Efficient Approach for the Localization of Copy-Move Forgeries Using PointRend with RegNetX

Mahmoud H. Farhan [1]* , *Khalid Shaker [1]* , *Sufyan Al-Janabi [2]*

[1] Department of Information Systems, College of CS & IT, University of Anbar, Ramadi, Iraq.
[2] Department of Networks Systems, College of CS & IT, University of Anbar, Ramadi, Iraq.
*Corresponding Author.

Abstract

Digital images are one of the modern era's dominant sources of information and communication. However, an image can simply be altered with the existence of several tools for image editing. These altered images can transmit across platforms of social media to influence some people in society and may have both positive and negative effects. Therefore, the development of technology becomes a necessary issue to detect and localize a forgery in an image. Copy-move forgery (CMF) is one of the most popular forgeries. In CMF, the new forgery image is created by copying a part presented in an image and placing it at a different location on the same image. This paper proposes a PointRend as a technique to localize copy-move forgeries. This work also presents the PointRend framework with a lightweight backbone model RegNetX (PointRend-RegNetX) to detect such forgeries. From the comparative analysis of the proposed technique with ResNet-50 backbone on two standard datasets, it has been shown that the proposed model (PointRend-RegNetX) is superior in MICC F-220 and MICC F-2000 datasets for images that contain copy-move forgeries. In instances segmentation of forged regions, the improved model (PointRend-RegNetX) has achieved a mean average precision (mAP) of 88.5% on MICC F-220 dataset and 86.4 % on MICC F-2000 dataset.

Keywords: Image forgery, copy-move forgery, deep learning, PointRend, RegNetX backbone

Introduction

Digital images are one of the dominant sources and are used in different areas, such as education systems, social media platforms, businesses, public health services, armed forces, judicial inquiries, political blogs, and so on. Besides that, the availability of editing tools for the images either commercially such as Photoshop, or even available for free such as GIMP. This permits anyone with a computer or mobile phone to easily manipulate any image without additional tools. As a consequence, altered images become a primary source for impacting individuals and society by spreading confusing news or information. Therefore, in digital image forensics, an effective and robust technique for detecting image forgery is of paramount importance [1].

The detection techniques of image forgery can be categorized into two types: active and passive. In an active method, such as a watermark or signature, certain information is embedded during the image creation. Image tampering with these methods is detected by analyzing the signature or watermark.

Although these methods allow protecting an image from stealing, their application is still limited because of the need for intervention of humans to recover the original image. On the contrary, manual processing does not require in passive techniques [2]. The changes in the whole image and local features are identified in passive technique detection. The validity of an image is determined by verifying its content and structure. Passive detection techniques are categorized into dependent and independent forgery detection. Dependent forgery detection techniques are holding two kinds of forgeries namely copy move and image splicing.

In a copy-move forgery (CMF), the new forgery image is created by copying a part presented in an image and placing it at a different location on the same image. While the merging of two or more images to produce a new image is known as Image splicing forgery [3]. Generally, the detection of CMF can be classified into three categories: keypoint-based techniques, block-based techniques, and deep learning-based techniques. Sometimes there is hybridization among two or more previous categories. On the other side, the localization of CMF is more challenging than the detection of CMF, where the detection seeks only to know if a whole image or part of it is fake or not. While localization is required to find the forged regions exactly [4]. The main contributions of this work can be summarized as follows: 1) Developing an architecture for copy-move forgeries localization based on the PointRend model with Resnet-50 as a backbone (PointRend-Resnet). 2) Replacing the backbone Resnet-50 with a lightweight model RegNetX 3.2G (PointRend-RegNetX). Then, the Evaluation of the two models on different two datasets namely MICC F220 and MICC F2000, Finally, the analysis of the proposed model PointRend-RegNetX in comparison with PointRend-Resnet on different standard datasets.

The remainder of this paper is structured as follows. The related work is reviewed in Section 2. The details of improved architecture are described in Section 3 and its subsection. The analysis of the experimental results is illustrated in Section 4. Finally, the conclusion of this work is presented in the last section.

Materials and Methods

Related Work

Deep learning helps researchers of computer vision to achieve better accuracy in different topics such as the classification of images, object identification, and semantic segmentation, compared to traditional techniques of computer vision [5, 6]. In addition, the algorithms of deep learning are more versatile and more domain-specific. The deep learning model detected specific elements automatically based on several type of deep neural networks, these types include: deep belief network, Autoencoder networks, and convolutional neural network (CNN). The most commonly network used in vision applications from these neural networks is a CNN, and most deep learning applications in the detection of CMF forgeries depends mainly on CNN, whereas CNNs offers an excellent performance, through the combination of convolution and rectification operations, i.e. linear and non-linear filtering.

Recently, many techniques based on the deep learning to detect CMF have been presented. Furthermore, the focus in the localization of image forgery has been on constructing a model instead of looking only at certain image features. Bondi et al. [7] detected and localized image forgery by exploiting CNN that extracted the features of camera model characteristic from image patches. Then, by clustering those features, the forged regions are detected and localized. Goel et al. [8] presented a deep learning approach to detect CMF which used a dual-branch CNN. For each branch, a different kernel sizes are employed in this approach to extracted multi-scale features. The two branches features are concatenated and passed to global max-pool layer to extracted the dominant features, which are used for CMF detection. Shi et al. [9] also used a dual CNN that combined CNN of Spatial domain (sub-SCNN) with CNN of Frequency domain (sub-FCNN). The image forgery located by Sub-SCNN of input image, while

Sub-FCNN of input image is used to take statistical features of three rates of DWT.

More specifically, pretrained CNN models that were learned over large datasets (such as ImageNet), and their weights have been used on other applications by fine-tuning them by modifying only the last layers. In computer vision areas, various deep-learning networks model have been proposed. These models included: AlexNet in 2012 [10], VGGNet [11] in 2014, GoogLeNet [12], and ResNet [13] in 2015. All of them achieved superior accuracy. Several pretrained CNNs have been used in recent years to detect a tampered image. The AlexNet model was used by Muzaffer and Ulutas [14] to detect the CMF in image. In this method, the features extracted using mid-level of pretrained AlexNet model. Samir et al. [15], presented an optimized AlexNet model for detection and localization of CMF. The optimized model replaced both Local Response Normalization and Maxout Function by Batch Normalization and Softmax Function. The VGG16 model used with Simple Linear Iterative Clustering (SLIC) algorithm by [16] to detect a CMF, and with Google Inception by [17] to detect and localize a CMF and generate mask of forgery regions. Mask R-CNN and the Sobel filter are utilized by Wang et al.[18] to detect and localize instances of copy-move and image splicing forgeries. By implementing the Sobel filter, the predicted

masks are able to detect gradients that closely resemble those of the actual mask. Ahmed et al.[19] has introduced a new method for detecting image forgeries using a novel deep learning backbone architecture named ResNet-conv. ResNet-conv is derived by substituting the feature pyramid network in ResNet-FPN with a group of convolutional layers. The study explores two ResNet variations, ResNet-50 and ResNet-101, for this purpose. Kadam et al.[3] has demonstrated the effectiveness of Mask R-CNN with MobileNet, a compact model, in detecting and recognizing instances of copy-move and image splicing forgeries. The researchers conducted a comparative evaluation of their proposed approach against ResNet-101 using seven widely-used benchmark datasets.

In this work, the use of PointRend models for detecting copy-move forgeries has shown great promise in terms of accuracy. Compared to previous studies, PointRend techniques have demonstrated improved performance in accurately localizing instances of image tampering and manipulation. Furthermore, the use of RegNetX as the backbone for PointRend models has further improved the accuracy and computational efficiency of the approach. RegNetX is designed to achieve high accuracy while being computationally efficient, making it suitable for deployment on resource-constrained devices.

Results and Discussion

PointRend is a modern algorithm that can detect accurately the target object and segment the target precisely. An Improved PointRend model is proposed and it applied as a vision system to localized copy-move forgery regions. This work uses the RegNetX network structure as a backbone instead of Resnet-50 which represents the original backbone. The framework of the improved PointRend is shown in Fig 1. The copy-move forgery image represents the input of the proposed model. The feature maps are extracted by applying the backbone (RegNetX) network followed by the feature pyramid network (FPN) on the input image. Then on these features maps, the region of interest

(ROI) is generated via the region proposal network (RPN) stage.

Afterward, the exact spatial locations of the ROIs are maintained using ROIAlign to output them in a fixed dimension. Then, the output of ROIAlign is divided into two branches: One of them is given as input to a fully connected network with two layers and returns the class and box of the object. While the other branch is fed into a point-rend head, which is represented as fine-grained features, and also passed to a shallow prediction head to generate coarse prediction features. Eventually, the combination of both is given as input into a multi-layer perceptron (MLP) to generated the mask of forgery region.
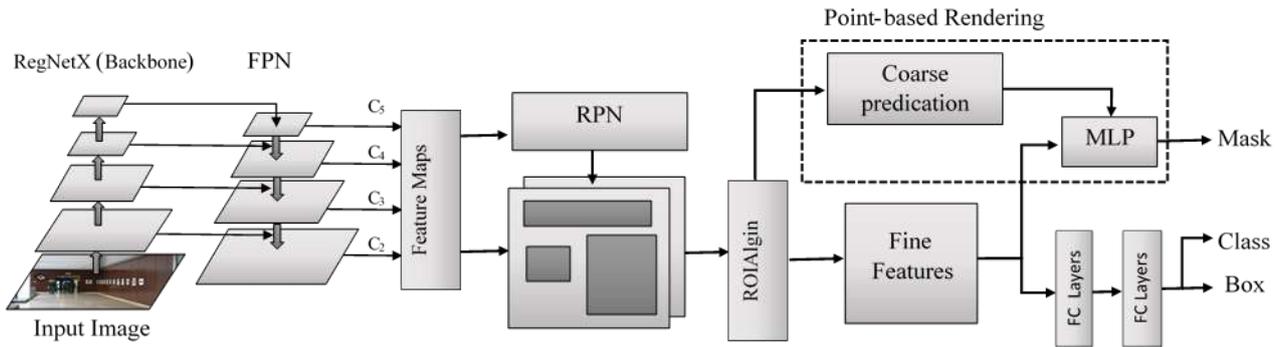
Figure 1. The structure of Overall framework of the improved PointRend

PointRend

PointRend technique (Point-based rendering) was presented by Kirillov et al. [20] to solve the image segmentation process by handling it as a rendering problem and obtaining a sharp and smooth boundaries. Any CNN based can be applied to the task of instance segmentation to handle the object boundaries from coarse-to-fine in an anti-aliasing mode [21]. This technique select a set of non-uniform points adaptively using a subdivision strategy to compute labels, and these points should be distributed widely in the image with high-frequency area [22]. Fig 2, shows the PointRend model that applied to perform the instance segmentation, which will be used to localize the region of the copy move forgery.
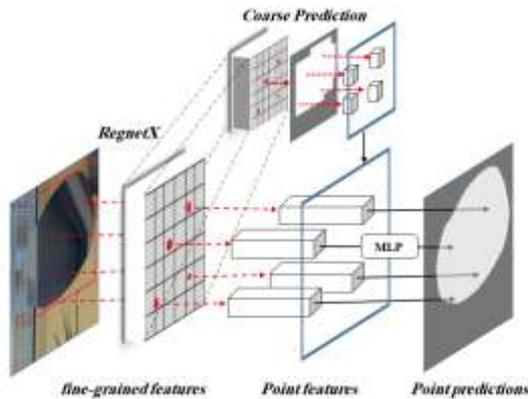
Figure 2. Instance segmentation using PointRend model.

The PointRend consists of three main modules which are: point selection, extraction of point-wise feature, and point head [20]. During the point selection module, the output image rendered iteratively with a coarse-to-fine mode. Firstly, rather than over-calculating all pixels, a few points are selected for prediction the segmentation label [23]. Then for the next iterations, PointRend used a bilinear interpolation operation for upsampling the predicted segmentation, and then select a set of points n_i^* on this denser grid by Eq 1 [21]:

where p(n_i) is the probability of this set of points n_i, which are relating to a binary mask. Point-wise features module is performed when the most uncertain points are selected, and it constructed based on these points by concatenating the features of coarse predicted and fine-grained. The features of coarse predicted is a 2-class prediction, which is in the region represents a vector of 2-dimensional for each point. Also, this feature conveys context that is more general and globalized. On the other side, the bilinear interpolation is used to extract the fine-grained features for each feature maps, which represents fine detailed segmentations vector [21]. Finally, PointRend used the point head module to predict the labels for each point based on previous point-wise features using a Multi-Layer Perceptron (MLP) neural network. There are three hidden layers in MLP network with 256 channels. In each layer, the input vector supplied with 256 output channels and the features of two coarse predictions for the next layer. ReLU has been used inside MLP network, while sigmoid applied to output of MLP [20]. Fig 3 illustrates one step of PointRend procedure. The bilinear interpolation used to perform upsampling by 2× on the grid (4×4). Then on the finer grid (8×8), N most uncertain points are selected (i.e. 28 point black dots), and recover the point-wise feature detailed.

$$n_i^* = \arg\min_{n_i} \left| p(n_i) - \frac{1}{2} \right| \qquad 1$$
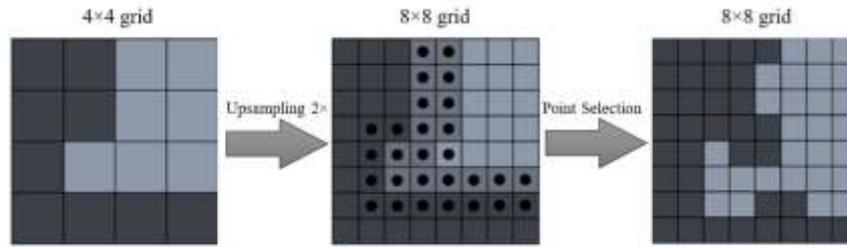
Figure 3. One Step of PointRend Process.

RegNetX

Neural Architecture Search (NAS) is a network search technology that has been very popular in recent years. NAS can find a set of best parameters (compound coefficient), and hence discover the best model under a convinced computational cost based on the neural structure search technology, but it requires high computing resources. A traditional NAS method relies on instances of an individual network which means one network at a time but has some weaknesses such as there being multiple methods to adjust the parameter and poor capability to generalize. Therefore, the scholars went towards estimating the overall network design space, which means estimating the relationship among a set of elements such as the depth and width of the network design comparable to the network goal [24]. NAS technology is used in RegNetX model. The structure of the RegNet network is shown in Fig 4.
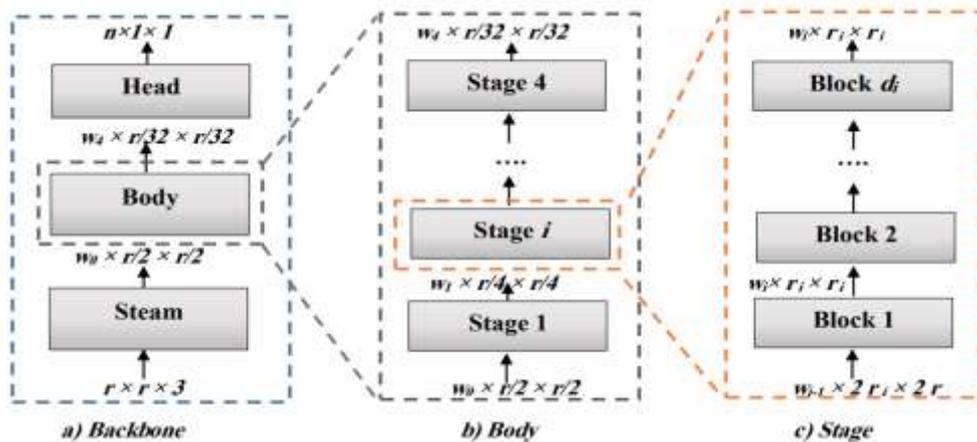
Figure 4. The Structure of RegNetX model: (a) Backbone; (b) body; (c) stage.

Mainly, the RegNetX backbone network is composed of three primary parts which are: stem, body, and head (see Fig 4a). The main focus of this model is on the body network while keeping the steam and head networks as simple as possible. The stem is composed of the following properties: the convolution layer that includes default Batch Normalization (BN), ReLU as activation function, 3 × 3 as convolution kernel size, 2 as a step length, and 32 convolution cores. Fig. 5b shows the body structure is a stack that composed of four stages. The width and height of the input matrix were reduced in half from the original matrix after each stage (see Fig 4c). A series of block stacks are composed to construct each stage. There are group convolutions and conventional convolutions with a step of 2 in the first block, and with a step 1 in the remaining blocks in each stage. Finally, in the classification network, the head is composed of two layers', which are global average pooling followed by full connected [24].

The block structure is shown in Fig 5, where Figs 5a , 5b show the case of step with stripe = 1 and stripe = 2, respectively. The block in RegNetX model can be seen as the same as the ResNext model block. There are two branches of the model block one with stripe 1 and another with stripe 2. In both branches, the block includes ordinary convolution (1 × 1) and group convolution (3 × 3) including BN and ReLU for each, followed by ordinary convolution (1 × 1) that includes BN. Downsampling is performed through an ordinary convolution (1 × 1) and group convolution (3 × 3) when the stripe equal to 2. The

resolution (height and width) of the characteristic matrix is represented by r in Fig 5. While s, w, g and b are a step distance, characteristic matrix channel, group width in the group convolution, and bottleneck ratio respectively. The r in input and output remains the same when s equals 1, while the output r down to half of the input r when s equals 2. Also, the output channel w_i is reduced to 1/b of the input channel w_{i-1} [24].
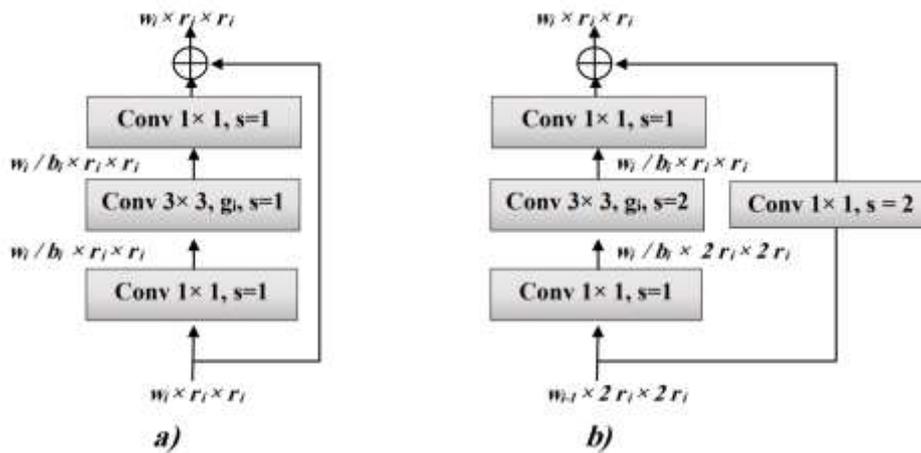
Figure 5. The Structure of the RegNetX block: (a) block with stride = 1; (b) block with stride = 2.

When comparing the RegNetX model with the standard backbone (e.g. ResNet), one can notice that it inherits its advantage by extra explores of the instances of designing space (stages and blocks) to obtain an appropriate structure of the whole backbone network, in addition to the shortcut connection of it. In this work, the hardware allowed only to use RegNetX-3.2GF as a backbone of the PointRend-RegNetX model for forgeries detection. Table 1 summarizes the hyperparameter settings that includes: the output width and number of blocks of each, and group ratio for both standard (ResNet) model and RegNetX(RegNetX-3.2GF) model. Obviously, this table below illustrates that RegNetX has compact output width and elastic growth ratio of the width of output between the successive stages. Furthermore, the size of RegNetX model is more lightweight compared to ResNet as a result of the implementation of the group convolution at each block [25].

Table 1. Comparison of ResNet and RegNetx models.

Properties		Backbone Network	
		ResNet-50	**RegNetx-3.2GF**
	S_1^*	256	96
Stage	S_2	512	192
Output Width	S_3	1024	432
	S_4	2048	1008
	S_1	3	2
	S_2	4	6
Num of Blocks	S_3	6	15
	S_4	3	2
Group Ratio(g_i)		-	48

*S1 to S4 indicated to the stages from 1 to 4

Feature Pyramid Network (FPN)
To increase the accuracy, the Feature Pyramid Network (FPN) has been used as a feature extractor. This step constructs new feature maps at a multi-scale that provide information better than those obtained by the ordinary feature pyramid. FPN is used because of the difficulty of small object detection [26]. When a backbone network is used alone, the most superficial layers (higher resolution) calculated low-level semantics (characteristics) such as contours, lines, etc. While the higher level semantic (lower resolution) was calculated in the deepest layers that were used to identify the objects. Therefore, FPN aims to solve the correlation between

resolution and semantic level by adding a second pyramid (top-down) to a backbone [27]. By adding the second pyramid of the same image the detection of the objects can be solved easily. The feature maps at several stages in FPNs are effectively fused, which utilized the shallow and deep feature maps to detect tampered objects [26]. The PointRend model was improved by integration between RegNetX and FPN, which enhanced accuracy with fewer parameters comparable to the conventional PointRend model that used Resnet.

Region Proposal Network (RPN)
Each of the feature maps is directed to the Region Proposal Network (RPN), which is a trivial deep neural network. The Region of Interest (ROI) is directly generated on the feature map together with RPN. RPN has a distinct structure involving classifiers and regressors. An unfixed size image is considered as input to RPN stage, which results in object score in conjunction with a proposal rectangular for a set of objects. It finds out whether the proposal rectangular fall into the foreground or the background. In the RPN, a fixed number of object boxes (known as anchor boxes) have been producing by using a sliding windows on the feature maps based on predefined scale and an aspect ratio for each pixel. The first coordinate correction has been performing for all the anchors which fall into to the foreground in a RPN. At the sliding window, an anchor correlated with a scale and aspect ratio has been positioned. The proposals of object are then being equipped with two connected layers one for object identification as a Classifier and another for generation of bounding box as a Regressor [26] (see Fig 6).
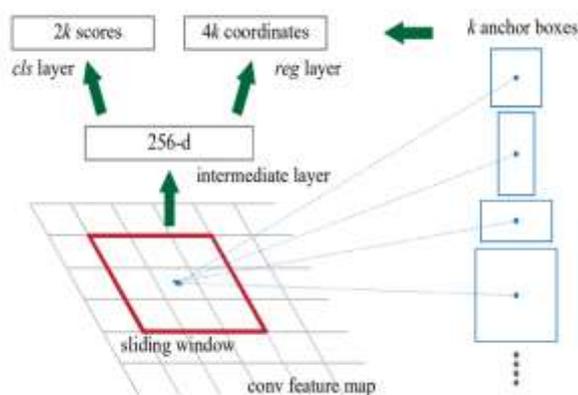
Figure 6. The architecture of RPN

RoIAlign
The Region of Interest (ROI) is generated on the feature map with RPN directly. The ROI alignment (ROIAlign) is vital for the representation of feature map, the prediction of class along bounding box of the object and segmentation mask, and conserving the regularity through the convolutional [26]. In first, the different ROIs are divided into equal number of bins. For each bin, The ROI boundaries may do not match the feature map boundaries nor the boundaries of the other bins. This problem is solved by the alignment of the ROI boundaries at different bins with the feature map boundaries [28]. ROIAlign used a bilinear interpolation to estimate the precise indexes of the feature maps. Smaller regions based predetermined number are producing by dividing the proposal ROI. Hence, four points are sampling for each region, and the feature value for each sample are computing using bilinear interpolation. Also, the feature map returned with a fixed size by maintaining exact spatial locations using ROIAlign [3] (see Fig 7).
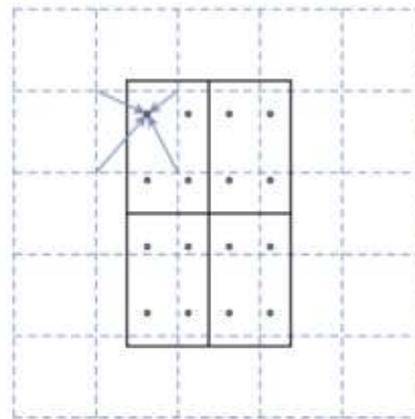
Figure 7. Alignment of a ROI. The ROI is represented by solid squares (2×2 in this case) in the dashed grid, which is a feature map [3].

Experimental Analysis
Two models of the PointRend technique (with Resnet50 and RegNetX 3.2G backbones) have been implemented and coded separately using PyTorch framework in Python language. The experiments have been executed on a Laptop with the following specifications: 64-bits processor of AMD Ryzen 74800H, RAM of 16 GB, and GPU driver of 4 GB (GTX 1650Ti). The optimizer used for each model is stochastic gradient descent (SGD) with learning rate equal to 0.02, momentum=0.9 and weight decay=0.00005. Both models are trained for 48 epochs with batch size equal to 10 in the training phase. At the end of every 2 epochs of training, the validation of system is performing.
There are two different dataset used to evaluated the models namely MICC F-220 and MICC F-2000

datasets [29]. These datasets split into two sets: training set with 70% of the whole dataset and the remaining treats as a test set. The obtained results of each model based on the test set. In following subsections, the comparison of both models was performed basing on the mean average precision at different IoU thresholds and losses functions.

Evaluation Metrics
For both models, the evaluation metrics of MS COCO are used to evaluate the effectiveness and performance of these models. The Average Precision (AP) of the result for object detection and instance segmentation defined over the IoU, which is calculated in object detection through the overlap ratio between predicted and ground truth bounding boxes, while it is calculated by the overlap ratio between predicted and ground truth masks as follows:

$$IoU_{BBox} = \frac{Bp \cap Bg}{Bp \cup Bg} \qquad 2$$

$$IoU_{mask} = \frac{Mp \cap Mg}{Mp \cup Mg} \qquad 3$$

where Bp, Bg, Mp and Mg are predicted bounding box, ground truth bounding box, mask and ground truth mask respectively. The value precision and recall over IoU threshold, are defined by object detection and instance segmentation via:

$$Precision = \frac{TP}{TP+FP} \qquad 4$$

$$Recall = \frac{TP}{TP+FN} \qquad 5$$

where TP, FP, and FN are true positive, false positive, and false negative, respectively. Hence, the value of AP calculated through the following Eq:

$$AP = \int_0^1 P(r)\, dr \qquad 6$$

where P and r are the values of precision and recall value. AP can be defining as the area under the precision-recall curve. There are three value of AP according to the threshold that will be used : 1) AP is computing by taking averaging of 10 IoUs thresholds with the ranges from 0.5 to 0.95 and stride between these threshold equal to 0.05. 2) AP_{50} calculated at IoU threshold equal to 0.5, and 3) AP_{75} which calculated at 0.75 IoU threshold. Another important metrics which will be used to comparing the performance the models in this work is mean Average Precision (mAP). The value of mAP is computed by calculating the mean value of a set of AP over different IoU thresholds for all classes, based on different challenges (object detection and instance segmentation). The obtained results are summarized in Table 2.

Table 2. Comparison of ResNet and RegNetx models performance.

Metrics		PointRend-ResNet		PointRend-RegNetX	
		MICC-F220	MICC-F2000	MICC-F220	MICC-F2000
Bounding Box of object detection	mAP	0.73	0.744	0.84	0.859
	mAP_{50}	0.973	0.92	1.00	0.989
	mAP_{75}	0.696	0.773	0.887	0.948
Instance Segmentation	mAP	0.778	0.738	0.885	0.864
	mAP_{50}	0.987	0.931	1.00	0.989
	mAP_{75}	0.845	0.772	1.00	0.955

The above analysis compares the performance of two popular object detection models, ResNet50 and RegNetX, on two different types of tasks, bounding box object detection and instance segmentation, using the mean Average Precision (mAP) metric at different IoU thresholds. The results showed that RegNetX outperformed ResNet50 on both types of tasks and on both datasets, achieving higher mAP scores and perfect accuracy at high IoU thresholds.
For the bounding box object detection task, RegNetX achieved mAP scores of 0.84 and 0.859 on the two datasets, while ResNet50 achieved scores of 0.73 and

0.744. RegNetX also achieved perfect mAP_{50} scores on both datasets, indicating perfect accuracy at high IoU thresholds, and very high mAP_{75} scores, indicating high accuracy at higher IoU thresholds. This suggests that RegNetX is a highly effective model for bounding box object detection tasks, especially when high accuracy is required.
Fig 8 shows the comparative analysis of mAP (1st column), mAP_{50} (2nd column), and mAP_{75} (3rd column) of Bounding Box of the object on different two standard datasets (MICC F220 and MICC F2000) using PointRend with Resnet-50 and

RegNetX as a backbone for two classes (original and copy). Where the x-axis for each plots represented the epochs values and y-axis represent the corresponds mean Average Precisions of bounding

box of the object. The first row represents the obtained results on MICC-F220 dataset, while the second row represents the result obtained on MICC F-2000.
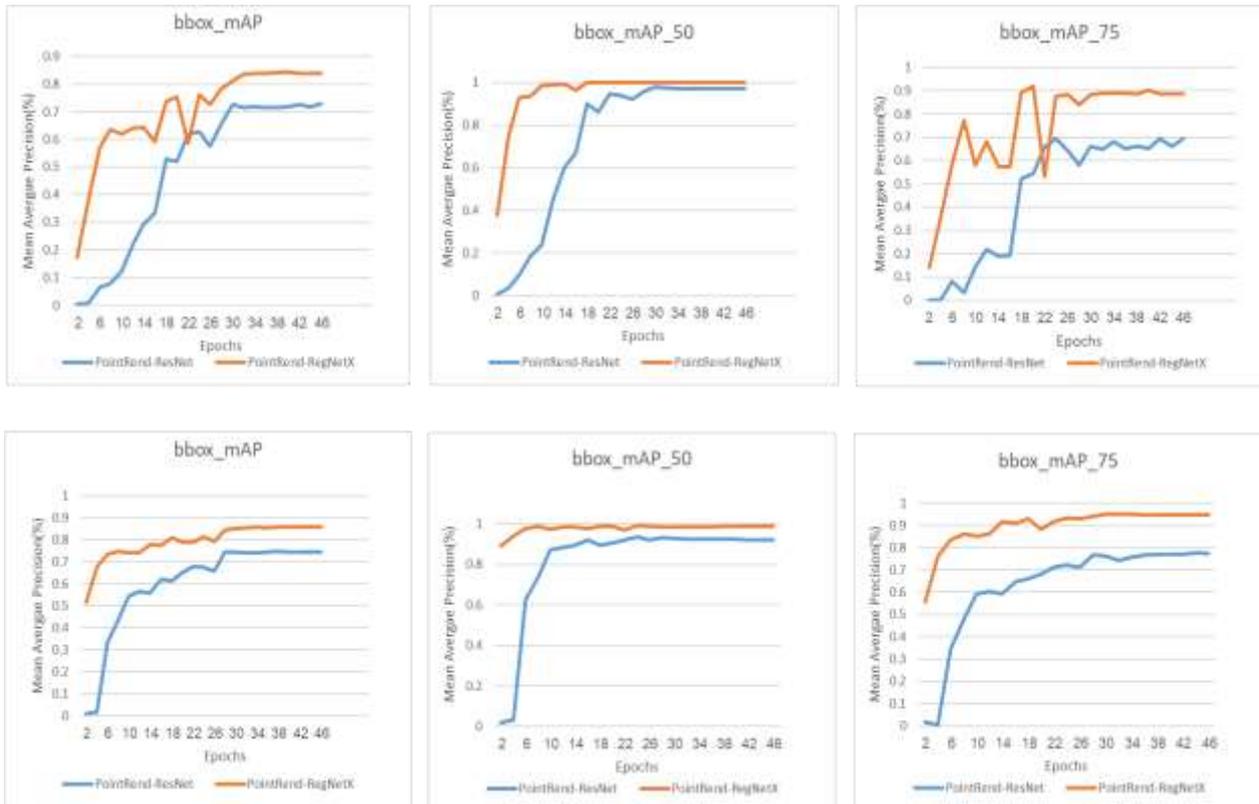
Figure 8. The mean Average Precision(mAP) of object Bounding Box(BBox) at different IoU thresholds of the two models for MICC-F220(First Row) and MICC-F2000 (Second Row) datasets.

For instance, segmentation task, RegNetX achieved mAP scores of 0.885 and 0.864 on the two datasets, while ResNet50 achieved scores of 0.738 and 0.778. RegNetX again achieved perfect mAP_{50} scores on both datasets, indicating perfect accuracy at high IoU thresholds, and very high mAP_{75} scores, indicating high accuracy at higher IoU thresholds. This suggests that RegNetX is also highly effective for instance segmentation tasks, which may be a more effective approach to object detection than traditional bounding box detection. Fig 9 shows the comparative analysis of mAP (1st column), mAP_{50} (2nd column), and mAP_{75} (3rd column) of instance segmentation on different two standard datasets (MICC F220 and MICC F2000) using PointRend with Resnet-50 and RegNetX as a backbone for two classes (original and copy). Here, the x-axis for each plots represented the

epochs values and y-axis represent the corresponds mean Average Precisions of instance segmentation. The first row represented the obtained results on MICC-F220 dataset, while the second row represented these result obtained on MICC F-2000. Figs 8, 9 illustrate the performance of the improved model(PointRend-RegNetX) exceeds the original model(PointRend-Resnet) on most IoU thresholds, and on different datasets. Also, the improved model achieves good results after the training with a few epochs comparable with the original model. Overall, these results suggest that RegNetX is a highly effective model for object detection tasks, achieving high accuracy and perfect scores at high IoU thresholds on both bounding box detection and instance segmentation tasks.
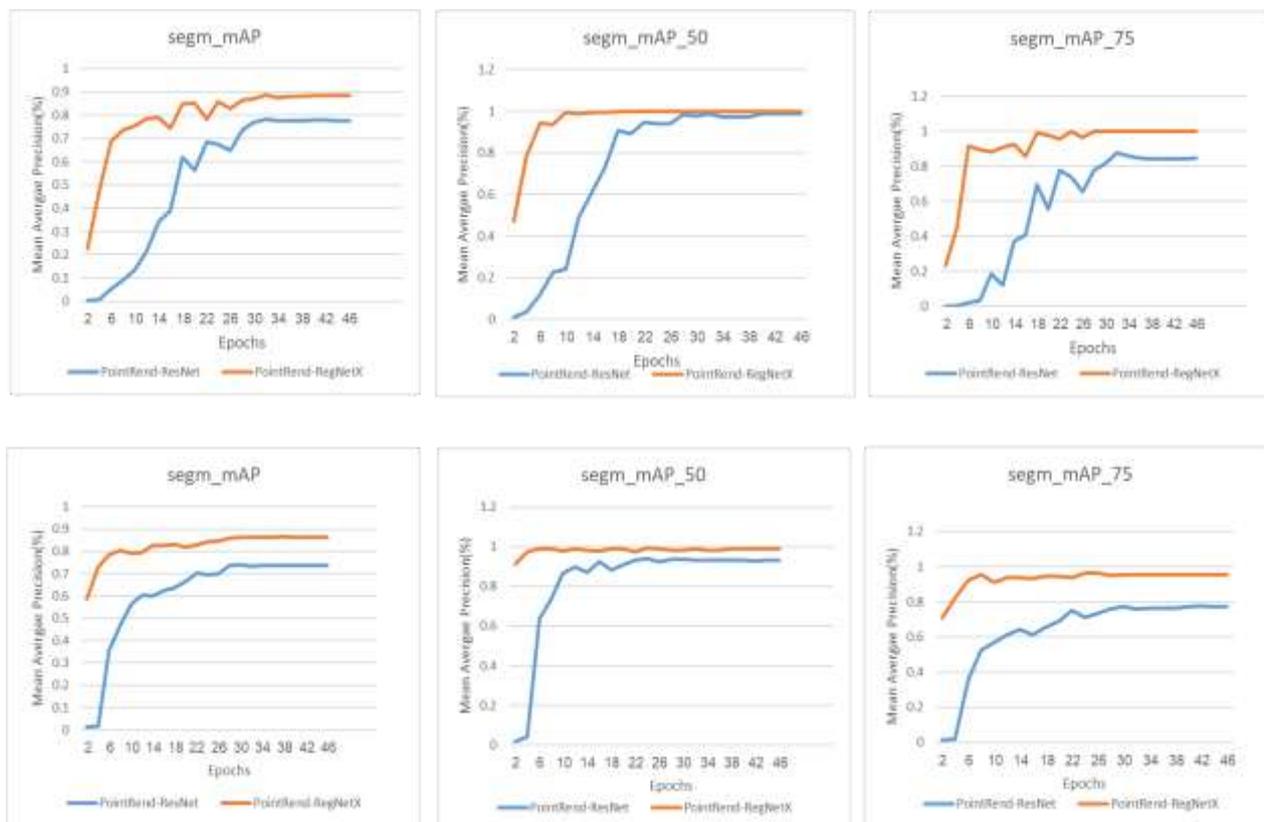
Figure 9. The mean Average Precision(mAP) of instances segmentation (segm) at different IoU thresholds of the two models for MICC-F220(First Row) and MICC-F2000 (Second Row) datasets.

Losses Functions

The overall architecture of the models consists of two stages mainly. In the first stage, RPN candidate the proposal bounding boxes for of original and forged regions. While the extraction of features from these proposal boxes involved in the second stage that also performing the classification, detection the bounding box location, and the mask for each proposal. The losses show the costs of the above neural networks training. In both models, there are seven losses that distributed as follows: 1) loss of RPN anchor class (L_{RPN}^{Cls}), 2) loss of RPN anchor bounding box (L_{RPN}^{BBox}), 3) loss of PointRend classification (L_{cls}), 4) loss of PointRend bounding box(L_{BBox}), 5) loss of PointRend Mask (L_{mask}) 6) loss of PointRend points (L_{point}) and 7) The overall loss(L).

In this paragraph a brief description of each them. L_{cls} illustrates how the model is good for predicting the correct class. This loss reflects if the model classifies the original and forged region and represents a Classification loss. Similarity, L_{RPN}^{Cls} predicted if an object is the particular object wanted to be found or not. L_{BBox} shows the distance between the bounding boxes of the predicted and the ground

truth. It used a Smooth_{L1} loss, and it represented a regression loss. Hence, it reflects how the model is good to detected whether the objects within an image. Hereby, L_{RPN}^{BBox} is used to illustrate how the model is well to detect the ROIs within an image. L_{mask} calculated based on masks corresponding to the right class for each ROI only. It used cross-entropy loss and reflecting the classification of pixel-wise. L_{point} illustrate how the model is good for predicting the uncertain selected points within the mask and it used cross-entropy loss. These losses are used for comparing the performance of the original model (PointRend-Resnet) and the improved model (PointRend-RegNetX).

In Figs 10, 11, 12, and 13, the x-axis represents the epochs values and y-axis represents the corresponds loss for both model. For each Fig, the first row represents the obtained results on MICC-F220 dataset, while the second row represents the result obtained on MICC F-2000, except the last Fig where the left for MICC-F220 and right for MICC F-2000. Fig 10, shows that the improved model is better for predicting the existence of the wanted object, and for detection the ROIs within an image. Also, the

superior of the improved model appears obviously in the remaining figures. Where the improved model is better for predicting the correct class and the distance between the bounding boxes of the predicted and the ground truth (see Fig 11). Also, it is better for predicating the masks corresponding to the right classes for each ROI only, and for predicting the uncertain selected points within the mask (see Fig 12). Finally, Fig 13, shows that the overall loss of the improved model is better the corresponding loss in the original model.
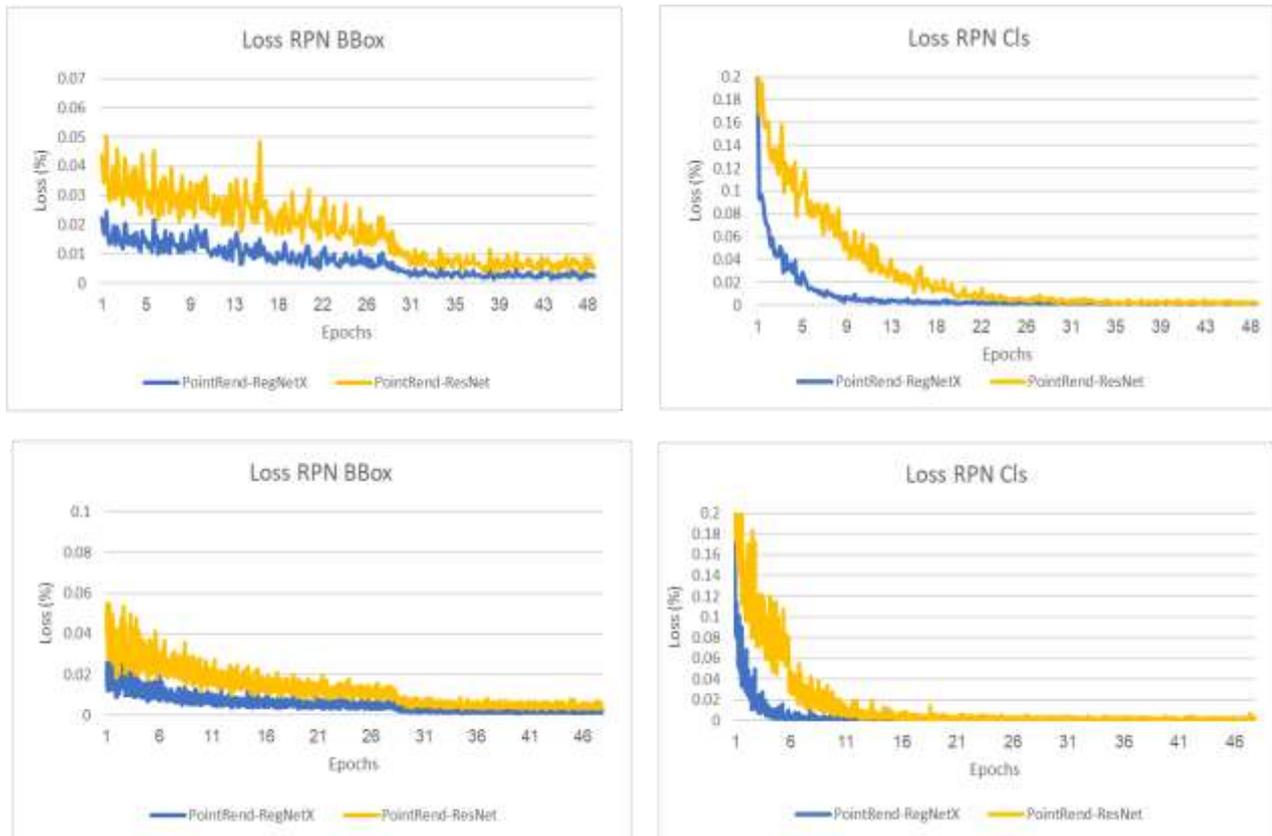
Figure 10. The obtained RPN losses of two models: L_{RPN}^{BBox} loss in first column, L_{RPN}^{Cls} loss in second column. MICC-F220(First Row) and MICC-F2000 (Second Row) datasets.
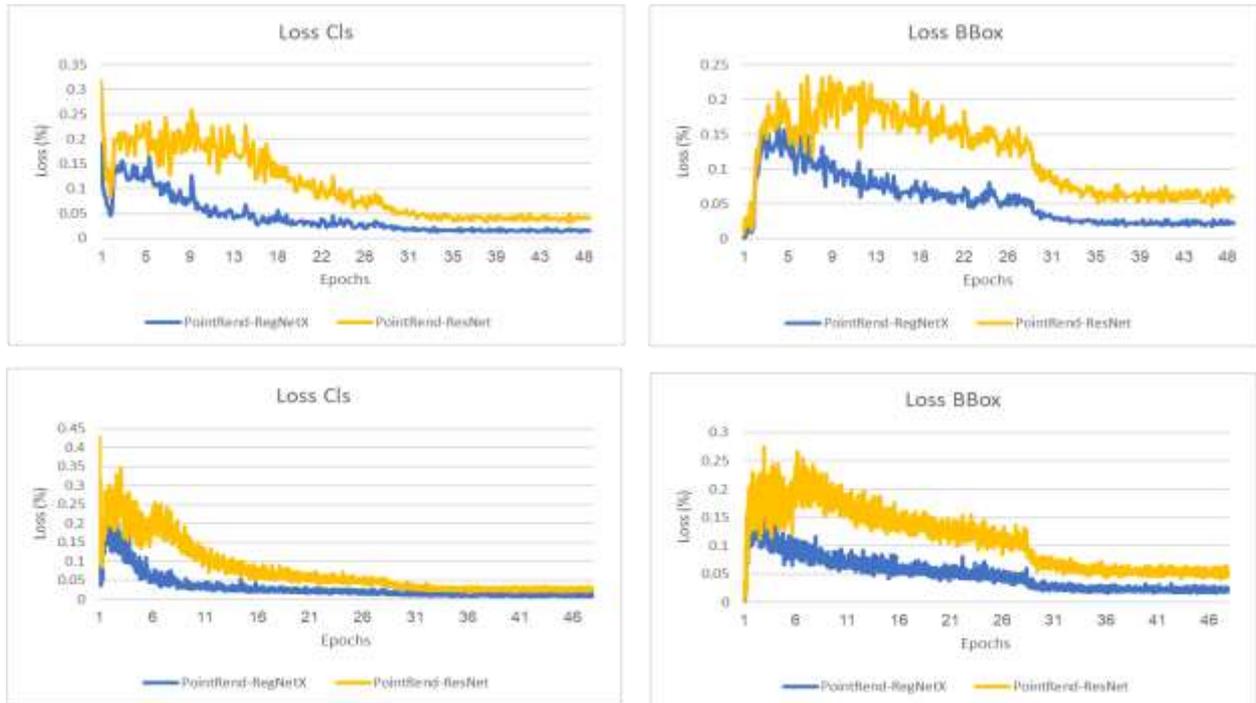
Figure 11. The obtained PointRend losses of two models: L_{Cls} loss in first column, L_{BBox} loss in second column. MICC-F220(First Row) and MICC-F2000 (Second Row) datasets.
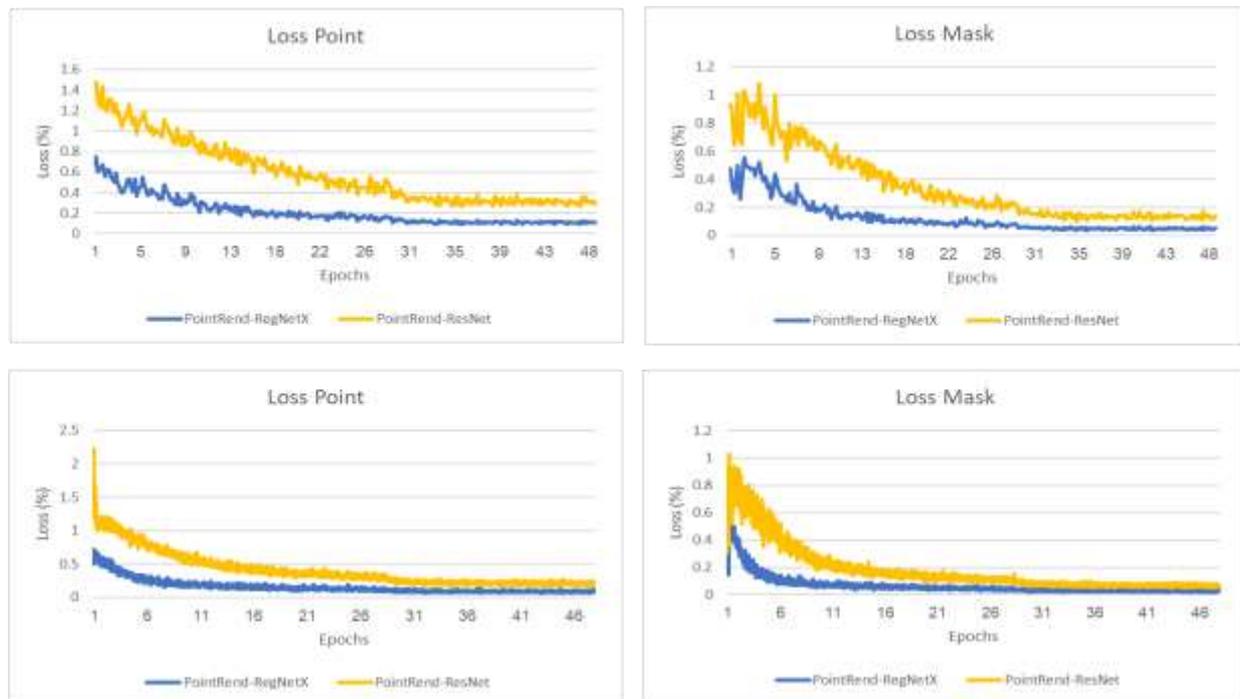
Figure 12. The obtained PointRend losses of two models: L_{point} loss in first column, L_{mask} loss in second column. MICC-F220 (First Row) and MICC-F2000 (Second Row) datasets.
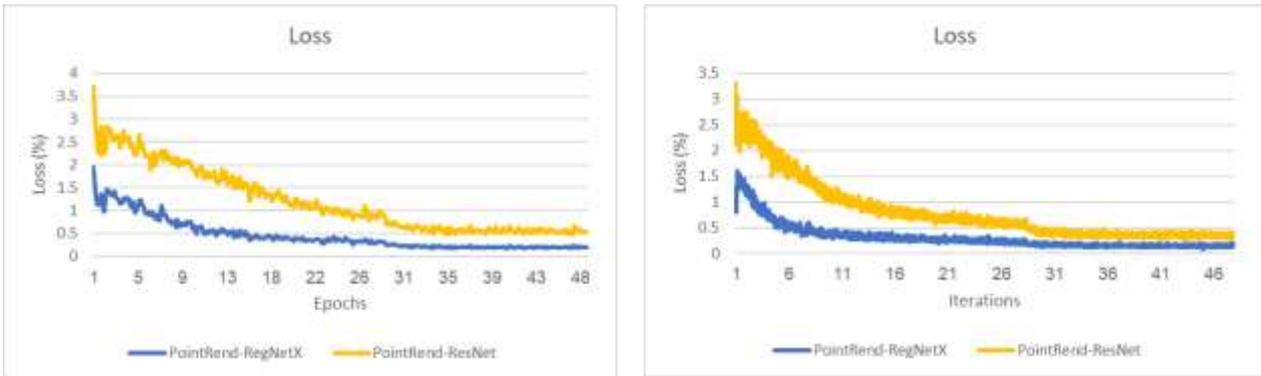
Figure 13. The overall losses obtained of two models. MICC-F220 (left) and MICC-F2000 (Right) datasets

A Comparison with Previous Studies

The obtained results showed that in most cases, the proposed PointRend's models outperformed the Mask R-CNN models [3]. Specifically, in the case of MICC F-220, PointRend-RegNetX achieved a higher mAP at all IoU levels compared to both MaskRCNN-ResNet-101 and MaskRCNN-MobileNet. For instance, at IoU 0.5, PointRend-RegNetX achieved a perfect score of 100% while MaskRCNN-MobileNet-V1 scored 93% and MaskRCNN-ResNet-101 scored 90%. Similarly, at IoU 0.75, PointRend-RegNetX achieved a score of 100% while MaskRCNN-MobileNet-V1 scored 83% and MaskRCNN-ResNet-101 scored 80% (see Fig 14).

In the case of MICC F-2000, the results are more mixed, but PointRend-RegNetX and PointRend-ResNet50 still perform comparably or better than the Mask R-CNN models. For instance, at IoU 0.5, PointRend-RegNetX achieved a score of 98.9% while MaskRCNN-MobileNet-V1 scored 74% and MaskRCNN-ResNet-101 scored 80%. At IoU 0.75, PointRend-RegNetX achieved a score of 95.5% while MaskRCNN-MobileNet V1 scored 60% and MaskRCNN-ResNet-101 scored 74% (see Fig 15).

Therefore, the PointRend models seem to be more effective than the Mask R-CNN models at instance segmentation tasks. PointRend's ability to refine object boundaries and accurately detect object edges may explain why it outperforms Mask R-CNN in these tasks. It is also worth noting that the choice of backbone network affects the performance of both PointRend and Mask R-CNN, with the RegNetX backbone performing better than the ResNet50 backbone in PointRend.
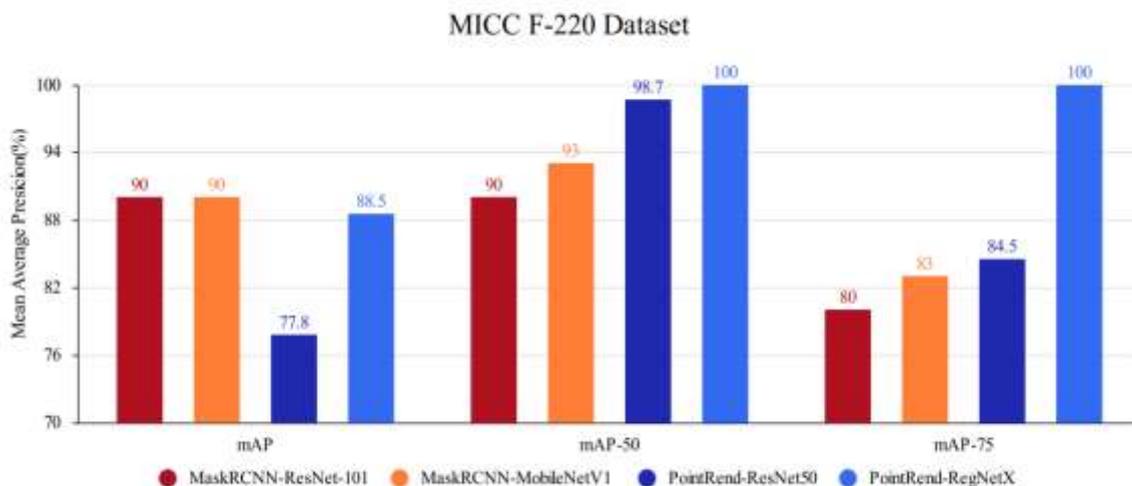
Figure 14. Comparison of mAP, mAP_{50}, and mAP_{75} for copy move using the proposed PointRend models with backbone Resnet50 and RegNetX against Mask RCNN models with backbone networks ResNet-101 and MobileNet V1 On MICC-F220 dataset
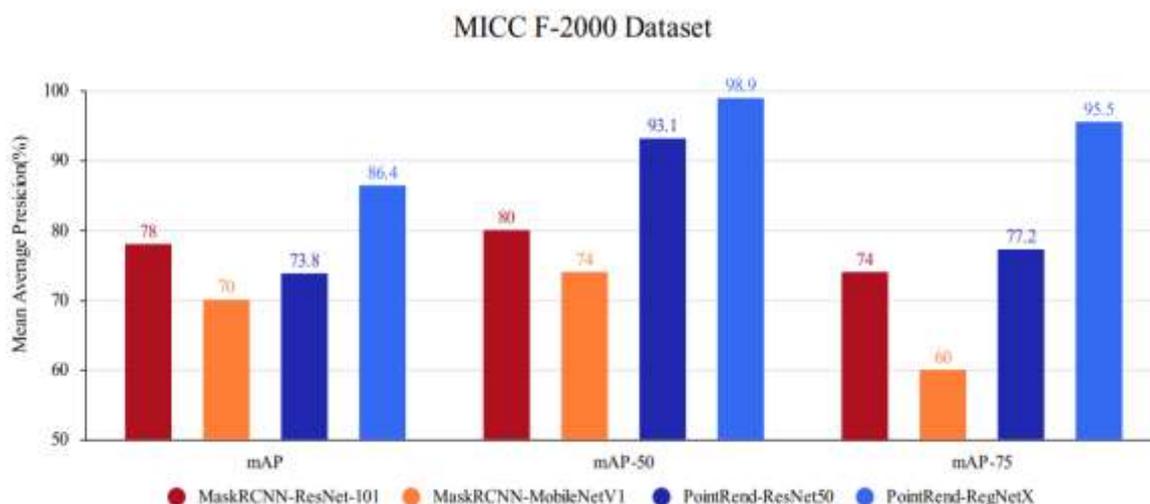
Figure 15. Comparison of mAP, mAP$_{50}$, and mAP$_{75}$ for copy move using the proposed PointRend models with backbone Resnet50 and RegNetX against Mask RCNN models with backbone networks ResNet-101 and MobileNet V1 On MICC-F2000 dataset

Conclusion

The paper presents an improved model for the task of localizing copy-move forgeries in images. The model is based on PointRend, which is a state-of-the-art instance segmentation model that can achieve high accuracy on complex scenes with a large number of objects. The improved model, called PointRend-RegNetX, replaces the original Resnet-50 backbone of PointRend with a more lightweight and efficient backbone (RegNetX). The two models, PointRend-Resnet and PointRend-RegNetX, are evaluated on two standard datasets, MICC F220 and MICC F2000. The results showed that PointRend-RegNetX outperforms PointRend-Resnet in most tests. In instances segmentation of forged regions, PointRend-RegNetX achieves a mean average precision (mAP) of 88.5% on MICC F220 and 86.4% on MICC F2000. In the detection of bounding box of the object, PointRend-RegNetX achieves a mAP of 84% on MICC F220 and 85.9% on MICC F2000. The experiments demonstrated that PointRend-

RegNetX provides a better percentage to localize a forged region in the image. Another interesting observation is that the proposed models perform better on the larger MICC F-2000 dataset than on the smaller MICC F-220 dataset. This suggests that the proposed models are better suited for more challenging and complex datasets. Moreover, the RegNetX model is more lightweight compared to ResNet model, and it has a compact output width and elastic growth ratio of the width of output between the successive stages. In future work, the modification on PointRend-RegNetX may be applied to achieve better results or investigated on other datasets that contain more challenging states. Additionally, this model may be applied to detect other types of image forgeries such as image splicing or image retouching. Overall, the improved model presented in this paper offers a promising solution for the problem of detecting copy-move forgeries in images.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been

included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee in University of Anbar.

Authors' Contribution Statement

M.H.F.'s contributions to the article paper include the following: Idea conception, framework design and implementation, database acquisition, and the writing of the manuscript. K.S. performed the following: Idea generation, analysis of the study, and revision. While S.J. accomplished the research analysis, revision, and proofreading.

References

1. Kasban H, Nassar S. An efficient approach for forgery detection in digital images using Hilbert–Huang transform. Appl Soft Comput. 2020; 97: 106728. https://doi.org/10.1016/j.asoc.2020.106728

2. Ahmed B, Gulliver TA, alZahir S. Image splicing detection using mask-RCNN. Signal Image Video Process. 2020; 14(5): 1035-42. https://doi.org/10.1007/s11760-020-01636-0

3. Kadam KD, Ahirrao S, Kotecha K. Efficient approach towards detection and identification of copy move and image splicing forgeries using mask R-CNN with MobileNet V1. Comput Intell Neurosci. 2022; 2022. https://doi.org/10.1155/2022/6845326

4. Abdalla Y, Iqbal MT, Shehata M. Convolutional neural network for copy-move forgery detection. Symmetry. 2019; 11(10): 1280. https://doi.org/10.3390/sym11101280

5. Harba ES, Harba HS, Abdulmunem IA. Advanced Intelligent Data Hiding Using Video Stego and Convolutional Neural Networks. Baghdad Sci J. 2021; 18(4):1317-1327. https://doi.org/10.21123/bsj.2021.18.4.1317

6. Khalaf M, Dhannoon BN. MSRD-Unet: Multiscale Residual Dilated U-Net for Medical Image Segmentation. Baghdad Sci J. 2022; 19(6 Supplement):1603-1611. https://doi.org/10.21123/bsj.2022.7559

7. Bondi L, Lameri S, Güera D, Bestagini P, Delp EJ, Tubaro S, editors. Tampering Detection and Localization Through Clustering of Camera-Based CNN Features. IEEE Conf Comp Vis Pattern Recognit Workshops. 2017. https://doi.org/10.1109/CVPRW.2017.232

8. Goel N, Kaur S, Bala R. Dual branch convolutional neural network for copy move forgery detection. IET Image Process. 2021: 656-65. https://doi.org/10.1049/ipr2.12051

9. Shi Z, Shen X, Kang H, Lv Y. Image manipulation detection and localization based on the dual-domain convolutional neural networks. IEEE Access. 2018: 76437-53. https://doi.org/10.1109/ACCESS.2018.2883588

10. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM. 2017; 60(6): 84-90. https://doi.org/10.1145/3065386

11. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 14091556. 2014. https://doi.org/10.48550/arXiv.1409.1556

12. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al., editors. Going deeper with convolutions. Proc IEEE conf comp vis pattern recognit. 2015. https://doi.org/10.1109/CVPR.2015.7298594

13. He K, Zhang X, Ren S, Sun J, editors. Deep residual learning for image recognition. Proc IEEE conf comp vis pattern recognit. 2016. https://doi.org/10.1109/CVPR.2016.90

14. Muzaffer G, Ulutas G, editors. A new deep learning-based method to detection of copy-move forgery in digital images. 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT); 2019: IEEE. https://doi.org/10.1109/EBBT.2019.8741657

15. Samir S, Emary E, El-Sayed K, Onsi H. Optimization of a pre-trained AlexNet model for detecting and localizing image forgeries. Information. 2020: 275. https://doi.org/10.3390/info11050275

16. Agarwal R, Verma OP. An efficient copy move forgery detection using deep learning feature extraction and matching algorithm. Multimed Tools Appl.. 2019: 1-22. https://doi.org/10.1007/s11042-019-08495-z

17. Wu Y, Abd-Almageed W, Natarajan P, editors. Image copy-move forgery detection via an end-to-end deep neural network. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV); 2018: IEEE. https://doi.org/10.1109/WACV.2018.00211

18. Wang X, Wang H, Niu S, Zhang J. Detection and localization of image forgeries using improved mask regional convolutional neural network. Math Biosci Eng. 2019; 16(5): 4581-93. https://doi.org/10.3934/mbe.2019229

19. Ahmed B, Gulliver TA, alZahir S. Image splicing detection using mask-RCNN. Signal Image Video Process. 2020; 14:1035-42. https://doi.org/10.1007/s11760-020-01636-0

20. Kirillov A, Wu Y, He K, Girshick R, editors. Pointrend: Image segmentation as rendering. Proceedings of the IEEE/CVF conf comp r vis pattern recognit. 2020. https://doi.org/10.1109/CVPR42600.2020.00982

21. Shan B, Fang Y. A cross entropy based deep neural network model for road extraction from satellite images. Entropy. 2020; 22(5): 535. https://doi.org/10.3390/e22050535

22. Duan E, Wang L, Wang H, Hao H, Li R. Feed weight estimation model for health monitoring of meat rabbits based on deep learning. Int J Agric Biol Eng. 2022; 15(1): 233-40. https://doi.org/10.25165/j.ijabe.20221501.6797

23. Dong Y, Zhang Y, Hou Y, Tong X, Wu Q, Zhou Z, et al. Damage Recognition of Road Auxiliary Facilities Based on Deep Convolution Network for Segmentation and Image Region Correction. Adv Civ Eng. 2022; 2022. https://doi.org/10.1155/2022/5995999

24. Li L, Zhang S, Wang B. Apple leaf disease identification with a small and imbalanced dataset based on lightweight convolutional networks. Sensors. 2021; 22(1): 173. https://doi.org/10.3390/s22010173

25. Zeng X, Wei S, Wei J, Zhou Z, Shi J, Zhang X, et al. CPISNet: delving into consistent proposals of instance segmentation network for high-resolution aerial images. Remote Sens. 2021; 13(14): 2788. https://doi.org/10.3390/rs13142788

26. Kadam K, Ahirrao S, Kotecha K, Sahu S. Detection and localization of multiple image splicing using MobileNet V1. IEEE Access. 2021; 9: 162499-519. https://doi.org/10.1109/ACCESS.2021.3130342

27. Peña Moliner D. Extending object classificatiion convolutional neural networks to custom logo detection: Universitat Politècnica de Catalunya; 2020. https://oa.upm.es/57088/

28. Jiao L, Zhang F, Liu F, Yang S, Li L, Feng Z, et al. A Survey of Deep Learning-Based Object Detection. IEEE Access. 2019; 7:128837-68. https://doi.org/10.1109/ACCESS.2019.2939201

29. Amerini I, Ballan L, Caldelli R, Bimbo AD, Serra G. A SIFT-Based Forensic Method for Copy–Move Attack Detection and Transformation Recovery. IEEE Transactions on Information Forensics and Security. 2011;6(3):1099-110. https://doi.org/10.1109/TIFS.2011.2129512

أسلوب كفوء لتحديد موقع تزوير Copy-Move باستخدام تقنية PointRend مع RegNetX

محمود هلال فرحان[1]، خالد شاكرجاسم[1]، سفيان الجنابي[2]

[1] قسم نظم المعلومات، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة الانبار، الرمادي، العراق.

[2] قسم انظمة الشبكات، كلية علوم الحاسوب وتكنولوجيا المعلومات، جامعة الانبار، الرمادي، العراق.

الخلاصة

الصور الرقمية هي أحد المصادر ذات الهيمنة العالية للمعلومات والاتصالات في العصر الحديث. ولكن يمكن بكل بساطة تغيير الصورة والتعديل عليها بسبب وفرة الادوات لتحرير الصور. يمكن أن تنتقل هذه الصور التي تم التعديل عليها عبر منصات وسائل التواصل الاجتماعي للتأثير على مجموعة من الأشخاص في المجتمع وقد تكون لها آثار إيجابية او سلبية. لتلك الأسباب أصبح تطوير تقنيات اكتشاف التزوير في الصور وتحديد موقعه مسألة ذات أهمية كبيرة. تزويرالـ Copy move (CMF) هو أحد أكثر عمليات التزوير شيوعًا. في هذا النوع CMF صورة التزوير الجديدة يتم إنشاءها عن طريق نسخ جزء معين من الصورة ووضعه في مكان اخر على نفس تلك الصورة. تقترح هذه الورقة البحثية تقنية PointRend كأسلوب لتحديد موقع الـ CMF. يقدم هذا العمل أيضًا باستخدام تقنية PointRend مع نموذج أساسي أقل حجما (RegNetX backbone) كنموذج مقترح (PointRend-RegNetX) لاكتشاف مثل هذه التزويرات. من التحليل المقارن للنموذج المقترح مع النموذج القياسي الذي يستخدم النموذج الأساسي ResNet-50 على مجموعتي بيانات قياسيتين، تبين أن النموذج المقترح (PointRend-RegNetX) قد تفوق على النموذج القياسي في كلا مجموعتي البيانات MICC F-220 و MICC F-2000 للصور التي تحتوي على تزوير الـ CMF. في حالات تحديد موقع المناطق المزورة، حقق النموذج المحسن (PointRend-RegNetX) متوسط معدل دقة (mAP) بنسبة 88.5 % على مجموعة البيانات MICC F-220 و 86.4 % على مجموعة البيانات MICC F-2000.

الكلمات المفتاحية: تزوير الصور، تزويرالـ Copy move (CMF) ، التعلم العميق، تقنية PointRend، النموذج الأساسي RegNetX.