \times 6$	1
8	Transition Layer 3	1×1 conv 2×2 avg_pool	1 2
9	Dense-Block 4	$\binom{1\times1 \text{ conv}}{3\times3 \text{ conv}} \times 3$	1
10	Classification layer	7×7 avgpool Fully Connected layer	

3.3.2. RPN unit

Here, the computed features from the DenseNet-41 are passed as input to the RPN unit to produce RoIs. The calculated RoIs are later used to recognize the forged content to accomplish the final segmentation task. The RPN unit consists of a 3×3 convolution layer to analyze the entire suspected sample to produce the required anchors [42]. The anchors are the bbox in the actual, varying in dimensions and scattered on the entire image. The RPN approach generates about 20k bboxs initially, which are intersected with each other, however, the RPN prediction module nominates the top anchors with the largest foreground (fg) value by employing the bbox regressor. Descriptively, the bbox with Intersection-over-Union (IoU) more than 0.7 are taken as the positive anchors (fg class), and the remaining are marked as negative (bg class). Such behavior of the RPN unit assists to select the more reliable set of RoIs that are used in the later stages of the model to accomplish the CMF detection and segmentation.

3.3.3. ROI classifier & Bbox regression

This module accepts the RoIs along with the feature maps as input (Fig. 4). This unit recognizes the forged region by discriminating them from the unforged image areas via improving the bbox estimation by pooling the RoIs to the fix-sized feature maps. More specifically, the RoIs areas do not match with the granularity of the feature maps as these are down-sampled k times as compared to the actual image size (via convolutions). Therefore, a ROIAlign layer is used to acquire the fix-sized feature vectors from the primary regions of varying sizes and to

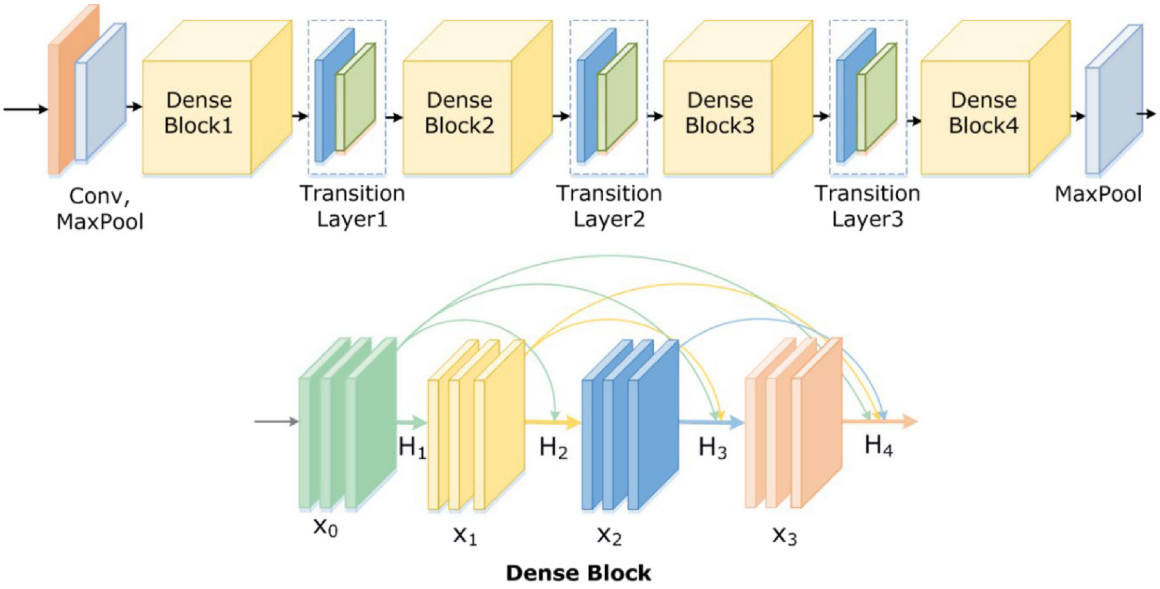

Fig. 3. Visual depiction of DenseNet-41 architecture and Dense Block.

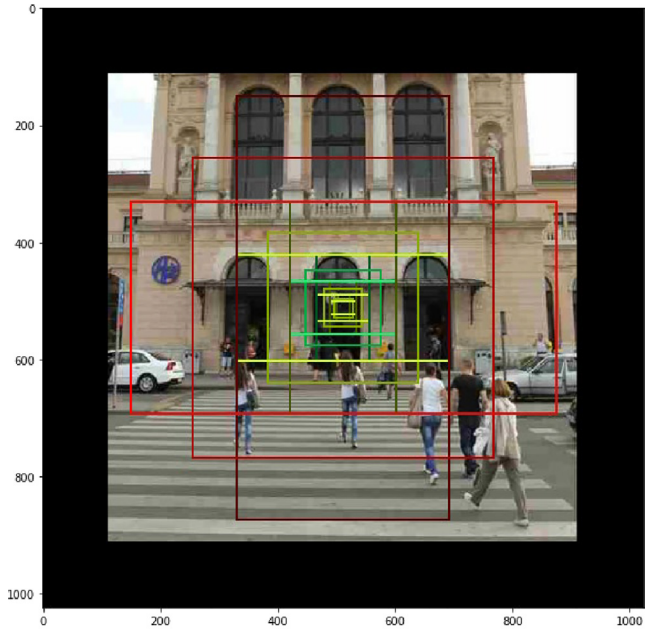

Fig. 4. Visual representation of the generated RoIs.

reset the feature map's size. The ROIAlign layer uses the bilinear interpolation operation to resolve the misplacement issue that occurred within the RoI pooling layer and, finally, the classifier and regression layers are used to get the ultimate localized CMF results.

3.3.4. Forgery segmentation

It is the last step of the proposed approach which accepts the positive areas (manipulated content) selected by the RoI classifier as input and generates the segmentation mask. The outcome masks are shown by floating numbers which assist to show detailed information in comparison to a binary mask. Initially, the GT_r masks are resized to a dimension of 28×28 to compute the loss against the estimated segmentation mask in the training phase. The scaled-down mask is later scaled up in the inference stage to match its size with Bbox and the final mask is generated to show the CMF content in the image.

3.4. Loss function

To optimize the CMFD performance, the proposed approach uses several loss functions [43] during the training process which are defined as:

$$L_{total}(CustomMask - RCNN) = L_c + L_b + L_m \tag{1}$$

Here, L_c, L_b, and L_m are showing the class label estimation, bbox enhancement, and segmented mask approximation losses, respectively. The L_c is computed as:

$$L_c = -\log p_v \tag{2}$$

Here, p is a $(d{+}1)$ dimensional vector related to the probability of a keypoint either belonging to a d class or bg. For all RoIs, $p = p_0, \ldots, p_d$ and p_v is showing the probability associated with a class v (Original/ manipulated). The L_b is computed as:

$$L_b(s_j, s_j^*) = \sum_{i \epsilon \{x,y,w,h\}} smooth_{L1}(s_i - s_i^*) \tag{3}$$

where,

$$smooth_{L1}(X) = \begin{cases} 0.5X^2 & \text{for } |X| < 1 \\ |X| - 0.5 & \text{else,} \end{cases} \tag{4}$$

More specifically, the Vector s_j is demonstrating four coordinate values of the estimated Bbox, while the s_j^* is showing the coordinate values of GT_r for all the positive anchors. The smooth-$L1$ method is effective and less sensitive to outliers in comparison to $L2$ loss. Finally, the segmentation mask is computed as:

$$L_m = -\frac{1}{N^2} \sum_{1 \le j,k \le N} \left[X_{jk}.logp_{jk}^d + \left(1 - X_{jk}\right).log\left(1 - p_{jk}^d\right) \right] \tag{5}$$

Here, X_{jk} is showing the value of a keypoint (j, k) in a GT_r mask with the dimension of $N \times N$ and p_{jk}^d shows the approximated value of the same keypoint for the learned mask.

4. Results and discussion

To analyze the CMF detection ability of the presented framework, we have evaluated the Custom Mask-RCNN with various

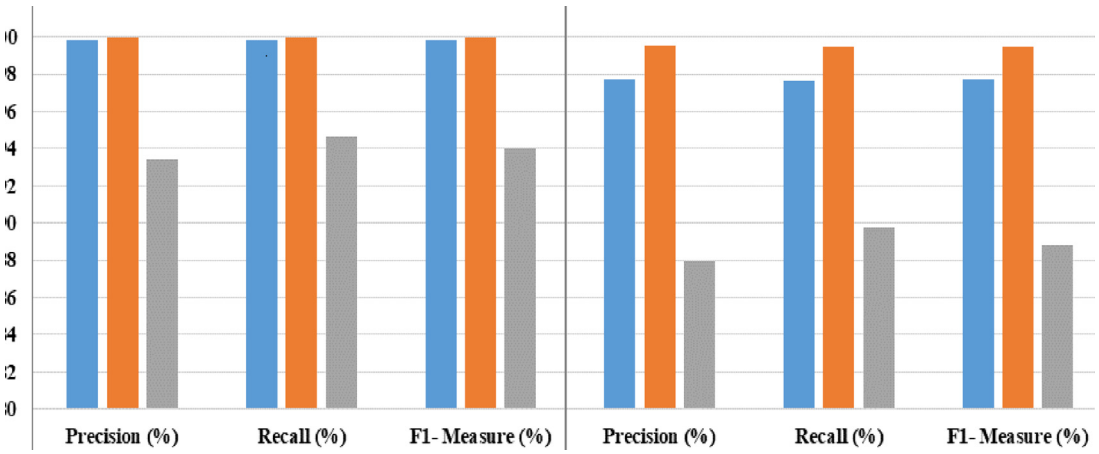

Fig. 5. Performance of the proposed method during model training and validation.

Table 2
Details of network implementation parameters.

Parameters	Value
Total used epochs	30
Value of learning rate	0.001
Selected batch size	6
The threshold value used for the confidence score	0.5
Value of Unmatched Threshold	0.5

Table 3
Distribution of the employed datasets used for training and testing.

	CoMoFoD		MICC-F2000		CASIA	
	Real	Forged	Real	Forged	Real	Forged
Total	260	13520	1300	700	7491	3274
Training	182	9464	910	490	5244	2292
Testing	78	4056	390	210	2247	982

post-processing operations i.e., blurring, noise, compression, scaling and rotational changes, etc. The model is implemented using Python with TensorFlow and executed on the Nvidia GTX1070 GPU-based system in a windows based operating system environment. For the training of the proposed model, we have used the pre-trained model on the MS-COCO database, which is then further trained using transfer learning on the CMFD datasets. The details of training parameters are defined in Table 2. We selected these parameters in our model as optimal results were attained under these settings. Moreover, the training and validation behavior of the proposed method for all employed datasets is given in Fig. 5, which is clearly showing the better learning behavior of the model.

4.1. Dataset

To validate the manipulation recognition performance of the proposed solution, we have utilized the CoMoFoD dataset [44] which is the largest and standard dataset of CMF. Moreover, we have also used the MICC-F2000 [45] and CASIA-V2,46 datasets for model evaluations.

In the CoMoFoD database, all the samples are manipulated by copying and pasting the content from the same sample to produce the CMF attacks. All the images contain five kinds of transformations which are: (i) translation: where no other transformation is employed and only object translation is performed, (ii) Scaling: in which the forged portion is both resized up or down and translated as well, (iii) Distortion: here the pasted content is distorted in its visual quality and translated as well, (iv) Rotation: where the copied object is rotated and translated as well, and (v) Combination: where two or more transformations are performed simultaneously. The above five types of image transformations contain 40 images and further, each sample has seven versions. Besides the case with no post-processing attacks, the samples are further manipulated with six types of post-processing attacks which are the presence of blurring, noise, compression, intensity variations, chrominance changes, and contrast adjustments. Further details about the dataset can be found in [45]. While the MICC-F2000 database comprises a total of 2000 samples, where 1300 samples are real and the remaining 700 are manipulated samples with a resolution of 512×512. The CASIA-V2 [46] database contains a total of 7491 colored original samples and 3274 copy-move forgery samples. It includes some uncompressed images in TIFF and BMP format, while others are in JPEG format with different compression levels. The size of samples in pixels varies from 320×240 to 800 × 600. For implementation, we have divided each dataset into 70% for training and 30% for testing separately (Table 3).

4.2. Performance metrics

To evaluate the CMFD performance of the introduced method, we employed the three measures namely Precision, Recall, and F1-score. Precision shows the percentage of predicted forged samples that are actually forged (i.e., locate samples that are already manipulated). Recall shows the percentage of actual forged samples that are properly predicted forged (i.e., the number of returned samples being identified as manipulated from all the forged). F1-score is a composite measure to calculate the accuracy of a forgery detection technique as it employs both Precision and Recall.

4.3. Evaluation of the proposed approach

We designed an experiment to compare the CMF performance of the proposed approach with the base models. We have tested the CMF detection performance of the Mask-RCNN model with base networks like VGG16 [47], VGG19 [48], ResNet50 [49], and ResNet101 [50], and attained results are shown in Table 4. The values shown in Table 4 are clearly exhibiting that the presented framework is more accurate to identify the CMF attacks in digital images as compared to other base models in terms of all employed performance measures. Moreover, the proposed approach is computationally efficient from all other base methods due to a smaller number of model parameters. The major reason for the enhanced CMF detection and classification performance of the

Table 4
Comparison of the proposed model with base approaches over the CoMoFoD dataset.

Model	Parameters (M)	Precision (%)	Recall (%)	F1-Score (%)	Time (s)
Mask-RCNN with VGG16	140.5	92.34	90.08	91.20	80
Mask-RCNN with VGG19	164.4	92.17	90.11	91.13	83
Mask-RCNN with ResNet-50	44.6	93.81	92.20	93.00	49
Mask-RCNN with ResNet-101	63.5	94.69	94.16	94.42	51
Proposed	27.1	**98.12**	**95.85**	**96.97**	**45**

proposed work is due to the reliable feature computation ability of the DenseNet-41 model which assists it in better capturing the underlying patterns in a viable manner. While in comparison the other base models like the VGG model are suffering from the high computational cost and model overfitting problem, while the ResNet models employ skip connection which fails to learn all the aspects of sample keypoints and results in performance degradation. The presented DenseNet-41 model better deals with the limitations of existing base models by employing dense layers which are capable of extracting a more reliable set of sample features and maintaining the computational complexity of the model as well by using a small number of framework layers.

4.4. Performance testing

We have selected the CoMoFoD dataset to compare the detection accuracy with the latest techniques due to its challenging nature. For this reason, we have designed seven types of experiments one with no post-processing attacks and the remaining with the six above-mentioned post-processing operations in the description of the dataset. To validate the forgery recognition performance of custom Mask-RCNN, we have chosen three studies [7,51,52] from the history using the same dataset with similar performance testing strategies as mentioned in [52]. To have a fair comparison, we have chosen both the ML and DL-based approaches to show the robustness of the presented framework. The approach [7] initially divided the input sample into multiple segments and then followed a two-step process to locate the manipulated regions. While the method in [51] worked without dividing the image into portions and computed the SURF descriptor-based features from the entire image on which the matching and clustering techniques were applied to locate the altered content. Whereas in [52], a DL approach namely Convolutional Kernel Network (CKN) was applied to learn the deep keypoints from the input samples and identify the forged areas. The comparative results with selected studies against all post-processing operations are elaborated in the subsequent sections.

4.4.1. Without post-processing attack

Initially, we experimented to validate the robustness of our work for the CMFD under the presence of image transformation operations only and visual results are reported in Fig. 6. It is depicted from Fig. 6 that our approach is proficient to detect all types of CMF attacks due to its effective localization power, which enables it to locate the varying forensic changes. To compare the results with the latest techniques, we have performed a performance analysis in Table 5. From the values shown in Table 5, it can be seen that the presented Custom Mask-RCNN performs better than the approaches in [7,51,52]. More specifically, for all transformation operations, the presented approach obtains an average value of 0.9792 for the precision metric while the methods in [7,51,52] show an average precision value of 0.556. Therefore, we can say that in the case of precision, the custom Mask-RCNN shows a 42.32% performance gain. While in the case of a recall, our technique shows an average value of 0.9548, whereas the other methods acquire an average value of 0.8103. Therefore, the Custom Mask-RCNN exhibits a 14.45% performance gain for the recall metric. For the F1 measure, the custom Mask-RCNN obtains an average value of 0.9668, whereas the competent methods show an average value of 0.5901, so, our work exhibits a 37.68% performance gain. From the reported results, it can be said that our approach is more robust to CMFD because of its power to better tackle the over-fitted network training data.

Table 5
Comparison with competent approaches for samples without post-processing attacks.

Reference	Transformation operation (over 40 images)	Precision (average)	Recall (average)	F1-Measure (average)
Li et al. [7]	Translation	0.4180	0.8327	0.4798
	Rotation	0.5594	0.8281	0.5978
	Scaling	0.5542	0.8492	0.6059
	Distortion	0.6425	0.9045	0.6961
Average	All operations	0.5435	0.8536	0.5949
Silva et al. [51]	Translation	0.4921	0.7754	0.5493
	Rotation	0.5532	0.7451	0.5573
	Scaling	0.4966	0.6971	0.5008
	Distortion	0.5814	0.7878	0.5650
Average	All operations	0.5308	0.7513	0.5431
Liu et al. [52]	Translation	0.4547	0.8023	0.5246
	Rotation	0.6833	0.9006	0.7174
	Scaling	0.5696	0.7516	0.5864
	Distortion	0.6631	0.8516	0.6987
Average	All operation	0.5926	0.8265	0.6317
Proposed	Translation	0.9787	0.9421	0.9600
	Rotation	0.9689	0.9647	0.9667
	Scaling	0.9795	0.9556	0.9674
	Distortion	0.9897	0.9568	0.9729
Average	All operation	0.9792	0.9548	0.9668

4.4.2. Brightness variation attack

The accurate CMFD system should be enabled to locate the manipulations from the digital images under the presence of severe intensity changes. Therefore, we have conducted an experiment to check the CMFD power of the presented solution for images with intense light variations, and pictorial results are shown in Fig. 7. One can visualize that the custom Mask-RCNN can accurately locate the forensic changes from the suspected samples under the occurrence of brightness changes due to its effective feature computation power. Moreover, the comparison with the latest methods is discussed in Table 6 where it is quite evident that our approach is more rigorous than the peer methods. For all transformation operations along with the presence of brightness variations, the custom Mask-RCNN shows 46.27%, 19.072%, and 41.95% performance gains for precision, recall, and F1-measure, respectively which shows the robustness of our technique.

4.4.3. Color reduction attack

We performed an experiment to test whether the proposed work can locate the CMF attacks from the samples suffering from the color reduction post-processing attack. To accomplish this, we have taken the images with the intense color reduction from the CoMoFoD dataset and evaluated them on the trained Mask-RCNN

Fig. 6. Visual representation of the CMFD with DenseNet41-Based Mask-RCNN for samples without post-processing attacks.

Fig. 7. Visual representation of the CMFD with DenseNet41-Based Mask-RCNN for samples under brightness variation post-processing attack with varying lower bound (LB) and upper bound (UB).

Table 6
Comparison with competent approaches for samples under the presence of brightness variation post-processing attack.

Reference	Transformation operation (over 40 images)	Precision (average)	Recall (average)	F1-Measure (average)
Li et al. [7]	Translation	0.3957	0.7942	0.4623
	Rotation	0.5601	0.8445	0.5933
	Scaling	0.5537	0.7860	0.5926
	Distortion	0.6464	0.8964	0.6892
Average	All operations	0.5389	0.8303	0.5843
Silva et al. [51]	Translation	0.4157	0.7429	0.4775
	Rotation	0.5272	0.7314	0.5333
	Scaling	0.3977	0.6305	0.4378
	Distortion	0.4834	0.7168	0.5137
Average	All operations	0.4560	0.7054	0.4906
Liu et al. [52]	Translation	0.4128	0.7848	0.4773
	Rotation	0.6075	0.9063	0.6531
	Scaling	0.5350	0.7290	0.5526
	Distortion	0.6342	0.7814	0.6609
Average	All operation	0.5474	0.8004	0.5859
Proposed	Translation	0.9684	0.9768	0.9726
	Rotation	0.9776	0.9684	0.9729
	Scaling	0.9819	0.9695	0.9757
	Distortion	0.9796	0.9629	0.9712
Average	All operation	0.9769	0.9694	0.9731

Table 7
Comparison with competent approaches for samples under the presence of Color reduction post-processing attack.

Reference	Transformation operation (over 40 images)	Precision (average)	Recall (average)	F1-Measure (average)
Li et al. [7]	Translation	0.3940	0.8361	0.4698
	Rotation	0.5692	0.8340	0.6174
	Scaling	0.5628	0.8969	0.6251
	Distortion	0.6352	0.9226	0.6955
Average	All operations	0.5403	0.8724	0.60195
Silva et al. [51]	Translation	0.4872	0.7884	0.5440
	Rotation	0.5432	0.6567	0.5066
	Scaling	0.5004	0.7092	0.5129
	Distortion	0.6147	0.7584	0.5813
Average	All operations	0.5364	0.7282	0.5362
Liu et al. [52]	Translation	0.4502	0.8448	0.5203
	Rotation	0.6583	0.8415	0.6891
	Scaling	0.5984	0.7832	0.6267
	Distortion	0.5984	0.7832	0.6267
Average	All operation	0.5763	0.8132	0.6157
Proposed	Translation	0.9715	0.9585	0.9649
	Rotation	0.9798	0.9518	0.9656
	Scaling	0.9849	0.9599	0.9722
	Distortion	0.9786	0.9637	0.9711
Average	All operation	0.9787	0.9585	0.9685

model. The obtained pictorial results are exhibited in Fig. 8, where it can be witnessed that the proposed solution can accurately locate the forensic manipulations. Moreover, the performance comparison with comparative techniques is presented in Table 7. The results are clearly depicting that our custom Mask-RCNN model is more accurate for CMFD under the color reduction post-processing attack as compared to its competitors [7,51,52] by exhibiting the performance gains of 42.77%, 15.39%, and 38.39% for precision, recall, and F1-measure, respectively.

4.4.4. Noise addition attack

It is impossible in real-world scenarios to avoid the occurrence of noise at the image acquisition step. Therefore, a CMFD framework should be capable of identifying the CMF attacks from the noisy samples as well. To check this, we conducted an analysis to verify the CMFD power of our model under the occurrence of noise in the input samples. The noisy images from the CoMoFoD dataset are tested and visual results are shown in Fig. 9. It is quite evident from the reported results (Fig. 9) that the Custom Mask-RCNN can easily cope with the noise attack in the images and can accurately identify the CMF attacks. Moreover, performance comparison with the comparative approaches is shown in Table 8. More precisely, in the case of precision, the custom Mask-RCNN shows an average performance gain of 39.31%, while for recall and F1-measure, it attains an average performance gain of 24.86%, and 39.35%, respectively. The reported values clearly demonstrate that the presented solution is vigorous to detect the CMF attacks even for noisy images due to the effective localization power of the custom Mask-RCNN.

Table 8
Comparison with competent approaches for samples under the presence of noise addition post-processing attack.

Reference	Transformation operation (over 40 images)	Precision (average)	Recall (average)	F1-Measure (average)
Li et al. [7]	Translation	0.4636	0.7563	0.5211
	Rotation	0.6202	0.8399	0.6528
	Scaling	0.5673	0.7438	0.5849
	Distortion	0.6806	0.7821	0.7013
Average	All operations	0.5829	0.7805	0.6150
Silva et al. [51]	Translation	0.4035	0.4550	0.3170
	Rotation	0.5924	0.6481	0.5265
	Scaling	0.6159	0.5115	0.4987
	Distortion	0.6828	0.5627	0.5270
Average	All operations	0.5736	0.5443	0.4673
Liu et al. [52]	Translation	0.5097	0.7819	0.5623
	Rotation	0.6385	0.8076	0.6578
	Scaling	0.5838	0.6840	0.5677
	Distortion	0.7380	0.8411	0.7627
Average	All operation	0.6175	0.7786	0.6376
Proposed	Translation	0.9776	0.9389	0.9578
	Rotation	0.9858	0.9539	0.9696
	Scaling	0.9889	0.9487	0.9684
	Distortion	0.9857	0.9578	0.9715
Average	All operation	0.9845	0.9498	0.9668

4.4.5. Compression attack

Another analysis is performed to check the CMFD accuracy of the presented work under the presence of the JPEG compression attack. For this reason, the compressed samples from the CoMoFoD dataset are taken and analyzed by the custom Mask-RCNN and pictorial results are exhibited in Fig. 10. It is obvious from Fig. 10 that the introduced framework is proficient to identify the manipulated regions from the compressed images. The performance analysis with various approaches is discussed in Table 9, from where it can be seen that the custom Mask-RCNN has acquired an average performance gain of 46.44%, 30.07%, and 48.19% for precision, recall, and F1-measure, respectively.

4.4.6. Blurring attack

The occurrence of blurring in digital samples is another inevitable attack, therefore the CMFD approaches must be robust to it. To analyze the manipulation detection ability of our method under the occurrence of blurring is checked in this experiment. The blurry images are tested by the custom Mask-RCNN and obtained outcomes are shown in Fig. 11, which clearly shows the effectiveness of the proposed solution to blurring attacks. Furthermore, the comparative analysis containing precision, recall, and F1-measure is discussed in Table 10 from where it can be

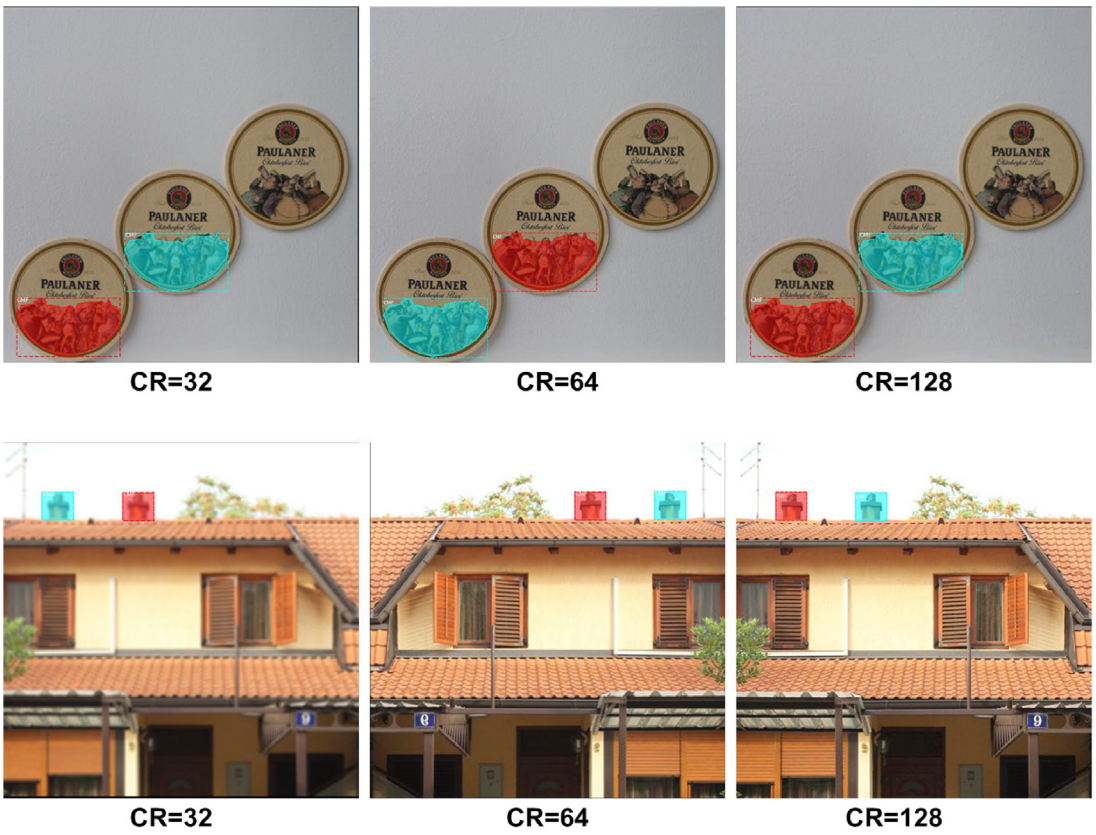

Fig. 8. Visual representation of the CMFD with DenseNet41-Based Mask-RCNN for samples under different levels of Color reduction (CR) post-processing attack.

Table 9
Comparison with competent approaches for samples under the presence of compression post-processing attack.

Reference	Transformation operation (over 40 images)	Precision (average)	Recall (average)	F1-Measure (average)
Li et al. [7]	Translation	0.3835	0.8473	0.4502
	Rotation	0.5809	0.8817	0.6285
	Scaling	0.5630	0.8448	0.6037
	Distortion	0.6490	0.8860	0.6902
Average	All operations	0.5441	0.8649	0.5931
Silva et al. [51]	Translation	0.3789	0.4122	0.3113
	Rotation	0.3990	0.3779	0.2708
	Scaling	0.4858	0.3567	0.3000
	Distortion	0.5625	0.3825	0.3571
Average	All operations	0.4565	0.3823	0.3098
Liu et al. [52]	Translation	0.4052	0.7260	0.4658
	Rotation	0.5977	0.8014	0.6369
	Scaling	0.5169	0.7248	0.5449
	Distortion	0.5963	0.7787	0.6175
Average	All operation	0.5290	0.7577	0.5663
Proposed	Translation	0.9693	0.9691	0.9692
	Rotation	0.9746	0.9691	0.9718
	Scaling	0.9764	0.9691	0.9727
	Distortion	0.9772	0.9691	0.9731
Average	All operation	0.9744	0.9691	0.9717

Table 10
Comparison with competent approaches for samples under the presence of blurring post-processing attack.

Reference	Transformation operation (over 40 images)	Precision (average)	Recall (average)	F1-Measure (average)
Li et al. [7]	Translation	0.3186	0.9206	0.3186
	Rotation	0.4481	0.8753	0.4481
	Scaling	0.4514	0.9096	0.4514
	Distortion	0.5022	0.9449	0.5022
Average	All operations	0.4301	0.9126	0.4301
Silva et al. [51]	Translation	0.4139	0.9008	0.4139
	Rotation	0.5183	0.7043	0.5183
	Scaling	0.5281	0.6994	0.5281
	Distortion	0.6243	0.8292	0.6243
Average	All operations	0.5211	0.7834	0.5211
Liu et al. [52]	Translation	0.3481	0.8270	0.3481
	Rotation	0.5114	0.8591	0.5114
	Scaling	0.4890	0.7836	0.4890
	Distortion	0.5715	0.8949	0.5715
Average	All operation	0.4800	0.8411	0.4800
Proposed	Translation	0.9796	0.9477	0.9634
	Rotation	0.9882	0.9398	0.9634
	Scaling	0.9877	0.9497	0.9683
	Distortion	0.9895	0.9595	0.9743
Average	All operation	0.9862	0.9492	0.9673

witnessed that our work is more effective than the peer methods. More specifically, approaches in [7,51,52] attain the average precision, recall, and F1-score of 0.4771, 0.8457, and 0.4771, whereas our framework acquires 0.9862, 0.9492, and 0.9673, respectively. Therefore, for all the transformation operations under the occurrence of blurring attack, the custom Mask-RCNN shows 50.91%, 10.34%, and 49.02% performance gains for precision, recall, and F1-score, respectively.

4.4.7. Contrast adjustment

We have performed another evaluation to check the robustness of CMFD under the varying image contrast values. The contrast variant samples from the CoMoFoD database are evaluated

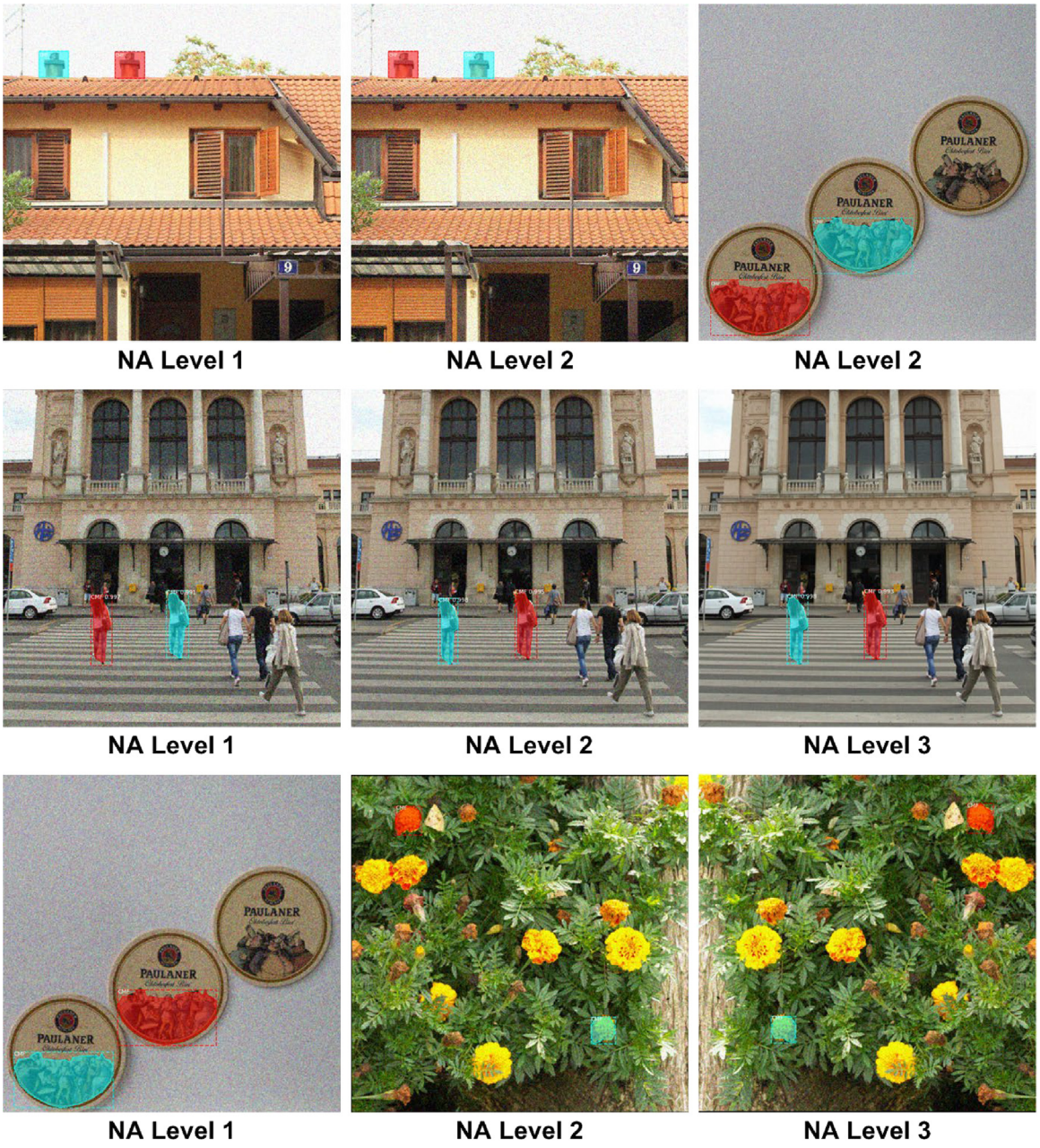

Fig. 9. Visual representation of the CMFD with DenseNet41-Based Mask-RCNN for samples under different levels of Noise addition (NA) post-processing attack.

and obtained results are exhibited in Fig. 12. One can see from the stated results (Fig. 12) that the introduced solution is effective to deal with the contrast variation attacks in the samples and can correctly localize the CMF attacks. Furthermore, performance results with the comparative techniques are discussed in Table 11. More explicitly, for the precision metric, the custom Mask-RCNN gives an average performance gain of 43.90%, while in the case of recall and F1-measure, it shows an average performance gain of 15.15%, and 39.71%, respectively. From the stated values, it is clear that the introduced model is robust to CMFD even under the image contrast variations because of the reliable feature calculation power of the custom Mask-RCNN.

4.4.8. Multi CMF attack

The robustness of the introduced framework is evaluated on a convincing scenario of CMF, where an object is copied many times in a sample to develop forensic alterations. The attained pictorial results are reported in Fig. 13. One can clearly depict from the obtained sample outcomes how well the introduced custom Mask-RCNN model performs under the existence of multi-CMF attacks.

4.5. Evaluation with the latest methods

The experimental analysis of the introduced framework against several post-processing attacks as illustrated in Section 4.3 has confirmed that our approach can effectively locate the manipulated samples from the suspected samples. To further demonstrate the CMFD power of our framework, we have tested it over the MICC-F2000, and CASIA-v2 datasets as well as against several latest approaches. For all the CoMoFoD, MICC-F2000, and CASIA-v2 databases, the comparative results against state-of-the-art approaches are shown in Table 12.

To assess the forgery detection accuracy of Custom Mask-RCNN against the CoMoFoD dataset, we have compared it with

Fig. 10. Visual representation of the CMFD with DenseNet41-Based Mask-RCNN for samples under levels of JPEG Compression (JC) post-processing attack.

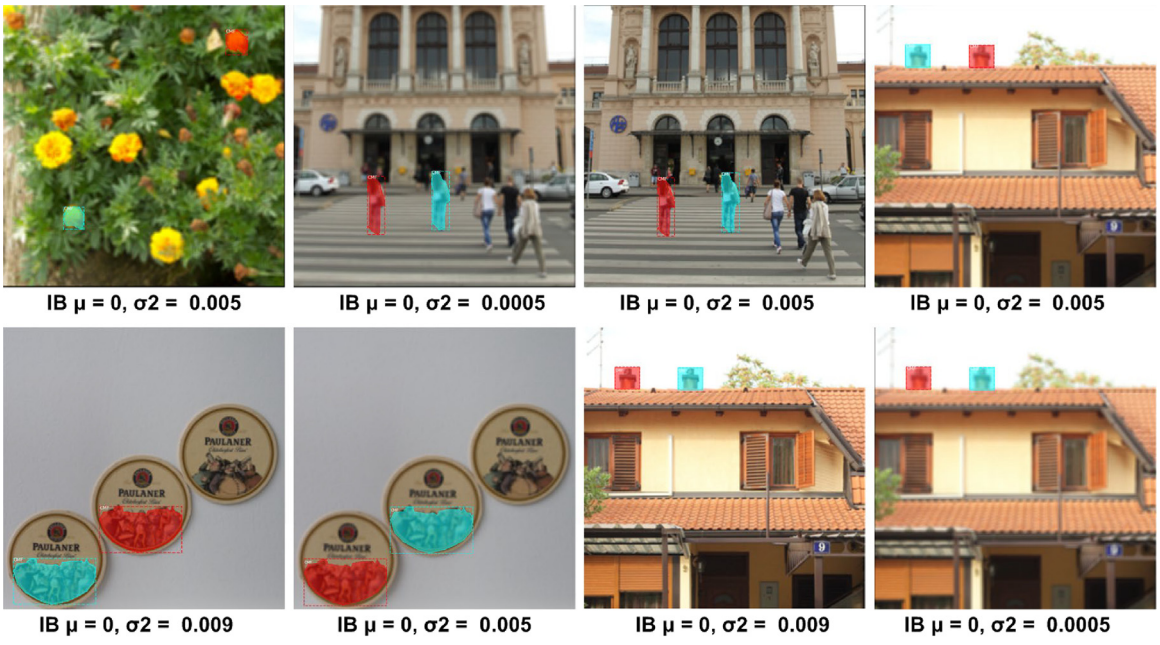

Fig. 11. Visual representation of the CMFD with DenseNet41-Based Mask-RCNN for samples under levels of image blurring (IB) post-processing attack.

the methods mentioned in [53–56]. Whereas, for the MICC-F2000 database, we have nominated the latest studies mentioned in [55,57–59]. The obtained values are clearly showing that our method is more robust and obtained the highest results than the methods in [53–59]. The techniques mentioned in [53,54,56] can accurately locate the manipulated content from the noisy and compressed images, however, these approaches are unable to work well for samples with intense rotational and scale variation. Whereas, the approach in [55] works well for image rotation and scale changes, however, it is unable to locate the forgery for the images with high color variations. Moreover, for the method in [57–59], the CMFD performance lacks images containing resizing and illumination changes. It can be seen from Table 12, that our proposed method shows improved results on the CASIA-V2 database as well in comparison to existing methods. The method presented in [26,61] employs CNN for the identification of CMF,

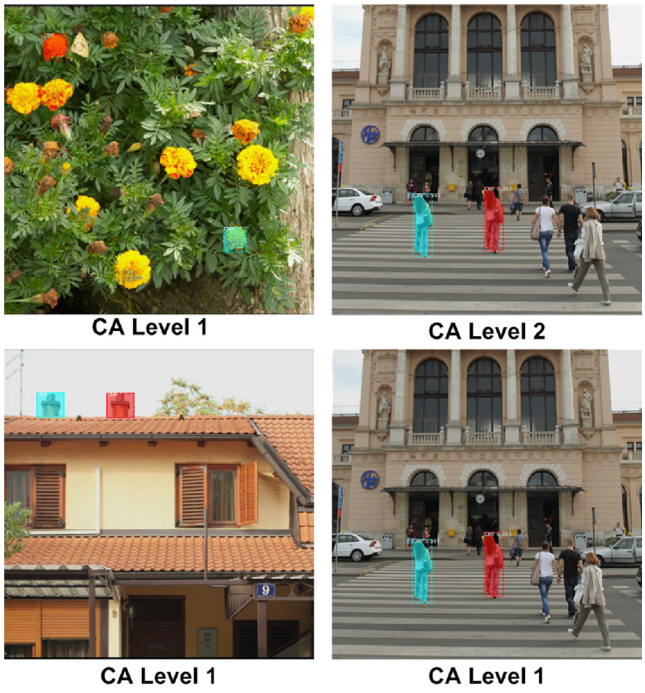

Fig. 12. Visual representation of the CMFD with DenseNet41-Based Mask-RCNN for samples under different levels of Contrast adjustment (CA) post-processing attack.

Table 11
Comparison with competent approaches for samples under the presence of contrast adjustment post-processing attack.

Reference	Transformation operation (over 40 images)	Precision (average)	Recall (average)	F1-Measure (average)
Li et al. [7]	Translation	0.4031	0.8706	0.4746
	Rotation	0.5447	0.8428	0.5972
	Scaling	0.5714	0.8400	0.5973
	Distortion	0.6445	0.8994	0.6941
Average	All operations	0.5409	0.8632	0.5908
Silva et al. [51]	Translation	0.5564	0.8586	0.5995
	Rotation	0.5722	0.6987	0.5321
	Scaling	0.5334	0.7228	0.5059
	Distortion	0.5334	0.7228	0.5059
Average	All operations	0.5488	0.7507	0.5358
Liu et al. [52]	Translation	0.4101	0.7218	0.4670
	Rotation	0.6157	0.8675	0.6590
	Scaling	0.5517	0.7987	0.5948
	Distortion	0.6610	0.8476	0.6924
Average	All operation	0.5596	0.8089	0.6033
Proposed	Translation	0.9879	0.9492	0.9682
	Rotation	0.9898	0.9591	0.9742
	Scaling	0.9992	0.9592	0.9788
	Distortion	0.9785	0.9691	0.9738
Average	All operation	0.9888	0.9591	0.9737

however, suffers from a model overfitting problem. The method in [60] performs boundary to pixel direction segmentation using deep features, however, it is sensitive to noise attacks. In comparison, the proposed Custom Mask-RCNN has better tackled the issues of the aforementioned techniques and can accurately locate the image forgeries with the occurrence of post-processing operations in samples. The main reason for the accurate performance of the DenseNet-41-based Mask-RCNN is because of its robust feature extraction power which has presented the image transformations effectively and improved the model recognition ability under the occurrence of numerous image distortions. In addition, the dense connections of DenseNet-41 extract the more descriptive image keypoints set which provides more details of the image. Consequently, the proposed solution outperforms the comparative approaches by a fair margin.

Table 12
Comparative analysis against the latest approaches.

Methods	Precision (%)	Recall (%)	F1-Measure (%)
CoMoFoD			
[53]	32.35	36.87	32.12
[54]	65.47	73.48	64.07
[55]	92.00	80.50	85.86
[56]	97.90	93.66	95.73
Proposed	**98.12**	**95.85**	**96.97**
MICC-F2000			
[57]	89.36	87.5	88.42
[58]	96.34	89.14	92.60
[55]	97.20	96.10	93.43
[59]	–	98.50	94.30
Proposed	**99.02**	**98.98**	**98.95**
CASIA-V2			
[60]	57.48	51.25	48.06
[26]	70.85	58.85	64.29
[61]	67.83	85.69	75.72
Proposed	**83.41**	**86.02**	**84.69**

4.6. Discussion

The forgeries introduced in the digital samples are imposing a serious threat to the authenticity of the images and restricted their usage in the processing of criminal cases. Therefore, researchers are proposing automated systems for the timely detection of real and manipulated images. However, the extensive changes in the brightness, position, scale, and orientation of various objects in digital images and the incidence of noise, blurring, compression, etc., are increasing the difficulties of the forgery identification approaches. In the presented work, we have proposed a deep learning model for the accurate and reliable detection of image forgeries under the presence of several image post-processing tasks.

We have presented a custom Mask-RCNN approach by proposing the Dense-41 model as the base network for detecting CMF from the input images. We have performed extensive experimentation on three challenging datasets of image forgery named the CoMoFoD, MICC-F2000, and CASIA-v2 and attained robust results with precision scores of 98.12%, 99.02%, and 83.41%, respectively. Moreover, we have accomplished a huge performance analysis of the presented work under the occurrence of several image distortions over the CoMoFoD dataset to explain better forgery recognition power of our DenseNet-41-based Mask-RCNN approach. More specifically, for the brightness changes and color reduction attacks, the proposed approach has attained precision values of 97.69% and 97.87%. While, for the noise addition and compression attacks, we have shown precision values of 98.45% and 97.44%. Whereas, for the blurring, and color adjustment attacks, the model has attained precision scores of 98.62% and 98.88%. The basic reason for the better CMFD performance of the proposed approach under the occurrence of the various post-processing attacks like brightness, color variations, noise, compression, blurring, contrast adjustment, and MCF is the better feature computation ability of the improved Mask-RCNN model which presented the sample information under all attacks in a viable manner. Furthermore, the presented model is translational, scale, and rotational invariant which empowers it to identify the

Fig. 13. Visual representation of the CMFD with DenseNet41-Based Mask-RCNN for samples under multiple CMF attacks.

forensic changes effectively under size, angle, and position variations of the copied region in the suspected samples. Additionally, the semantic mask generation ability of the Mask-RCNN model improves its recall ability which enables it to detect multiple CMF attacks from the digital samples effectively. Moreover, we have trained the model with optimal hyper-parameters which optimize the framework performance and enable it to perform effectively on three challenging datasets CoMoFoD, MICC-F2000, and CASIA-v2. The performance analysis under the incidence of the image post-processing attack is clearly exhibiting the robustness of the DenseNet-41-based Mask-RCNN approach for the CMFD. However, small performance degradation is found for images with huge light alterations as the proposed approach ignores some low-level features computation which lacks to fully capture all information of the suspected samples under such variations. In the future, we plan to overcome such limitations by employing other DL approaches. Furthermore, we plan to test other DL approaches for the detection of CMFD. Moreover, we plan to present a unified model capable of detecting both the CMF and image splicing attacks.

5. Conclusion

In our work, a DL model namely the custom Mask-RCNN has been presented to locate the forensic alteration made in the digital samples. More specifically, the DenseNet-41 network is proposed at the feature calculation level of the Mask-RCNN model for deep key points computation. The inclusion of the DenseNet-41 as the base network assists the proposed model in better identifying a robust set of image features that are later localized, segmented, and classified by the Mask-RCNN approach. Extensive experimentation in contrast to the recent competitive techniques is conducted to show the efficacy of our CMFD approach. The reported results confirm that our approach is capable of accurately identifying the forged content under the existence of noise, blurring, compression, contrast and brightness alterations, and color reduction. Moreover, the approach is competent to detect multiple CMF attacks and proficient in terms of time complexity as well. The major need for forensic analyzers is the correct identification of the forged area in a suspected sample which can assist them to process a legal claim. Hence, due to the accurate recognition and localization of CMF by the proposed DenseNet-41-based Mask-RCNN model, we can say that this approach can play a significant part in the area of digital image analysis. Even though the presented framework is proficient to perform well under the presence of image post-processing attacks, however, a little performance degradation has been observed for images with huge light variations. Therefore, in the future, we plan to overcome this limitation by applying other deep learning methods. Moreover, we plan to investigate the generalization power of our method by designing a cross-dataset evaluation.

CRediT authorship contribution statement

Tahira Nazir: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing. **Marriam Nawaz:** Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Momina Masood:** Methodology, Validation, Formal analysis, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Ali Javed:** Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Standard dataset is used which is publicly available.

Acknowledgment

This work was supported by the Punjab Higher Education Commission of Pakistan via Award No. PHEC/ARA/PIRCA/20527/21.

References

[1] A. Kumar, et al., Markov feature extraction using enhanced threshold method for image splicing forgery detection, in: Smart Innovations in Communication and Computational Sciences, Springer, 2019, pp. 17–27.
[2] S. Agarwal, S. Chand, Image forgery detection using co-occurrence-based texture operator in frequency domain, in: Progress in Intelligent Computing Techniques: Theory, Practice, and Applications, Springer, 2018, pp. 117–122.
[3] M.H. Alkawaz, et al., Detection of copy-move image forgery based on discrete cosine transform, Neural Comput. Appl. 30 (1) (2018) 183–192.
[4] A. Parveen, Z.H. Khan, S.N. Ahmad, Block-based copy–move image forgery detection using DCT, Iran J. Comput. Sci. 2 (2) (2019) 89–99.
[5] I.J. Cox, et al., Digital Watermarking, vol. 53, 2002.
[6] T. Mahmood, et al., A survey on block based copy move image forgery detection techniques, in: 2015 International Conference on Emerging Technologies, ICET, IEEE, 2015.

[7] J. Li, et al., Segmentation-based image copy-move forgery detection scheme, IEEE Trans. Inf. Forensics Secur. 10 (3) (2014) 507–518.
[8] T.K. Huynh, et al., A survey on image forgery detection techniques, in: The 2015 IEEE RIVF International Conference on Computing & Communication Technologies-Research, Innovation, and Vision for Future, RIVF, IEEE, 2015.
[9] T. Nazir, et al., Digital image forensic analysis using hybrid features, in: 2021 International Conference on Artificial Intelligence, ICAI, IEEE, 2021.
[10] M.A. Qureshi, M. Deriche, A bibliography of pixel-based blind image forgery detection techniques, Signal Process., Image Commun. 39 (2015) 46–74.
[11] H. Kasban, S. Nassar, An efficient approach for forgery detection in digital images using Hilbert–Huang transform, Appl. Soft Comput. 97 (2020) 106728.
[12] L. Darmet, K. Wang, F. Cayre, Disentangling copy-moved source and target areas, Appl. Soft Comput. 109 (2021) 107536.
[13] A. Dixit, S. Bag, Adaptive clustering-based approach for forgery detection in images containing similar appearing but authentic objects, Appl. Soft Comput. 113 (2021) 107893.
[14] S. Tinnathi, G. Sudhavani, An efficient copy move forgery detection using adaptive watershed segmentation with AGSO and hybrid feature extraction, J. Vis. Commun. Image Represent. 74 (2021) 102966.
[15] Q. Lyu, et al., Copy move forgery detection based on double matching, J. Vis. Commun. Image Represent. (2021) 103057.
[16] P. Niu, et al., Fast and effective keypoint-based image copy-move forgery detection using complex-valued moment invariants, J. Vis. Commun. Image Represent. (2021) 103068.
[17] R. Agarwal, O.P. Verma, Robust copy-move forgery detection using modified superpixel based FCM clustering with emperor penguin optimization and block feature matching, Evol. Syst. (2021) 1–15.
[18] G. Tahaoglu, et al., Improved copy move forgery detection method via L* a* b* color space and enhanced localization technique, Multimedia Tools Appl. (2021) 1–38.
[19] M. Nawaza, et al., Single and multiple regions duplication detections in digital images with applications in image forensic, J. Intell. Fuzzy Syst..
[20] C. Lin, et al., Copy-move forgery detection using combined features and transitive matching, Multimedia Tools Appl. 78 (21) (2019) 30081–30096.
[21] C.-C. Chen, W.-Y. Lu, C.-H. Chou, Rotational copy-move forgery detection using SIFT and region growing strategies, Multimedia Tools Appl. 78 (13) (2019) 18293–18308.
[22] A. Roy, et al., Copy-move forgery detection with similar but genuine objects, in: Digital Image Forensics, Springer, 2020, pp. 65–77.
[23] K.B. Meena, V. Tyagi, A copy-move image forgery detection technique based on tetrolet transform, J. Inf. Secur. Appl. 52 (2020) 102481.
[24] N. Goel, S. Kaur, R. Bala, Dual branch convolutional neural network for copy move forgery detection, IET Image Process. 15 (3) (2021) 656–665.
[25] X. Wang, et al., Detection and localization of image forgeries using improved mask regional convolutional neural network, Math. Biosci. Eng. 16 (5) (2019) 4581–4593.
[26] J.-L. Zhong, C.-M. Pun, An end-to-end dense-inceptionnet for image copy-move forgery detection, IEEE Trans. Inf. Forensics Secur. 15 (2019) 2134–2146.
[27] Y. Zhu, et al., AR-net: Adaptive attention and residual refinement network for copy-move forgery detection, IEEE Trans. Ind. Inform. 16 (10) (2020) 6714–6723.
[28] R.E. Yancey, N. Matloff, P. Thompson, Multi-linear faster RCNN with ELA for image tampering detection, 2019, arXiv preprint arXiv:08484.
[29] G. Tahaoglu, et al., Ciratefi based copy move forgery detection on digital images, Multimedia Tools Appl. (2022) 1–36.
[30] G. Yue, et al., SMDAF: A novel keypoint based method for copy-move forgery detection, IET Image Process. (2022).
[31] Y. Gan, J. Zhong, C. Vong, A novel copy-move forgery detection algorithm via feature label matching and hierarchical segmentation filtering, Inf. Process. Manage. 59 (1) (2022) 102783.
[32] M.A. Elaskily, et al., A novel deep learning framework for copy-moveforgery detection in images, Multimedia Tools Appl. 79 (27) (2020) 19167–19192.
[33] Y. Rodriguez-Ortega, D.M. Ballesteros, D. Renza, Copy-move forgery detection (CMFD) using deep learning for image and video forensics, J. Imaging 7 (3) (2021) 59.
[34] A. Dutta, A. Gupta, A. Zisserman, Vgg Image Annotator (VIA). [cited Dec, 2020]; Available from: http://www.robots.ox.ac.uk/~vgg/software/via.
[35] M. Nawaz, et al., Skin cancer detection from dermoscopic images using deep learning and fuzzy k-means clustering, Micros. Res. Tech. 85 (1) (2022) 339–351.
[36] T. Nazir, A. Irtaza, V. Starovoitov, Optic disc and optic cup segmentation for glaucoma detection from blur retinal images using improved mask-RCNN, Int. J. Opt. 2021 (2021).
[37] S. Albahli, et al., An improved faster-RCNN model for handwritten character recognition, 46 (9) (2021) 8509–8523.
[38] S. Albahli, et al., An improved faster-RCNN model for handwritten character recognition, Arab. J. Sci. Eng. 46 (9) (2021) 8509–8523.
[39] S. Albahli, et al., Recognition and detection of diabetic retinopathy using densenet-65 based faster-rcnn, Comput. Mater. Contin. 67 (2021) 1333–1351.
[40] M. Masood, et al., A novel deep learning method for recognition and classification of brain tumors from MRI images, Diagnostics 11 (5) (2021) 744.
[41] A.I. Xavier, et al., Object detection via gradient-based mask R-CNN using machine learning algorithms, Machines 10 (5) (2022) 340.
[42] M. Masood, et al., Brain tumor localization and segmentation using mask RCNN, Front. Comput. Sci. 15 (2021) 1–3.
[43] K. Ogurtsova, et al., IDF Diabetes Atlas: Global estimates for the prevalence of diabetes for 2015 and 2040, Diabetes Res. Clin. Pract. 128 (2017) 40–50.
[44] D. Tralic, et al., CoMoFoD—New database for copy-move forgery detection, in: Proceedings ELMAR-2013, IEEE, 2013.
[45] K. Rathi, M. Urvashi, DATASET for Image Forgery Detection.
[46] J. Dong, W. Wang, T. Tan, Casia image tampering detection evaluation database, in: 2013 IEEE China Summit and International Conference on Signal and Information Processing, IEEE, 2013.
[47] D. Theckedath, R. Sedamkar, Detecting affect states using VGG16, ResNet50 and SE-ResNet50 networks, SN Comput. Sci. 1 (2) (2020) 1–7.
[48] M. Bansal, et al., Transfer learning for image classification using VGG19: Caltech-101 image data set, J. Ambient Intell. Humaniz. Comput. (2021) 1–12.
[49] I.Z. Mukti, D. Biswas, Ransfer learning based plant diseases detection using ResNet50, in: 2019 4th International Conference on Electrical Information and Communication Technology, EICT, IEEE, 2019.
[50] S.-L. Lin, Application combining VMD and ResNet101 in intelligent diagnosis of motor faults, Sensors 21 (18) (2021) 6065.
[51] E. Silva, et al., Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes, J. Vis. Commun. Image Represent. 29 (2015) 16–32.
[52] Y. Liu, Q. Guan, X. Zhao, Copy-move forgery detection based on convolutional kernel network, Multimedia Tools Appl. 77 (14) (2018) 18269–18293.
[53] Y. Lai, et al., An improved block-based matching algorithm of copy-move forgery detection, Multimedia Tools Appl. 77 (12) (2018) 15093–15110.
[54] O.M. Al-Qershi, B.E. Khoo, Enhanced block-based copy-move forgery detection using k-means clustering, Multidimens. Syst. Signal Process. 30 (4) (2019) 1671–1695.
[55] M. Nawaz, et al., Image authenticity detection using DWT and circular block-based LTrP features, CMC-Comput. Mater. Continua 69 (2) (2021) 1927–1944.
[56] N. Kumar, T. Meenpal, Salient keypoint-based copy–move image forgery detection, Aust. J. Forensic Sci. (2022) 1–24.
[57] I. Amerini, et al., Copy-move forgery detection and localization by means of robust clustering with J-Linkage, Signal Process., Image Commun. 28 (6) (2013) 659–669.
[58] D.M. Uliyan, M.T. Alshammari, Investigation of image forgery based on multiscale retinex under illumination variations, Forensic Imaging (2020) 200385.
[59] W. Ye, et al., A two-stage detection method of copy-move forgery based on parallel feature fusion, EURASIP J. Wirel. Commun. Netw. 2022 (1) (2022) 1–22.
[60] Q. Li, et al., Image copy-move forgery detection and localization based on super-BPD segmentation and DCNN, Sci. Rep. 12 (1) (2022) 1–12.
[61] Y. Wu, W. Abd-Almageed, P. Natarajan, Image copy-move forgery detection via an end-to-end deep neural network, in: 2018 IEEE Winter Conference on Applications of Computer Vision, WACV, IEEE, 2018.