A Deep Framework for Copy-Move Image Forgery Detection and Classification using Improved Faster Regional Convolutional Neural Network

Anjani Kumar Rai[1*], Subodh Srivastava [2]
[1,2] Department of Electronics and Communication Engineering, National Institute of Technology, Patna -800005, India

Abstract

Copy-move forgery manipulates the originality of digital images. It impacts the information related to judicial reports, military operations, journalism, and intelligence. However, the accuracy of traditional detection methods relies on handcrafted feature extraction and geometrical transformation, which requires a high level of human involvement. Moreover, the loss of fine feature details and false bounding boxes also affects the detection accuracy. Thus, an improved, faster regional convolutional neural network has been proposed to overcome these challenges. To advance the efficacy, a modified backbone is proposed. It employs ResNet101 for fine feature learning. Additionally, a generalized intersection over union as a new bounding box loss is proposed to achieve precise detection of forged instances. The effectiveness of the proposed method has been assessed on MICC-F220, modified CASIA, CoMoFoD, and COVERAGE. The comparative quantitative exhibit shows that the proposed method attains an accuracy gain over Faster-RCNN, RetinaNet, detection transformer, and YOLOv8 by +16.77%, +13.78%, +8.82% and +5.71%. The comparative qualitative performance reveals that the proposed method facilitates the precise detection and classification of multi-dataset copy-move image forgery in a single framework.

Keywords: Copy-Move Forgery, Detection, Faster Regional Convolutional Neural Network, Generalized Intersection over union, ResNet101

1. Introduction

Digital images serve as an important means of conveying messages. It controls the function of our day-to-day activities, such as medical images, social media, documentation, judicial reports, surveillance, remote sensing, the defence sector, and other image-based operations [1]. The artificial intelligence-based tool reduces the human effort. However, this advancement led to image manipulation in terms of image editing, deep fake, morphing, adjustment, and retouching. Furthermore, the human eye finds it difficult to visualize the modifications made to the image attributes [2]. Mobile services have emerged as a convenient source for acquiring images. Besides this, social networking tempers and distributes the acquired digital images. Thus, image manipulation is a multifaceted challenge that affects the authenticity of digital images.

Digital image forgeries are divided into active and passive image forgeries [3]. The active image forgeries require prior knowledge of the actual image to recognize the manipulated region of interest (RoI). To prevent this, watermarking and digital signatures on images are common approaches. In passive image forgeries, the image characteristics are manipulated in terms of copy-move forgery (CMF), image splicing, and counterfeiting [4]. CMF is a pixel-based passive form of image manipulation. In CMF, some fragments of an image are copied and pasted into the other parts of the same image [5]. This kind of forgery poses a serious risk to the authenticity of digital evidence, media reports, and data security. Therefore, the detection and classification of CMF is a major research problem. The classification process categorizes the images as real and fake, while detection implies the area of manipulated RoI within the image.

The copy-move forgery detection and classification (CMFDC) techniques are grouped as traditional, machine learning (ML), and deep learning (DL) based methodologies [6]. The performance of traditional methodologies such as Speeded-Up Robust Features (SURF), Histogram of Oriented Gradients (HOG), Gray Level Co-occurrence Matrix (GLCM), and Discrete Cosine Transform (DCT) relies on image attributes. The ML-based methods' efficacy depends on segmentation of forged instances along with proper extraction and selection of features. However, these methods require high human involvement and rely on handcrafted features. Moreover, they are not adaptive to fine feature variations [7]. Besides this, DL-based methods for CMFDC do not necessitate extraction of features and selection. However, the effectiveness of its detection is contingent upon the precise spatial, local, and global feature representation. Moreover, the efficacy of ML- and DL-based methodologies depends upon the datasets utilized. The

training on limited CMF datasets results in model overfitting. Thus, to address these challenges, an improved faster regional convolutional neural network (IFRCNNet) has been proposed. The primary objective of the proposed method is to achieve precise detection and classification of the CMF in a single framework. The proposed deep framework is composed of an advanced backbone module for fine feature representation, a new bounding loss function for precise detection of forged instances. The key contributions of the proposed IFRCNNet are as follows:

- A novel IFRCNNet deep framework is proposed for the CMF detection and classification for digital forged images.
- A modified backbone module integrating ResNet101 is proposed for fine feature learning.
- A generalized intersection over union based new bounding box loss function is introduced for precise detection of forged RoIs from the digital images.
- An extensive comparative visual and statistical analysis has been presented for the validation of the proposed IFRCNNet on openly accessible datasets.

The paper is organized as follows: Section 2 reviews the existing literature, Section 3 describes the proposed methodology, Section 4 presents the research findings and comparative analysis, and Section 5 concludes the study

2. Related Work

The literature has presented numerous studies on CMFDC that utilize ML and DL principles. Kiruthiga et al. [8] presented a faster regional convolutional neural network (Faster RCNN) for CMFDC. Furthermore, a Bayer filter was utilized to improve spatial features of Faster RCNN. The method was trained for the COVERAGE CMF dataset, and the accuracy achieved was 94%. However, the method was developed for smaller and single CMF datasets. Moreover, the accuracy was also moderate. Besides this, no qualitative analysis was illustrated for CMFDC. Basavaraj et al. [9] employed YOLOv3, ZF-Net, and Hybrid Teacher-Learning-Based Optimization (HTLBO) to perform CMFDC. The HTLBO was applied to tune the parameters of ZF-Net. The method was iterated for multiple image datasets (MISD), and the obtained accuracy was 95.8%. However, the model merely combined various techniques without introducing any new approaches. Furthermore, the baseline models were used without any modification. Tamilselvi [10] adapted Faster RCNN with the Naive Bayes classifier and SURF for image tampering analysis. It was primarily designed for CASIA v1.0 CMF datasets. The attained accuracy was 92%. However, no comparative visual and statistical data was presented to judge the efficacy of the method. Moreover, the integration of DL, ML, and traditional methods was unclear. Huynh et al. [11] utilized YOLOv5 in combination with K-means clustering and a modified Zernike moment for the purpose of CMFDC. The model first performed object detection followed by the copy-move forgery classification. The method was trained for MICC-F600 CMF datasets, and the achieved accuracy was 95.37%. However, no comparative qualitative analysis was presented for the validation of the work. Moreover, the baseline model was used as it was. Li et al. [12] adapted ConvNeXt-founded for document forgery detection. The method was trained for a self-prepared dataset known as a document image dataset. The achieved F1 score was 0.939. However, the model was designed for a smaller dataset. Moreover, the method performs only classification. Atak et al. [13] used vision transformer with variational autoencoder network for CMF classification. The method was developed for the CASIAv2 and COVERAGE CMF datasets. The obtained area under score was 91.4%. However, the efficacy of the method was a function of patch selection. Sabeena et al. [14] had used the convolutional block attention module (CBAM) for CMF detection. Moreover, feature map self-correlation with Atrous spatial pyramid pooling (ASPP) was applied to generate the mask of the forged instances. The method was iterated for copy-move forgery detection (CoMoFoD) in the CMF dataset. The attained precision was 98.5%. However, the method was developed for smaller datasets. Latnekar [15] adapted Faster-RCNN with logit normalization and distance intersection over union-centered for CMFDC. The method was trained for MICC and CASIA CMF datasets. The achieved accuracy was 96.72%. However, no comparative qualitative analysis was demonstrated for the validation of the work. Moreover, the baseline model was used without any refinement. Qiu et al. [16] had integrated Faster-RCNN with the Hilbert–Huang spectrogram and the similarity computation module to perform CMF detection from the audio dataset. The method was developed for TIMIT, the Arabic Speech Corpus, and the Chinese speech CMF dataset. The average accuracy was 83.00%. However, the Faster R-CNN was used without any modification. Su et al. [4] had integrated ConvNeXt, CBAM, ASPP, and an edge supervision network for copy move forgery detection. The COVERAGE, CASIA, and

NIST16 datasets were utilized to evaluate the efficacy of the technique. The achieved accuracy was 98.1%. However, no image forgery classification was accomplished by this model; only a mask was obtained. Wei et al**.** [17] employed Faster-RCNN with an edge detector for CMFCD. Moreover, ResNet101 was used as a backbone module. The method was designed for NIST2016, CASIA, and Columbia CMF. The mean average precision achieved was 81.00%. However, no comparative qualitative analysis was presented to judge the effectiveness of the method.

Literature reveals that existing CMFDC methods have some limitations. It is observed that most of the existing methods were developed for single CMF datasets. Furthermore, the existing methods primarily employed either Faster-RCNN or a variant of YOLO for CMFDC. Besides these, most of the existing methods used a baseline model without any modification. Traditional methods such as SURF had limited effectiveness when it came to fine feature details. Additionally, ML-based performance depends on the handcrafted features. Apart from this, they require a high level of human intervention. The existing Faster-RCNN-based methods have limited performance in terms of precise detection of forged RoIs. Additionally, the existing methods lack the integration of the attention module for fine feature representation learning. This architectural refinement makes them less effective in identifying subtle and finely forged regions. Overall, the major limitations of the existing methods are listed as dependence on the baseline model without any architectural advancement; lack of an attention module for fine feature learning; usage of a smaller CMF dataset for the development of the model; and failure to offer precise detection of forged instances. Thus, to overcome these issues, a novel IFRCNNet has been proposed for CMFDC in a simultaneous framework.

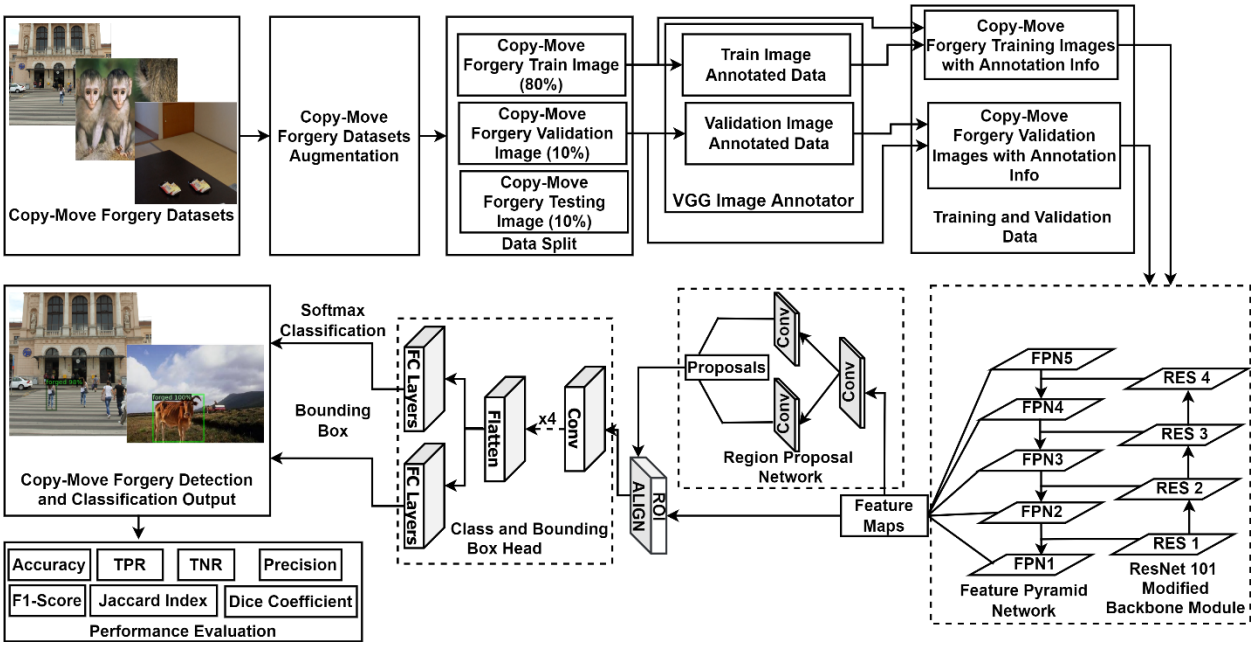

Fig.1: Schematic outline of the proposed improved faster regional convolutional neural network for copy-move forgery detection and classification

3. Method and Models

Fig. 1 illustrates the schematic outline of the proposed IFRCNNet. The design of the proposed framework is initiated with the acquisition of openly accessible copy-move forgery datasets. Afterwards, the acquired data has been augmented to balance the number of images. Later, the augmented data is divided into training, validation, and test data in a ratio of 80%:10%:10. Furthermore, the training and validation images are annotated with the help of the visual geometric image annotator (VIA) [18]. To develop the proposed IFRCNNet, two major advancements have been done at the architectural level. The first advancement involves modifying the backbone by using ResNet101 to enhance fine feature learning. The second advancement is a generalized intersection over union (GIoU) employed as a new bounding box loss to achieve precise detection of forged instances.

3.1 Dataset

The openly accessible CMF datasets such as copy-move forgery detection (CoMoFoD) [19], MICC-F220 [20], Modified CASIA [21], and the copy-move forgery database with similar but genuine objects (COVERAGE) [22] are used for the development of the proposed IFRCNNet. The CoMoFoD has 260 images with variable image dimensions. The MICC-F220 has total 220 images divided equally as real and forged with a varying image size. The modified CASIA dataset has 800 images with an image dimension of 384×256. The COVERAGE dataset consists of total 200 images. Table 1 exhibits the complete description of CMF datasets.

Table 1: Description of copy-move forgery datasets

Dataset	Images	Format	Dimensions	URL
CoMoFoD [19]	260	JPEG	512×512, 3000×2000	https://www.vcl.fer.hr/comofod/#:~:text=CoMoFoD%20database%20for%20a%20copy,%2C%20scaling%2C%20combination%20and%20distortion.
MICC-F220 [20]	220	JPEG	722×480	https://lci.micc.unifi.it/labd/2015/01/copy-move-forgery-detection-and-localization/
Modified CASIA [21]	800	JPEG	384×256	https://ieee-dataport.org/open-access/modified-casia
COVERAGE [22]	200	JPEG	400×486	https://github.com/wenbihan/coverage

3.2 Proposed Improved Faster Regional Convolutional Neural Network for Copy-Move Forgery Detection and Classification

Copy-move forgery is a prevalent form of image manipulation that has high impacts on day-to-day activities and on society. The proposed IFRCNNet's progression commences with data acquisition, augmentation, partitioning, and annotation. These operations are commonly designated as data preparation steps as seen in Fig. 1. Let D demonstrates the CMF data obtained after these operations. Therefore, it can be mathematically expressed using Eq. (1) as follows:

$$D = D_{Train} \cup D_{Val} \cup D_{Test} \quad (1)$$

where, $D_{Train} = (X_i, Y_i)$, $D_{Val} = (X_j, Y_j)$, and D_{Test} are the train, validation, and test CMF images. X_i, and X_j are the training, and validation images along with their annotated labels Y_i, and Y_j.

The architecture of the proposed IFRCNNet includes a backbone, feature pyramid network (FPN), region proposal network (RPN), RoI align, classification head, and bounding box head, as illustrated in Fig.1. It is an advanced form of the conventional Faster-RCNN [23]. VGG16 is employed for the initial feature extractor and serves the role of the backbone module in conventional Faster-RCNN. However, it has certain architectural limitations that impact its effectiveness as a backbone module. Its architecture is comprised of sequential 3×3 convolutional layers, which lack residual connections. Moreover, it results in vanishing gradients and restricts the fine features representation. To overcome these limitations, the proposed framework adopts ResNet101 [24] as the backbone. It is used due to its residual learning properties and deep hierarchical blocks to preserve gradient flow. Therefore, the feature maps obtained from ResNet101 may be expressed as:

$$C_l = f_{ResNet101}(x_i), l \in \{2,3,4,5\} \quad (2)$$

where, x_i is the input image and C_l signifies the hierarchical convolutional features obtained through various ResNet101 layers.

The various layers of ResNet101 includes residual learning, succeeded by a convolutional ($Conv$), batch normalization, and $ReLU$ activation. Let F_k represents the feature maps at level k of the FPN. It reads as:

$$F_k \in \mathbb{R}^{C \times H_k \times W_k}, k \in \{2,3,4,5\} \quad (3)$$

where, P_k exhibits the acquired features at different scales. C, H_k and W_k represents the channel, height, and width of the feature map. $\mathbb{R}$ shows the real valued features. The modified backbone module with FPN [25] shows a top-down feature hierarchy for feature refinement. The pyramid levels are computed as:

$$F_{k-FPN} = Conv_{1\times1}(U_P(F_{k+1}) + C_l); l \in \{2,3,4,5\} \quad (4)$$

where, $U_P(.)$ is the up sampling, C_l is the feature maps at a various residual stage and F_{k-FPN} is the hierarchical features.

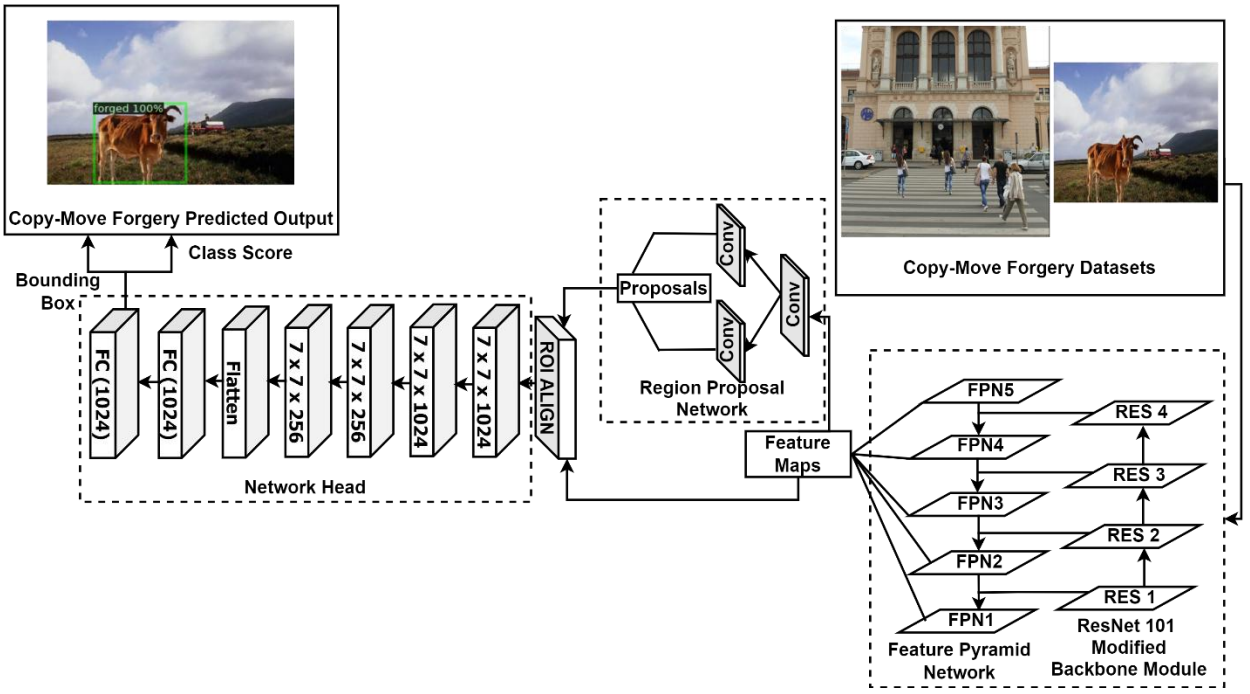

Fig.2: Architectural details of the Proposed improved faster regional convolutional neural network

Furthermore, these feature maps fed to RPN where it develops region proposals or anchors using feature maps. It highlights the probable forged instances within the image. It is mathematically articulated as:

$$P_k = (x_c, y_c, w, h) \tag{5}$$

where, P_k is probable anchor or region proposal, (x_c, y_c) are the center coordinates, and (w, h) are width and height. Afterwards, for the obtained proposals is processed through the RoI-align which produces a fixed size RoI features. It may be mathematically expressed as:

$$F_{RoI}^{(j)} = RoIalign(F_{k-FPN} * P_k) \tag{6}$$

In subsequent step, the obtained feature maps are fed to the network head which comprised of class and box heads. The class head offers a confidence score which reflect the likeliness or probability that a predicted region actually belongs to either forged or real class. The bounding box (BB) signifies the region of the forged instances. The network head includes four successive $Conv$ layers, followed by a flatten layer, and two fully connected (FC) layers. In the hidden layers, $ReLU$ and max pooling occur after each of the four $Conv$ layers. The loss function of the Faster-RCNN [23] clubs the BB, and class loss. The class loss is assessed with respect to binary cross entropy loss while $L1_{smooth}$ is utilized for BB loss by the conventional Faster-RCNN. The complete loss function equation is expressed as:

$$L_{Faster-RCNN} = L_{CLS} + L_{BOX} \tag{7}$$

Here, the L_{CLS} is evaluated as:

$$L_{CLS} = \frac{1}{N_{cls}} \sum_i -m_i^* log m_i^* - (1 - m_i^*)\, log(1 - m_i^*) \tag{8}$$

The BB loss function L_{BOX} is articulated as:

$$L_{BOX} = \frac{\lambda}{N_{box}} \sum_i L1_{smooth}\,(z_i - z_i^*) \tag{9}$$

From the Eqs. (7), and (8), m_i, and z_i are the predicted probability, and bounding box of the target; m_i^*, and z_i^*, are the ground-truth label, and bounding box of the target. However, $L1_{smooth}$ have limited performance corresponding to precise CMF detection. Moreover, it is less sensitive to the outliers. Thus, to refine the efficacy of BB detection, GIoU) [26] is utilized aa a new BB loss function for the proposed IFRCNNet. It is statistically expressed as:

$$GIOU = \left| \frac{z_j^{(i)} \cap \hat{z}_j^{(i)}}{z_j^{(i)} \cup \hat{z}_j^{(i)}} \right| - \frac{\left| t \setminus \left(z_j^{(i)} \cup \hat{z}_j^{(i)} \right) \right|}{|t|} \tag{10}$$

where, t is the minimal closing box which entirely encircles both $z_j^{(i)}$ and $\hat{z}_j^{(i)}$. $\left|t \setminus \left(z_j^{(i)} \cup \hat{z}_j^{(i)}\right)\right|$ is the outermost region to the union of $z_j^{(i)}$ and $\hat{z}_j^{(i)}$ but lies within t. Therefore, the new BB loss function is expressed as:

$$L_{BOX_{new}} = \sum_{i} \sum_{j \in \{x,y,w,h\}} GIOU\left(z_j^{(i)}, \hat{z}_j^{(i)}\right) \tag{11}$$

Thus, the modified loss function combining class and BB loss of IFRCNNet are demonstrated as:

$$L_{\text{IFRCNNet}} = L_{CLS} + L_{BOX_{new}} \tag{12}$$

4. Result and Discussion

The proposed IFRCNNet has been trained on openly accessible CMF datasets, as mentioned in Section 3.1. It has been observed that the available CMF datasets have an uneven number of real and forged images. Moreover, they exhibit less in quantity. Both these issues restrict the performance of the proposed method and may lead to model overfitting. Therefore, to address this issue, the principle of augmentation [27] has been utilized. The primarily used augmentation operations are as follows: rotation, horizontal flip (HF), vertical flip (VF), HF followed by VF, VF followed by HF, rotation with HF, and rotation with VF. Besides this, such images that have lost the information due to the augmentation are removed to balance the datasets. Table 2 illustrates a complete overview of CMF datasets in terms of before augmentation (BF AUG), after augmentation (AF AUG), removed, and total number of images.

Table 2: Overview of the CMF datasets BF and AF augmentation

Dataset	BF AUG	AF AUG	Removed	Total
CoMoFoD [19]	260	1040	40	1000
MICC-F220 [20]	220	1100	100	1000
Modified CASIA [21]	800	1600	600	1000
COVERAGE [22]	200	1200	200	1000
Total				4000

4.1 Experimental Setup and Performance Metrics

The proposed IFRCNNet is trained for the Python programming language. The system environment includes an Intel(R) Xeon(R) processor with 128 GB of RAM and an NVIDIA RTX A2000 series graphics card with 12 GB. To train the proposed IFRCNNet, the batch size, epochs, and learning rate have been set at 16, 1000, and 0.0001, respectively. Moreover, the Adam optimizer is applied to update the weights of hyperparameters.

The effectiveness of the proposed IFRCNNet has been assessed in terms of the classification and detection quantitative metrics [28], [29]. Table 3 summarizes such metrics. These metrics are evaluated with respect to confusion matrices, which are composed of quantitative information about actual and predicted classes. Moreover, literature shows that, for effective performance, the values of these metrics should be high.

Table 3: Performance Evaluation Metrics

Metrics	Mathematical Formula
Accuracy (ACC)	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$
True positive rate (TPR)	$TPR = \frac{TP}{TP + FN}$
True negative rate (TNR)	$TNR = \frac{TN}{TN + FP}$
Precision (PRC)	$PRC = \frac{TP}{TP + FP}$
F1-Score (F1-SC)	$F1 - SC = \frac{2 * PC * TPR}{PC + TPR}$
Dice Coefficient (DCF)	$DCF = \frac{2\|Y \cap Y^*\|}{\|Y\| + \|Y^*\|'}$
Jaccard Index (JAC)	$JAC = \frac{\|Y \cap Y^*\|}{\|Y \cup Y^*\|'}$
	where, Y and Y^* are the predicted and actual BB coordinates.

4.2 Qualitative Analysis

To accomplish the comparative qualitative analysis, the existing methods such as Faster-RCNN [23], RetinaNet [30], detection transformer (DETR) [31], and YOLOv8 [32] have been considered. The rationale behind opting for these methods is that they all perform detection and classification. Moreover, these methods are trained and tested for the same datasets under the same experimental setup. Figures 3, 4, 5, and 6 illustrate the comparative qualitative analysis for the CoMoFOD [19], MICC-F220 [20], modified CASIA [21], and COVERAGE [22] datasets. The first image represents the test image. The second are reference annotated images, which exhibit the presence of CMF instances. The third to seventh images of these figures depict the results derived from Faster-RCNN, RetinaNet, DETR, YOLOv8, and the proposed IFRCNNet, respectively.

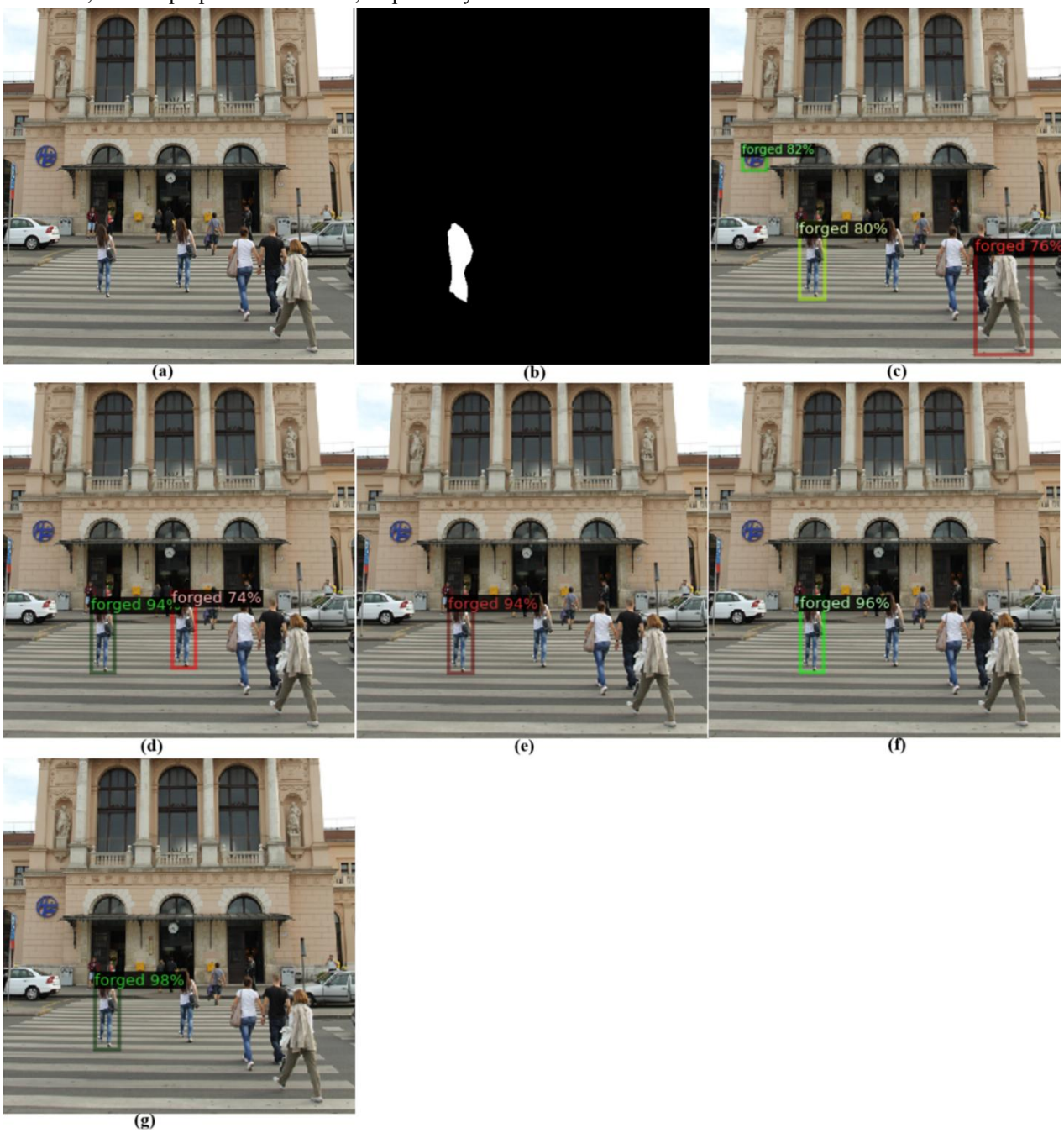

Fig. 3: Comparative qualitative study for CoMoFoD data where (a) is the test image; (b) is the reference image; (c), (d), (e), (f) and (g) are the result of Faster-RCNN, RetinaNet, DETR, YOLOv8, and the proposed IFRCNNet

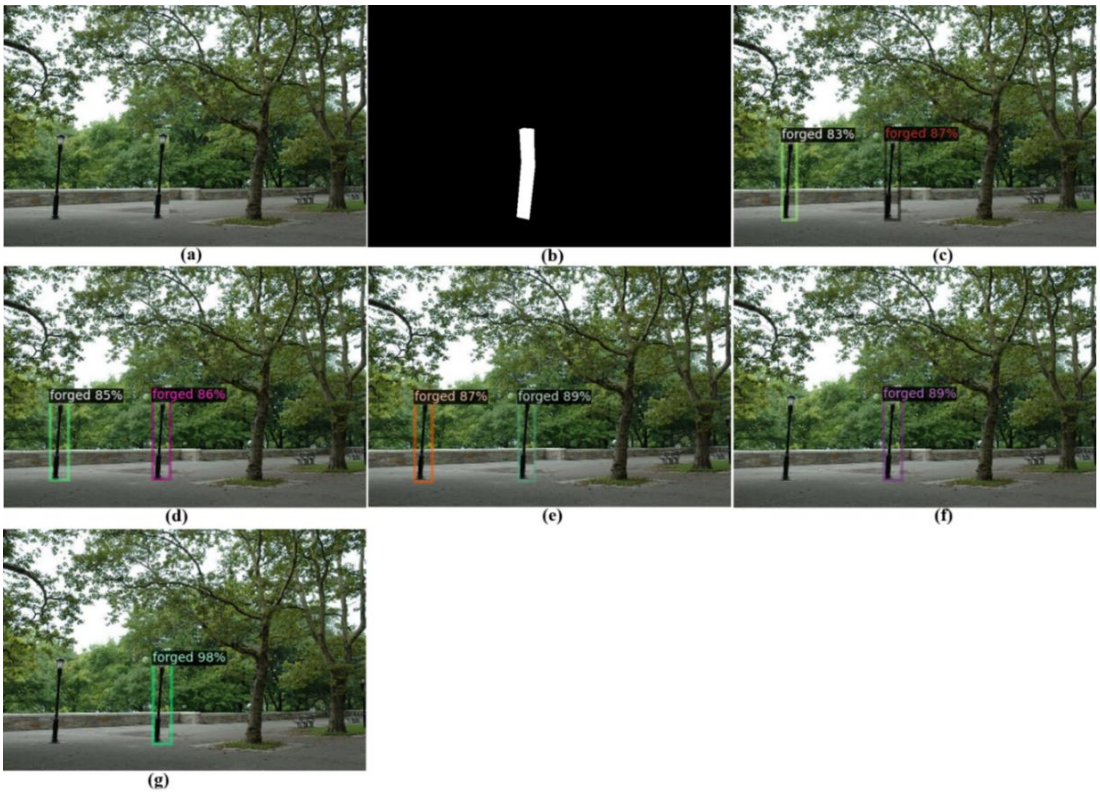

Fig. 4: Comparative qualitative analysis for MICC-F220 data where (a) is the test image; (b) is the reference image; (c), (d), (e), (f) and (g) are the result of Faster-RCNN, RetinaNet, DETR, YOLOv8, and the proposed IFRCNNet

Fig.5: Comparative qualitative study for modified CASIA data where (a) is the test image; (b) is the reference image; (c), (d), (e), (f) and (g) are the outcome of Faster-RCNN, RetinaNet, DETR, YOLOv8, and the proposed IFRCNNet

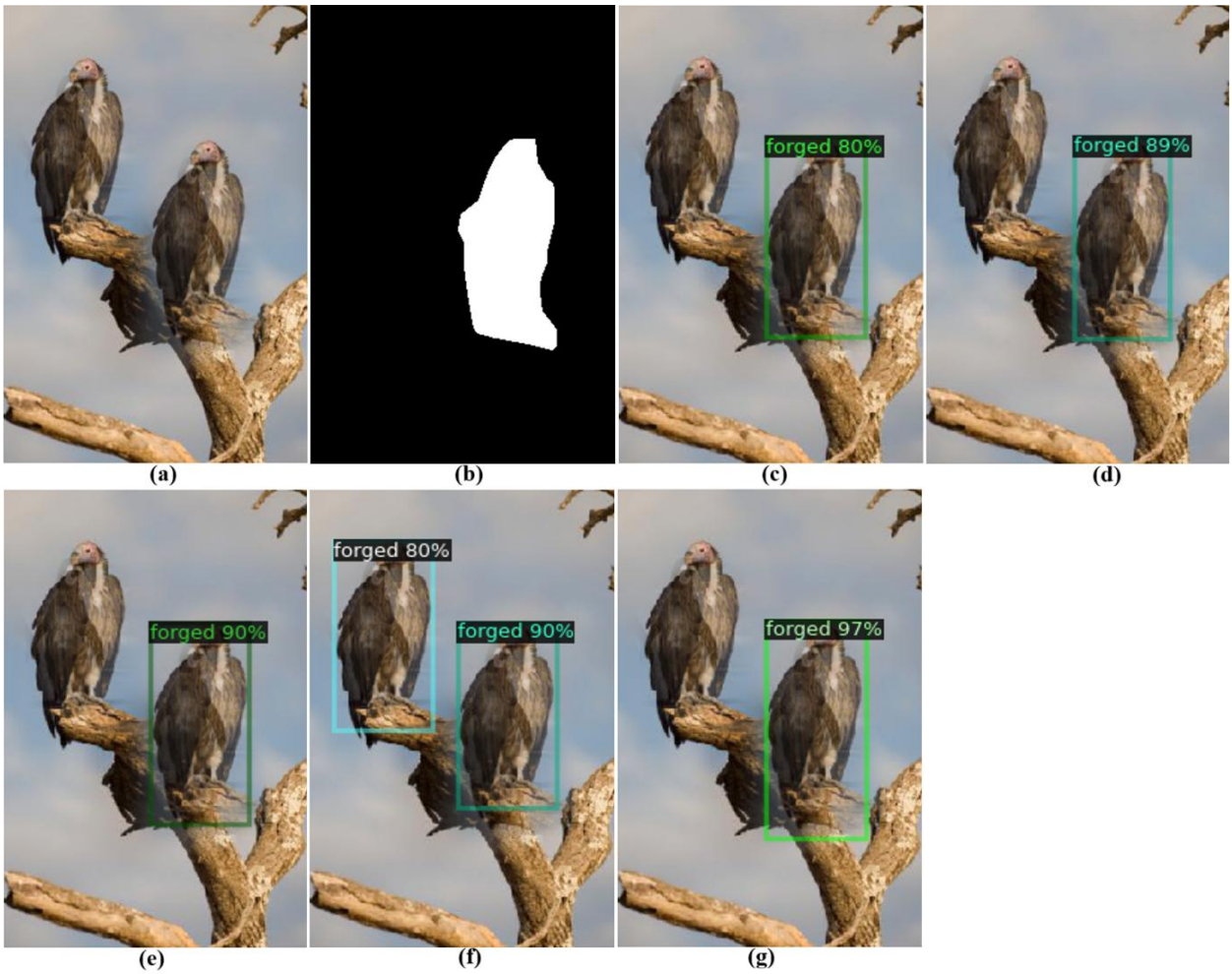

Fig.6: Comparative qualitative evaluation for COVERAGE dataset where (a) is the test image; (b) is the reference image; (c), (d), (e), (f) and (g) are the result of Faster-RCNN, RetinaNet, DETR, YOLOv8, and the proposed IFRCNNet

Fig. 3 presents a comparative qualitative evaluation of the existing object-detection method and the proposed method on the CoMoFoD dataset. Fig. 3(a) displays the original test image, while Fig. 3(b) highlights the forged region with the corresponding ground truth mask. Fig. 3(c) demonstrates the output obtained from Faster R-CNN. It is observed that it has high false detections with varying confidence scores in the range of 76%–82%. Fig. 3(d) shows the result of RetinaNet. It achieves a better bounding box compared to Faster-RCNN, with confidence values varying between 74% and 94%. Figs. 3(e) and 3(f) demonstrate the outcome of DETR and YOLOv8, respectively. They have single-forged instance detection and classification, with confidence scores of 94% and 96%, respectively. In contrast, the result of the proposed IFRCNNet, shown in Fig. 3(g), achieves the most precise detection with a higher classification confidence score of 98%. Fig. 4 depicts a comparative qualitative analysis on the MICC-F220 dataset. Fig. 4(a) presents the test image, and Fig. 4(b) shows the corresponding ground-truth forged region. Figs. 4(c), 4(d), and 4(e) illustrate the outcome of Faster R-CNN, RetinaNet, and DETR. It is observed that these methods have false detection and classification of forged instances. Moreover, the result of YOLOv8 is demonstrated in Fig. 4(f). It has a moderate classification confidence score. In view of this, the result of the proposed IFRCNNet outperforms all the methods, as shown in Fig. 4(g). Fig. 5 presents a qualitative comparison for the modified CASIA dataset. Fig. 5(a) illustrates the test image, and Fig. 5(b) shows the reference ground truth for the forged instance. Figs. 5(c) to 5(f) show the output of Faster R-CNN, RetinaNet, DETR, and YOLOv8. All these methods detected the forged instances accurately, but they have limited performance in terms of classification confidence score. They achieved the confidence scores of

74%, 76%, 88%, and 93%, respectively. Besides this, the result obtained by the proposed IFRCNNet, as depicted in Fig. 5(g), achieves a high classification score of 100%. Fig. 6 illustrates a qualitative comparison on the COVERAGE dataset. Fig. 6(a) shows the test image comprises the CMF instance, and Fig. 6(b) is the ground-truth forged region. Figs. 6(c), 6(d), and 6(e) demonstrate the result of Faster R-CNN, RetinaNet, and DETR. It is observed that they achieved accurate detection of forged instances. However, the classification confidence score is limited to the 80%, 89%, and 90% confidence scores. Moreover, the result of YOLOv8 as exhibited in Fig. 6(f) has false detection of forged instances with the classification confidence score of 80% and 90%. Apart from this, the proposed IFRCNNet yields the most accurate detection with a high classification confidence of 97%, as shown in Fig. 6(g).

4.3 Quantitative Analysis

The quantitative analysis has been evaluated with correspond to the metrices presented in Table 3. The quantitative analysis is accessed for the 10% CMF test images. The comparative quantitative study has been presented between the existing object detection methods and the proposed IFRCNNet. The existing methods such as Faster-RCNN, RetinaNet, DETR, and YOLOv8 are primarily utilized for this purpose. The analysis has been demonstrated with respect to mean ± standard deviation (SD), the 95% confidence interval (CI), and accuracy percentage improvement valuation along with paired t-tests and p-value. The results are demonstrated in Tables 4, 5, and 6, respectively.

Table 4: Comparative quantitative analysis amid existing object detection method and proposed IFRCNNet in terms of mean ± standard deviation

Model	**Metrics**						
	ACC	TPR	TNR	PRC	F1-SC	JAC	DCF
Faster-RCNN [23]	84.40 ± 0.30	84.70 ± 0.23	82.80 ± 0.41	82.98 ± 0.91	82.38 ± 0.42	79.62 ± 0.13	84.46 ± 0.12
RetinaNet [30]	86.62 ± 0.17	85.60 ± 0.42	89.20 ± 0.31	89.42 ± 0.71	85.67 ± 0.12	82.18 ± 0.27	86.52 ± 0.21
DETR [31]	90.58 ± 0.34	90.26 ± 0.31	89.42 ± 0.16	89.47 ± 0.29	89.68 ± 0.34	86.18 ± 0.51	89.46 ± 0.36
YOLOv8 [32]	93.24 ± 0.32	92.56 ± 0.23	92.38 ± 0.27	92.16 ± 0.41	92.36 ± 0.21	86.28 ± 0.21	92.36 ± 0.23
Proposed IFRCNNet	98.56 ± 0.14	98.70 ± 0.08	98.90 ± 0.05	98.92 ± 0.06	98.80 ± 0.07	98.84 ± 0.20	98.90 ± 0.05

Table 5: Comparative quantitative analysis between existing object detection method and the proposed IFRCNNet using 95 % CI

Model	**Metrics**						
	ACC	TPR	TNR	PRC	F1-SC	JAC	DCF
Faster-RCNN [23]	84.40 ± 0.70	84.70 ± 0.48	82.80 ± 0.42	82.98 ± 0.72	82.38 ± 0.58	79.62 ± 0.12	84.46 ± 0.29
RetinaNet [30]	86.62 ± 0.38	85.60 ± 0.43	89.20 ± 0.81	89.42 ± 0.47	85.67 ± 0.29	86.18 ± 0.14	86.56 ± 0.31
DETR [31]	90.58 ± 0.31	90.26 ± 0.38	89.42 ± 0.51	89.47 ± 0.36	89.68 ± 0.12	86.18 ± 0.31	89.46 ± 0.24
YOLOv8 [32]	93.24 ± 0.30	92.56 ± 0.41	92.38 ± 0.21	92.16 ± 0.52	92.36 ± 0.41	86.28 ± 0.42	92.36 ± 0.15
Proposed IFRCNNet	98.56 ± 0.23	98.70 ± 0.04	98.90 ± 0.08	98.92 ± 0.07	98.80 ± 0.08	98.84 ± 0.09	98.90 ± 0.06

Table 6: Comparative improvement performance analysis of the proposed IFRCNNet over baseline methods

Baseline Model	**Attributes**		
	Δ ACC (%)	t-value	p-value
Faster-RCNN [23]	+ 16.77	16.26	1.63×10^{-24}
RetinaNet [30]	+ 13.78	27.48	1.47×10^{-20}
DETR [31]	+ 8.82	13.19	1.34×10^{-22}
YOLOv8 [32]	+ 5.71	18.41	1.38×10^{-22}

Table 4 presents a comprehensive quantitative comparison between the proposed IFRCNNet and Faster R-CNN, RetinaNet, DETR, and YOLOv8. The analysis uses evaluation metrics such as ACC, TPR, TNR, PRC, F1-SC, JAC, and DCF. The results reveal YOLOv8 and DETR achieving an ACC of 93.24 ± 0.32% and 90.58 ± 0.34%, respectively. Moreover, RetinaNet and Faster R-CNN achieve an ACC of 84.40 ± 30% and 86.62 ± 0.17%. In contrast, the proposed IFRCNNet achieves an ACC of 98.56 ± 0.14%, TPR of 98.70 ± 0.08, TNR of 98.90 ± 0.05%, PRC of 98.92 ± 0.06, F1-SC of 98.80 ± 0.07, JAC of 98.84 ± 0.20%, and DCF of 98.90 ± 0.05%. Moreover, it is observed that the proposed IFRCNNet attains a smaller SD (±0.14). It demonstrates that it maintains near-uniform performance across all evaluation conditions. Table 5 presents a comparative analysis of existing object detection models and the proposed

IFRCNNet using 95% confidence intervals (CI). It is observed that Faster R-CNN shows a CI of ±0.70 for ACC and ±0.72 for PRC. It implies a high degree of uncertainty, and its performance can differ across different subsets of forged images. Moreover, RetinaNet and DETR exhibit moderate CIs between ±0.38 and ±0.81. It highlights fluctuations in sensitivity (TPR) and specificity (TNR). Besides this, YOLOv8 shows CIs at ±0.52 for PRC and ±0.41 for TPR. It signifies its mean performance is satisfactory. However, the stability of its results is not consistently reliable. In contrast, the proposed IFRCNNet demonstrates substantially narrower 95% CIs across all metrics: ACC: ±0.23, TPR: ±0.04, TNR: ±0.08, PRC: ±0.07, JAC: ±0.09, and DCF: ±0.06. The narrower CI confirms the greater confidence that the estimated performance metrics correctly represent the model's true behavior on unseen data. Table 6 presents the statistical significance evaluation of the proposed IFRCNNet against four baseline object detection frameworks using percentage improvement in accuracy (ΔACC), paired t-test values, and p-values. The proposed IFRCNNet outperforms Faster R-CNN, RetinaNet, DETR, and YOLOv8 with ACC values of +16.77%, +13.78%, +8.82%, and +5.71%. The t-values ranging from 13.19 to 27.48 are substantially higher than the typical critical threshold for statistical significance at 95% confidence ($|t| > 2$). Furthermore, the corresponding p-values 1.63×10^{-24}, 1.47×10^{-20}, 1.34×10^{-22}, and 1.38×10^{-22} are extremely small. It is far below the standard significance level of 0.05. These p-values provide overwhelming statistical evidence that the performance improvements are highly significant. Moreover, it effectively eliminates the hypothesis that the proposed IFRCNNet and the baseline models perform equally.

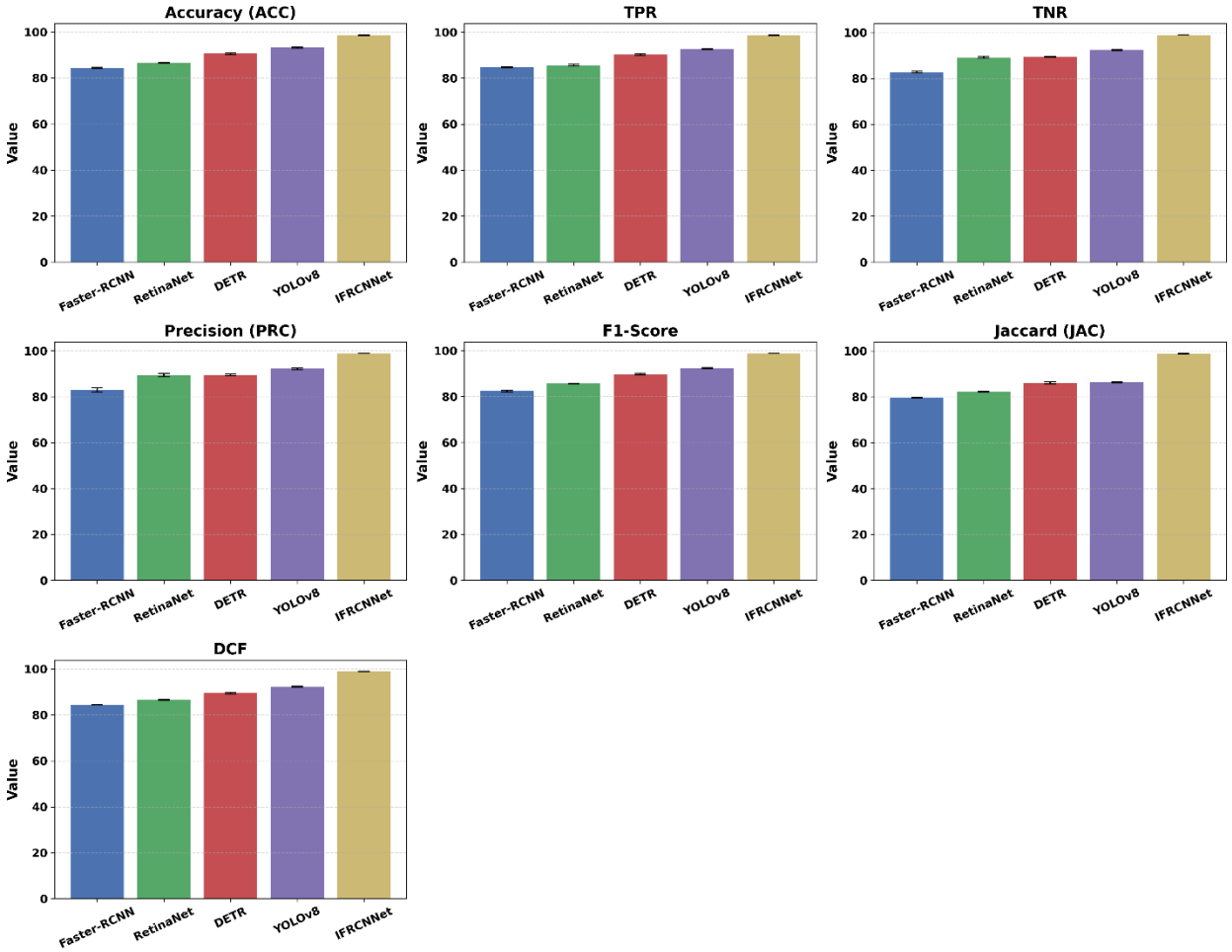

Fig.7: Comparative quantitative bar plot analysis between existing object detection methods and the proposed IFRCNNet across seven evaluation metrics

Figure 7 presents a comprehensive comparative bar plot that analyzes the existing object detection methods and the proposed IFRCNNet. This plot includes data for all seven metrics listed in Table 4. It is observed that Faster R-CNN has limited performance across nearly all metrics. Furthermore, the RetinaNet and DETR demonstrate moderate improvements but still exhibit fluctuation in TPR, TNR, and JAC. Besides this, YOLOv8 is achieving satisfactory mean values. However, the error bars reveal that its performance is not consistently stable across different samples. In view of this, the proposed IFRCNNet delivers the highest mean values on every metric with the smallest error bars. It demonstrates superior accuracy, sensitivity, specificity, precision, and spatial localization capability. Moreover, the reduced standard deviation confirms IFRCNNet's generalizability and reliability across the heterogeneous CMF dataset. Overall, the qualitative and quantitative analysis confirms that the proposed method outperforms the existing object detection method.

5. Conclusion

This paper presents a novel IFRCNNet deep framework for CMF detection and classification. The deep framework comprised a modified backbone module employing ResNet101 for fine feature extraction. Moreover, a generalized intersection over union was used as a new bounding box loss to achieve the precise detection of the forged region of interest. These two major refinements were introduced at the architecture level. The efficacy of the proposed IFRCNNet was assessed for the CoMoFoD, MICC-F220, modified CASIA, and COVERAGE CMF datasets. Moreover, augmentation was employed to increase and balance the uneven number of images of these datasets. The quantitative study was presented in terms of mean ± SD, 95% CI, paired t-tests, and p-values. The proposed IFRCNNet attained a mean accuracy of 98.56 ± 0.14. The proposed method surpasses the existing object detection methods, such as Faster-RCNN, RetinaNet, DETR, and YOLOv8, by an accuracy gain of +16.77%, +13.78%, +8.82%, and +5.71%. Furthermore, the qualitative analysis demonstrates that the proposed IFRCNNet has high classification accuracy and a better detection of forged instances. Moreover, it can be concluded that the proposed IFRCNNet is the best-suited deep approach to detect and classify the CMF instances in a single framework. Future work may concentrate on enhancing accuracy and generalizability for other forms of image forgery.

References

[1] Q. Zeng, H. Wang, Y. Zhou, and R. Zhang, "Dual-decoding branch contrastive augmentation for image manipulation localization," *Knowledge-Based Syst.*, vol. 309, no. June 2024, p. 112776, 2025, doi: 10.1016/j.knosys.2024.112776.

[2] J. Ravula and N. Singh, "GAN-ViT-CMFD: A novel framework integrating generative adversarial networks and vision transformers for enhanced copy-move forgery detection and classification with spectral clustering," *Intell. Syst. with Appl.*, vol. 26, no. February, p. 200524, 2025, doi: 10.1016/j.iswa.2025.200524.

[3] D. Pawar, R. Gowda, and K. Chandra, "Image forgery classification and localization through vision transformers," *Int. J. Multimed. Inf. Retr.*, vol. 14, no. 1, pp. 1–11, 2025, doi: 10.1007/s13735-025-00358-8.

[4] L. Su, C. Dai, H. Yu, and Y. Chen, "IFE-Net: Integrated feature enhancement network for image manipulation localization," *Image Vis. Comput.*, vol. 153, no. September 2024, p. 105342, 2025, doi: 10.1016/j.imavis.2024.105342.

[5] X. Yang, X. Chai, Z. Gan, L. Cao, and Y. Zhang, "MSHRT-Net : Multi-scale hierarchical residual transfer network for image manipulation detection and localization," *Neurocomputing*, vol. 648, no. February, p. 130788, 2025, doi: 10.1016/j.neucom.2025.130788.

[6] R. Elbarougy, O. Abdelfatah, G. M. Behery, and N. M. El-Badry, "Deep learning-based method for detection of copy-move forgery in videos," *Neural Comput. Appl.*, vol. 8, 2025, doi: 10.1007/s00521-025-11009-8.

[7] M. Zanardelli, F. Guerrini, R. Leonardi, and N. Adami, "Image forgery detection: a survey of recent deep-learning approaches," *Multimed. Tools Appl.*, vol. 82, pp. 17521–17566, 2022, doi: 10.1007/s11042-022-13797-w.

[8] M. A. Kiruthiga, H. Rudravaram, M. Priyanka, and J. Sangeetha, "Enhanced copy-move forgery detection using dual stage faster R-CNN and bayer filter analysis," in *AIP Conference Proceedings*, 2025, vol. 3300, no. 1, p. 20096.

[9] C. Basavaraj and P. V. B. Reddy, "Hybrid Teacher Learning Optimization Enabled Deep Learning for Copy-Move Image Multiple Forgery Detection," 2025, doi: 10.1177/24056456251320116.

[10] M. Tamilselvi, "Evaluation of Superior Accuracy of Faster RCNN over Conventional Forensic Techniques in Image Tampering Detection," *2024 2nd Int. Conf. Self Sustain. Artif. Intell. Syst.*, pp. 70–74, 2024, doi: 10.1109/ICSSAS64001.2024.10760523.

[11] K. Huynh and T. Ly, "An efficient model for copy-move image forgery detection," vol. 18, no. 2, pp. 181–195, 2025, doi: 10.1108/IJWIS-04-2022-0088.

[12] W. Li, B. Li, K. Zheng, S. Li, and H. Li, "Document image forgery detection and localization in desensitization scenarios," *Signal Processing*, vol. 238, no. May 2025, p. 110123, 2026, doi: 10.1016/j.sigpro.2025.110123.

[13] I. G. Atak and A. Yasar, "Image forgery detection by combining Visual Transformer with Variational Autoencoder Network," *Appl. Soft Comput.*, vol. 165, no. January, p. 112068, 2024, doi: 10.1016/j.asoc.2024.112068.

[14] M. Sabeena and L. Abraham, "Convolutional block attention based network for copy - move image forgery detection," *Multimed. Tools Appl.*, vol. 83, no. 1, pp. 2383–2405, 2024, doi: 10.1007/s11042-023-15649-7.

[15] U. Latnekar, "An efficient image forgery and region detection using LogDIoU-Faster RCNN," *Sens. Imaging*, vol. 24, no. 1, p. 25, 2023.

[16] X. Qiu, C. Shi, X. Li, M. Wu, S. M. Abdullahi, and Y. Liu, "S-Faster R-CNN : Intraspectral Similarity Learning for Audio Copy-Move Forgery Localization in IoT Security," *IEEE Internet Things J.*, vol. 12, no. 15, pp. 30185–30202, 2025, doi: 10.1109/JIOT.2025.3569678.

[17] X. Wei, Y. Wu, F. Dong, J. Zhang, and S. Sun, "Developing an image manipulation detection algorithm based on edge detection and faster r-cnn," *Symmetry (Basel).*, vol. 11, no. 10, p. 1223, 2019.

[18] A. Dutta and A. Zisserman, "The {VIA} Annotation Software for Images, Audio and Video," 2019, doi: 10.1145/3343031.3350535.

[19] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, "CoMoFoD - New database for copy-move forgery detection," *Proc. Elmar - Int. Symp. Electron. Mar.*, pp. 49–54, 2013.

[20] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-based forensic method for copy-move attack detection and transformation recovery," *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3 PART 2, pp. 1099–1110, 2011, doi: 10.1109/TIFS.2011.2129512.

[21] Y. Zheng, Y. Cao, and C. H. Chang, "A PUF-Based Data-Device Hash for Tampered Image Detection and Source Camera Identification," *IEEE Trans. Inf. Forensics Secur.*, vol. 15, no. X, pp. 620–634, 2020, doi: 10.1109/TIFS.2019.2926777.

[22] R. B. Wen, Y. Zhu, "COVERAGE – A Novel Database for Copy-move Forgery Detection," *Ieee*, pp. 161–165, 2016.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Adv. Neural Inf. Process. Syst.*, vol. 28, pp. 91–99, 2015.

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2016-Decem, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[25] T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 936–944, 2017, doi: 10.1109/CVPR.2017.106.

[26] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2019-June, pp. 658–666, 2019, doi: 10.1109/CVPR.2019.00075.

[27] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[28] P. Kumar, A. Kumar, S. Srivastava, and Y. Padma Sai, "A novel bi-modal extended Huber loss function based refined mask RCNN approach for automatic multi instance detection and localization of breast cancer," *Proc. Inst. Mech. Eng. Part H J. Eng. Med.*, p. 09544119221095416.

[29] A. Kumar, P. Kumar, M. Mahto, and S. Srivastava, "Breast Cancer Detection and Localization Using a Novel Multi Modal Approach," *IEEE Trans. Instrum. Meas.*, vol. 74, no. Ml, pp. 1–13, 2024, doi: 10.1109/TIM.2024.3502883.

[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

[31] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable Detr: Deformable Transformers for End-To-End Object Detection," *ICLR 2021 - 9th Int. Conf. Learn. Represent.*, pp. 1–16, 2021.

[32] R. Varghese and M. Sambath, "Yolov8: A novel object detection algorithm with enhanced performance and robustness," in *2024 International conference on advances in data engineering and intelligent computing systems (ADICS)*, 2024, pp. 1–6.