IMAGE INPAINTING DETECTION USING CNN

A Major Project Report Submitted By

Zangam Teja (19116103)

Under the guidance of

Dr. Toshanlal Meenpal

(Associate Professor)

Department of Electronics and Communication Engineering

National Institute of Technology Raipur

G.E. Road Raipur-492010(C.G.).

NATIONAL INSTITUTE OF TECHNOLOGY, RAIPUR

CERTIFICATE

This is to certify that the Major Project work entitled “Image Inpainting Detection Using CNN” is the bonafide work done by Zangam Teja (19116103) under our guidance and supervision. This report is submitted following the completion of the Major project during the academic session of August -December 2022.

Head of the Department	Project Guide
Dr. A.S. Raghuvanshi Head of Department Department of ECE NIT Raipur	Dr. Toshanlal Meenpal Associate Professor Department of ECE NIT Raipur

NATIONAL INSTITUTE OF TECHNOLOGY, RAIPUR

CERTIFICATE BY THE EXAMINER

This is to certify that the Major Project work entitled “Image Inpainting Detection Using CNN” is the bonafide work done by Zangam Teja (19116103) under our guidance and supervision. This report is submitted following the completion of the Major project during the academic session of August-December 2022.

Examiner 1

Name:
Date:

Examiner 2

Name:
Date:

NATIONAL INSTITUTE OF TECHNOLOGY RAIPUR

Declaration

I certify that

a) The work comprised in the thesis is original and is done by myself under the supervision ofmy supervisor.

b) The work has not been submitted to any other institute for any degree or diploma.

c) I have followed the guidelines provided by the Institute in writing the thesis.

d) Whenever I have used materials (data, theoretical analysis, and text) from other sources, I have given due credit to them in the text of the thesis and given their details in the references.

e) Whenever I have quoted written materials from other sources, I have put them under quotationmarks and given due credit to the sources by citing them and giving the required details in the references.

(19116103)
Zangam Teja

Acknowledgment

Every project, big or small, succeeds in large part because of the efforts of a group of outstanding people who have always offered beneficial advice or given a hand. I am grateful for all of the people who have contributed to the success of our project by providing inspiration, support, and guidance. We are highly indebted to our guide Dr. Toshanlal Meenpal for his invaluable help and guidance during the course of the project. We acknowledge his constant encouragement by giving his critics of our work. We are grateful to him for having given us support and confidence. Our special thanks goto Dr. A.S. Raghuvanshi, Head of the Department of Electronics and Communication Engineering, NIT Raipur, for providing us with the best facilities in Department. We'd want to express our gratitude to all of our teachers for providing a firm foundation for our studies and subsequent research. They have been tremendous sources of inspiration for us, and we sincerely appreciate them. We'd also like to express our gratitude to our friends, particularly my classmates, for all of the smart and thought-provoking talks that inspired us to think beyond the obvious. Last but not least, I'd like to express my gratitude to my parents for teaching in me the value of hard work.

Zangam Teja (19116103)

Table of Contents

List of Figures and Tables

Abstract

In multimedia forensics, image forgery detection has become a challenging issue. The removal of objects from photographs is one of the most popular counterfeit techniques. Using patch-based image inpainting, object elimination renders the image visually appealing and physically plausible while leaving no discernible traces. Researchers have always had trouble locating the Inpainted region because of this. In order to identify patch-based inpainting procedures, this research suggests a hybrid CNN. We achieved an outstanding result in detecting the Inpainted region. To train on and validate our findings, we built an Inpainted picture database and we evaluated our performance using the IoU metric.

1. Introduction

Digital images are now widely employed in many areas of daily life, including news, advertising, publishing, etc., thanks to the quick development of computer technology and the Internet. The ability to access, duplicate, and modify images more effectively than at any other moment in recent memory is made possible by the availability of potent image-processing software. In the era of computerized interactive media, the unwavering quality (authenticity), as well as the legitimacy of visual data, have been major concerns. Digital picture forensics has received tremendous attention and has recently seen a range of applications because it is such a promising method. [1]. JPEG compression, histogram balance, median clarifying, copy-move forgery, and other criminology methods are being developed and have required remarkable work from experts. Picture inpainting plans to use already-existing image data to successfully replace image areas that have been destroyed or removed in a way that is physically feasible. It's a cutting-edge, effective image-altering technology. It represents a substantial achievement in the rapidly developing field of computer vision and image processing. There are many uses for image inpainting, including image restoration (like removing text or scratches), photo editing (like removing objects), virtual restoration of digital artwork (like removing breaks), image coding and transmission recovery, and more[8].

Fig 1.1: An instance of picture inpainting is used to remove objects, exhibiting the actual image (left) and the inpainted image (right).

It's a cutting-edge, effective image-altering technology. It represents a substantial achievement in the rapidly developing field of computer vision and image processing. True classification and feature extraction are often the two components of the classification issue. The inpainting activity hardly gets the features with high segregation since it leaves no indisputable trail. The main method of contrast used in standard inpainting forensics is the similarity features between image patches. patchwork made of unpainted areas overpainted [1]. They both have certain disadvantages in common, including high computing costs for the extraction of features and significant false positive rates in uniform image areas. In this study, we propose to recognize the inpainting modification and differentiate the inpainting position by using a convolutional neural network.

2. Literature Survey

It has been proposed that image inpainting be used to enhance image quality or repair damaged areas. [2]. When an image changes, undesired elements can be eliminated via image inpainting. Built on the region-filling method, image inpainting has been split into two groups. The first is dispersion-based inpainting, in which neighborhood structure is flawlessly proliferated from the outside to the inside of the openings [3]. This method is suitable for filling small gaps while maintaining the current state of design in the area. If bigger areas need to be recovered, this tactic does not work effectively. The second method is patch-based inpainting, which involves duplicating patches from the known piece of the image and sewing them together to fill the gaps. [6] In that, we execute area duplication, this technique—known as the block duplication method—is comparable to a copy-move. When compared to district duplication, block duplication appears to be more ambiguous.

Inpainting forensics has become a testing ground for specialists because the improved inpainting procedure leaves no perceptual artifacts. Wu et al. [8] proposed prior blind inpainting location, which relied on zero-network naming and fluffy enrollment. Zero-network naming was utilized in dubious locale to give a matching level of the blocks and fluffy part transport had been registered to recognize the altered areas. The primary downside of this strategy was a manual choice of dubious areas. Chang and associates suggested a two-stage scanning computation for accelerating the search for suspicious areas to get over the abovementioned drawback. [9] multi-area relations had sifted the false alarm for this situation. Discovery precision was still constrained by comparable patches. Utilizing focal pixel planning for fix search and patch detection helped enhance this work by removing deceiving issue patches. [10 -11]

Convolutional neural networks (CNN) have made significant progress in the fields of image processing and computer vision. In the world of forensics, experts have started to adopt CNN for better execution. Zhu and his colleagues. [12] presented fascinating CNN-based forensics painting concept. A CNN network based on encoder-decoders was used, and a name lattice was used to improve CNN preparation. For ascertaining the successful misfortune, weighted cross entropy had been utilized. Exploratory outcomes had accomplished better execution. Driven by these works we have meant to plan a powerful identification calculation for Patch-based inpainting. Some of the main works of literature are tabulated as follows:

Table 2.1: Some recent literatures

YEAR	AUTHOR	TITLE, JOURNAL NAME	DESCRIPTION
2018	X. Zhu, et. al.	A deep learning approach to patch-based image inpainting forensics. Signal Processing: Image Communication.[1]	Used CNN based encoder decoder model for inpainting investigation. Reduces computation cost.
2021	Wang, X., Niu, S. and Wang, H	A survey on image tampering and its detection in real-world photos.[2]	The most prevalent types of picture tampering, publicly accessible image tampering datasets, and cutting-edge tampering detection methods are covered in this study.
2017	Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H.,	Mobilenets: Efficient convolutional neural networks for mobile vision applications.[5]	The mobile net is the topic of this essay, which illustrates how mobile nets work well for a variety of uses.
2015	Ronneberger, O., Fischer, P. and Brox, T.,	U-net: Convolutional networks for biomedical image segmentation.[4]	In order to utilise the given annotated examples more effectively, the U-net and training technique heavily rely on data augmentation.

3. Related works

3.1 Inpainting:

An artwork's missing or damaged portions are filled in during the conservation process of inpainting in order to create a full image. In image restoration, this cycle is frequently applied. It can be used with both analog and digital art forms, including paintings in oil or acrylic, prints from photographs, models, and digital photos and videos. Its foundations are in actual artwork, such as paintings and models.

Inpainting is credited with being invented by Pietro Edwards (1744–1821), Director of the Restoration of the Public Pictures in Venice, Italy. Edwards focused his restoration efforts on the objectives of the artist by employing a standardized method.
Inpainting must pass a few ethical conditions before it can be deemed real. The ethical acceptability of the total amount of inpainting done and the sort of inpainting chosen is influenced by a variety of thoughts and decisions. The ethical demands of inpainting are centered on authenticity, reversibility, and verification, similar to the majority of assurance prescriptions.
Such compensation should be reversible and should not fraudulently affect the known aesthetic, intellectual, and physical aspects of the cultural property, notably by eliminating or hiding original material, said the International Council on the Preservation of Cultural Property. [7]
In an era of museum tourism, new technologies, and a need for flawless photographs devoid of flaws, conservators' ethical procedures to preserve the integrity of originals continue to face challenges. However, inpainting could also be used maliciously to alter image content, which causes a crisis of faith in image content. [1]
There are two primary groups of inpainting detection methods now in use:

- The diffusion-based
- The patch-based. [1]

In this project patch based Inpainted images are used for the detection of inpainting.

3.2 Patch-based inpainting

Image inpainting is a method used to cosmetically restore areas of an image that have been effectively damaged or removed [1]. The dispersion-based and patch-based approaches can be used to categorize existing methods into two basic classes. By addressing imperfect differential conditions or equivalent dispersion frameworks, the prior diffuse image data from well-known locations into obscure regions. They have excellent workmanship when painting lengthy, thin parts, but struggle to recreate the surface and restore vastly missing areas. The patch-based methods internally disperse the picture patches from the known positions into the distant area, patch by patch. When painting large, hidden areas, they typically yield better inpainting results. Consequently, the subject of our consideration is patch-based picture inpainting forensics.

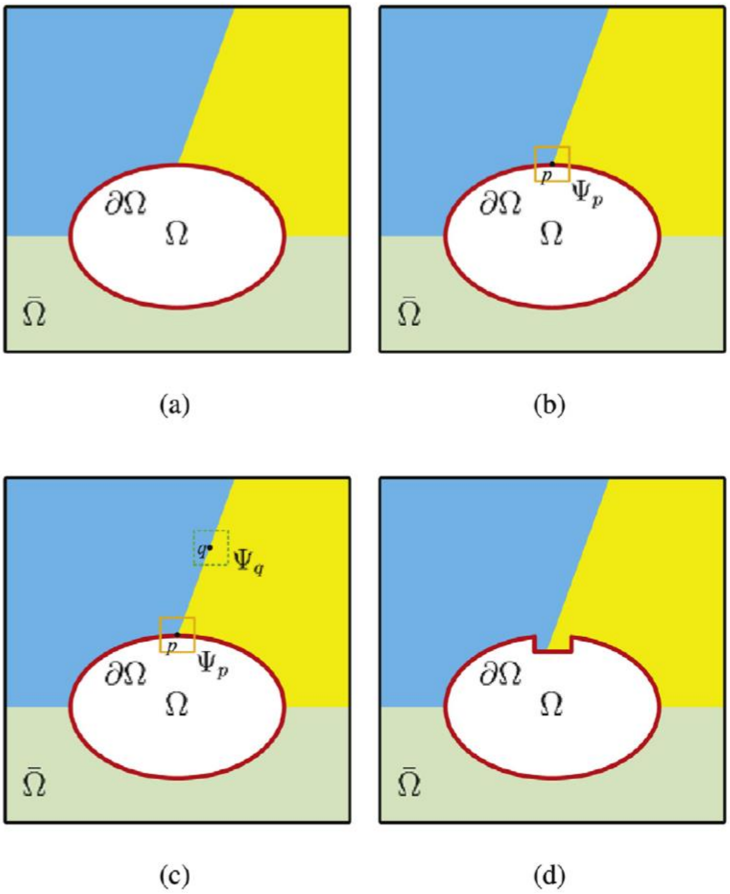


Fig 3.1 (a): missing region Ω of original image b) Selection of a patch $\Psi_P$with the major precedence c) Probing the matching patch Ψ_P for Ψ_q in the recognized area Ω
d) inpainting the area and updating the interior

The critical cycle for patch-based inpainting, using Criminisi's approach as an example, is shown as follows. The inpainting system is briefly understood as surveys given the image with the unidentified area shown in Fig. 3.1.

Stage 1: calculating the need for each internal point $\partial\Omega$ of the unidentified area Ω,
We make a decision regarding the circumstance at the location where the greatest need is p. The patched image Ψ_p focused at p is chosen to be used as the continuing patch that will be Inpainted, as shown in Fig 3.1 (b).

Stage 2: We search the entire known region, as suggested by a specific patch-matching principle, what's more, figure out the most comparable patch Ψq of Ψp, as displayed in Fig. 3.1(c).

Stage 3: The fatal flaw in Ψp will be occupied by utilizing a comparing picture element in Ψq, and the inside $\partial\Omega$, the acknowledged re n $\bar{\Omega}$ furthermore, the unidentified area Ω is undeniably refreshed in Fig 3.1.

Stage 4: It is possible to completely Inpainting the unidentified area and produces the Inpainted image by performing Stages 1 through 3 iteratively. The area needs calculation and the plan for the area matching standard continues to take into account the structure and surface data. This realizes the presentation benefit of Criminisi's method. The fundamental idea of Criminisi's method served as the foundation for many later inpainting techniques.

Fig 3.2: Inpainting object removal forgery example (here the player is removed and inpainting operation is performed). Original image on the left side and the Inpainted image on the right side

4. Proposed Methodology

In this study, we use convolutional neural network-based techniques to find the Inpainting region of patch-based. The encoder-decoder network topology used in the construction of the CNN enables us to forecast the likelihood of inpainting for each pixel in a picture. In this, project a hybrid model is built for the detection of Inpainted images. Here Pre-trained Mobile net is used as an encoder and a part of the U-net is used as a decoder. The image dataset is taken in the following way to train the model.

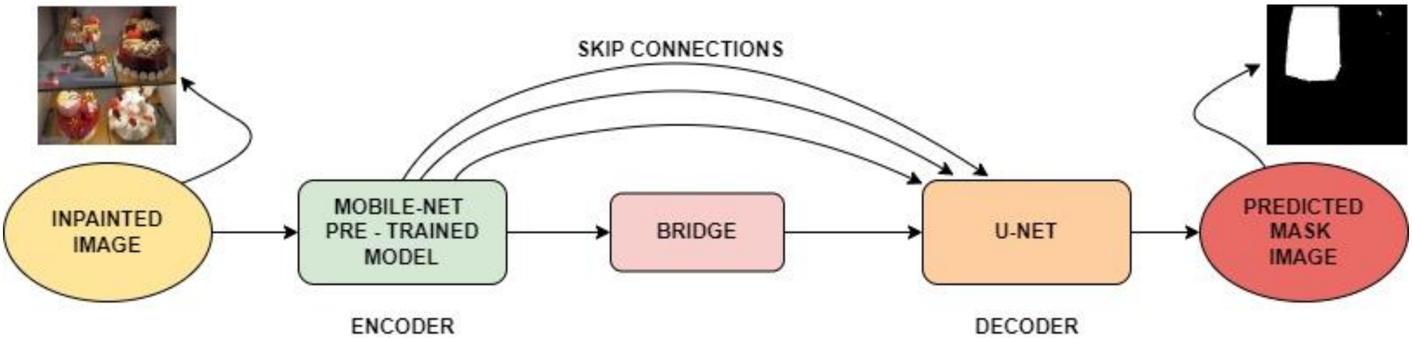


Fig 4.1: Block Diagram of Model Architecture

4.1 Model Architecture

In this project, a hybrid model is implemented, it contains U-net and mobile net model architecture. A Pretrained Mobile net is used as an encoder part. A part of the U-net is used as a decoder part. And there are some skip connections connecting in between encoder and decoder. Basically, this skip connection are used for the feature maps to concatenate from the layers of an encoder to the layers of a decoder. These skip connections helps to resolve the problem of gradient vanishing and more context is given to the later layers of the Convolution neural network.

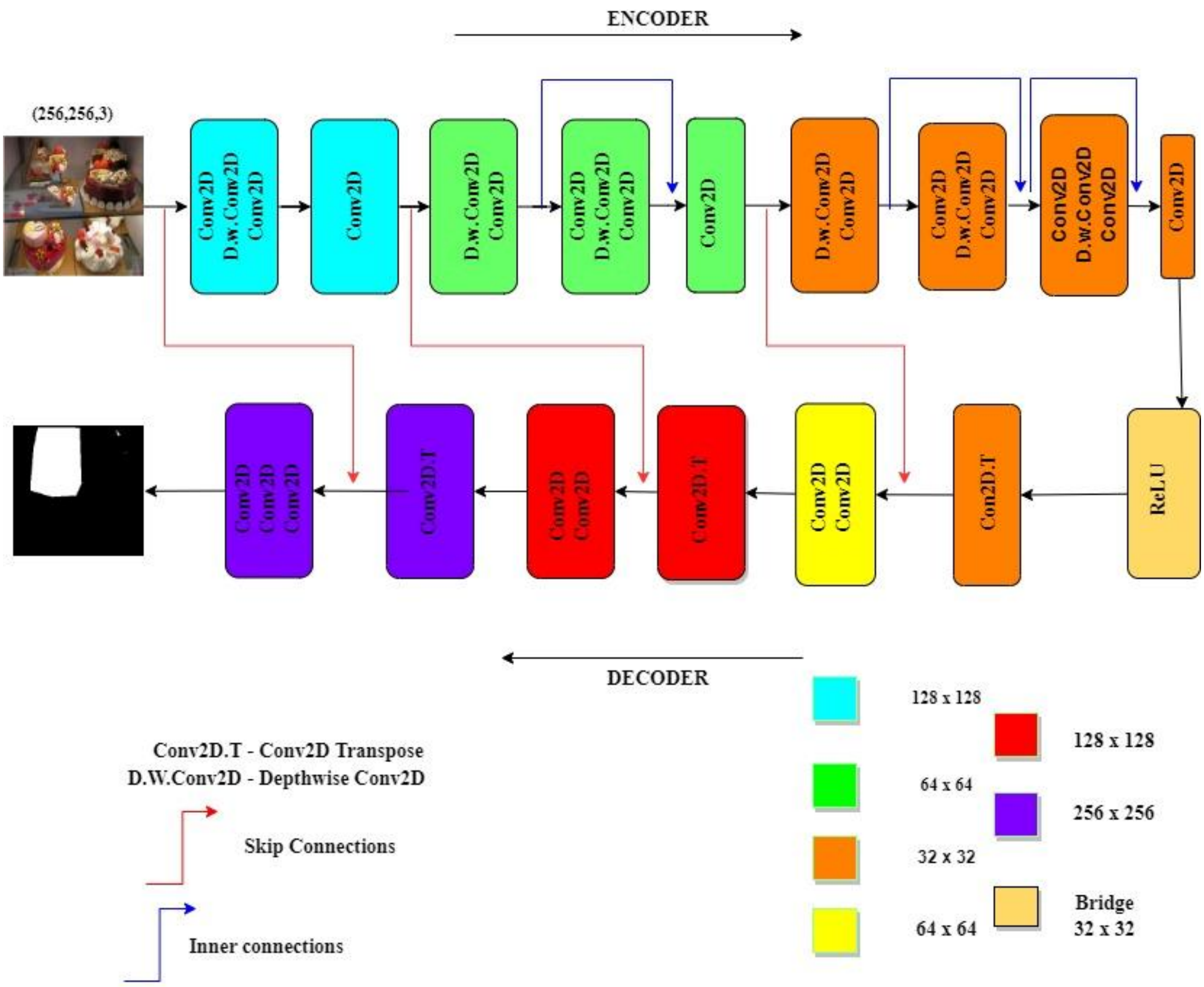


Fig 4.2: Proposed Model Architecture

4.1.1 <u>Encoder</u>

As an encoder part, we are using a mobile net pre-trained model. Neural networks have revolutionized several facets of artificial intelligence, enabling tasks like picture recognition to be tested with near-perfect accuracy. In any case, the drive to further develop precision often includes some major disadvantages: Modern networks demand significant computing resources that are beyond the capabilities of many embedded and mobile applications. A convolutional neural network architecture called MobileNetV2 was created for portable electronics. It is based on a bottleneck layer structure with an inverted residual structure. [5] The intermediate expansion layer filters feature using quick depth-wise convolutions as a source of non-linearity. The MobileNetV2 design consists of a 32-filter initial fully convolution layer, followed by 19 further bottleneck layers.

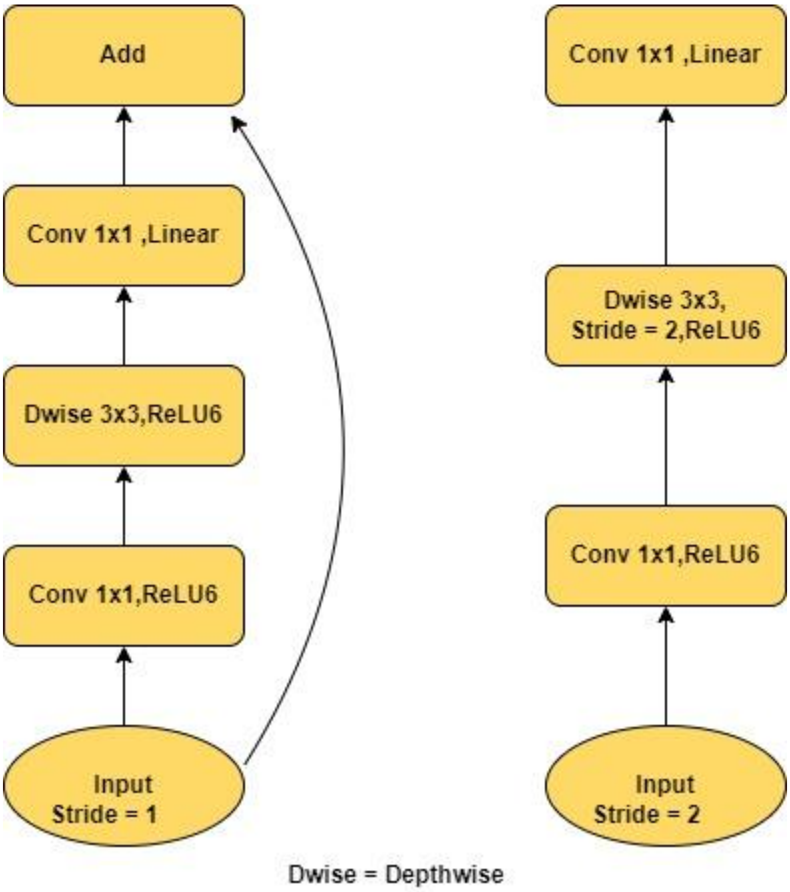


Fig 4.3: MobileNetV2 Convolutional Blocks

In MobileNetV2 [5], the blocks are of two kinds. One is a residual block of stride 1. One more is a block of stride 2 for downscaling. There are 3 layers for the blocks of two kinds. This time, the foremost layer is 1×1 convolution (with ReLU6). [5] The following layer depth-wise convolution. The final layer is another convolution of 1×1 however with no non-linearity. It is guaranteed that on the off chance that ReLU is utilized once more, the profound networks just have the force of a straight classifier on the non-zero volume part of the resulting space. As we are using a mobile net v2 pre-trained model as an encoder. which is trained with the ImageNet dataset. ImageNet is a huge dataset of marked photographs envisioned for computer vision research. In light of statistics about the dataset recorded on the ImageNet landing page, there are somewhat more than 14 million pictures in the dataset, somewhat more than 21,000 gatherings or classes, and somewhat more than 1 million pictures that have bounding boxes explanations.

Advantages of using Mobile Net as an Encoder:

- When compared to a non-pre-trained model, using a pre-trained encoder speeds up convergence.
- In comparison to a model that has not been trained beforehand, a pre-trained encoder aids the model in achieving high performance.

4.1.2 <u>Decoder</u>

In a variety of image identification tasks, deep convolutional networks have lately surpassed the state of the art. While convolutional networks have existed for quite some time, their success has been limited by the size of the available preparation sets and the size of the thought-about networks. Krizhevsky et al. made a significant breakthrough by directing the preparation of an enormous network with 8 layers and a large number of boundaries on the ImageNet dataset with 1 million preparation pictures. From that point forward, much larger and deeper networks were built.
U-Net [6], advanced from the conventional convolutional neural network, was first planned and applied in 2015 to handle biomedical pictures. As a general convolutional neural network zero in its undertaking on picture characterization, where information is a picture and result is one mark, however in biomedical cases, it requires us not exclusively to recognize whether there is an illness, yet in addition to confining the area of irregularity.

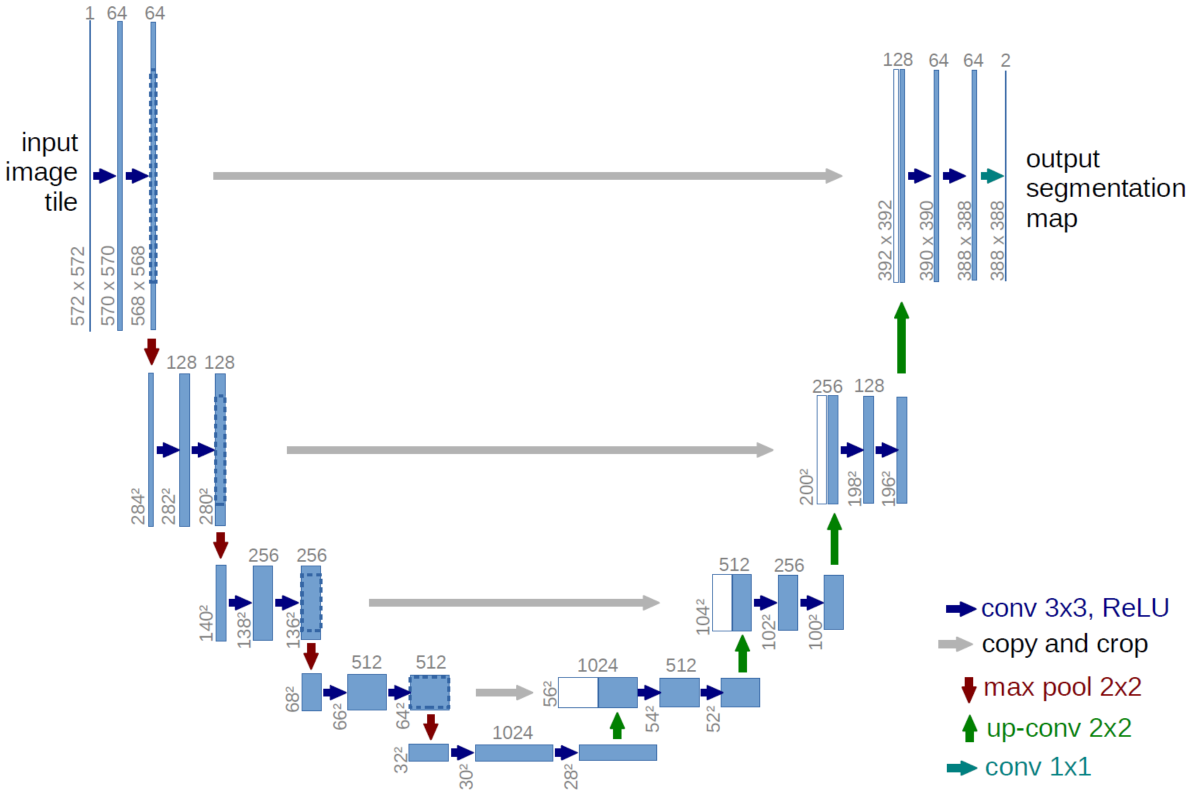


Fig 4.4: U-Net Architecture

U-Net architecture is built and modified to yield better segmentation with less training data. It is built using the fully convolutional network (FCN), which means that only convolutional layers are used and no dense layers are used at all.

Advantages of using U-net as a decoder:

- To acquire general information combining localization and context, which is necessary to forecast a successful segmentation map, the U-Net combines the location information from the down-sampling path.
- No Dense layer is used.

4.1.3 Loss Function

Loss functions are numerical equations that compute how far the forecasts veer off from the real qualities. Higher loss values recommend that the model is making a significant error, though lower loss values suggest that the forecasts are somewhat accurate. The objective is to decrease the loss function as much as could reasonably be expected. The loss function is utilized by models to get familiar with the trainable parameters, like weights and biases. In this project, we have used the focal loss function as loss function.

Focal loss focuses on the examples that the model gets wrong rather than the ones that it can confidently predict, ensuring that predictions on hard examples improve over time rather than becoming overly confident with easy ones.

Focal loss achieves this through something called Down Weighting. Down weighting is a technique that reduces the influence of easy examples on the loss function, resulting in more attention being paid to hard examples. This technique can be implemented by modifying the Cross-Entropy loss by a factor.[12]

$$Focal\ Loss = -\sum_{i=1}^{i=n}(i - p_i)^{\gamma} log_b(p_i)$$

Where γ (Gamma) is the focusing parameter to be tuned using cross-validation.

5. **Experiments and Results**

5.1 **Training :**

We have used a exampler – based dataset [8] to train the proposed model which has 1900 colored Inpainted images of 256x256 . The Inpainted region in the database ranges in size and shape 2% - 40% size of the image. Out of a total of 1900 photos, we randomly selected 1500 for training and 500 for testing. We scaled our database images to the suggested model's input size of 224x224 before utilizing them for training. In order to identify the used tampered region for each inpainted image, we constructed a ground truth label matrix. As the inpainted region in the training samples is relatively small in comparison to the uninpainted region, a binary focal loss with gamma value 2 was used as a loss function. The CNN parameters were iteratively updated using sgd. Parameters of Learning rate = 0.001, Momentum = 0.9, decay = 0.0016The model had been trained with 100 epochs by the above parameters and the plots of accuracy and loss of shown in fig 5.2 and fig 5.3

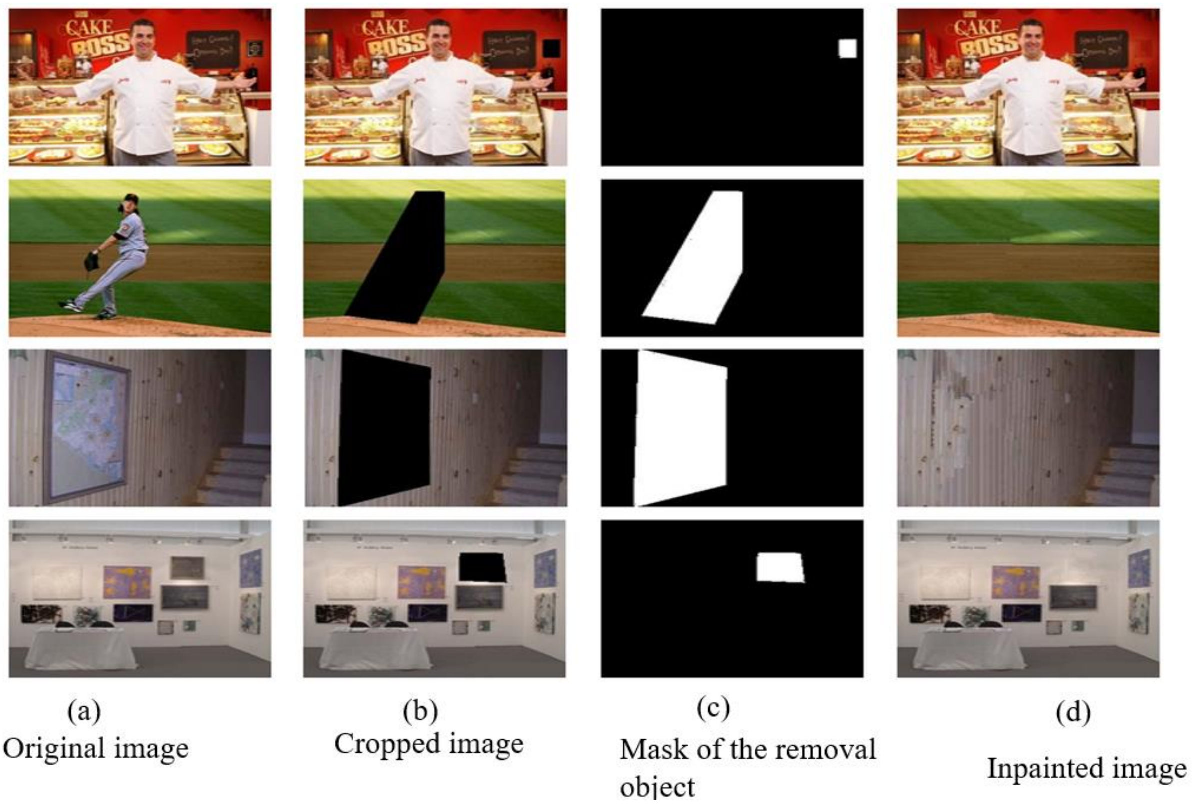


Fig 5.1: Inpainted Image Dataset: (a) original image, (b) cropped image, (c) removed object mask, (d) Inpainted image

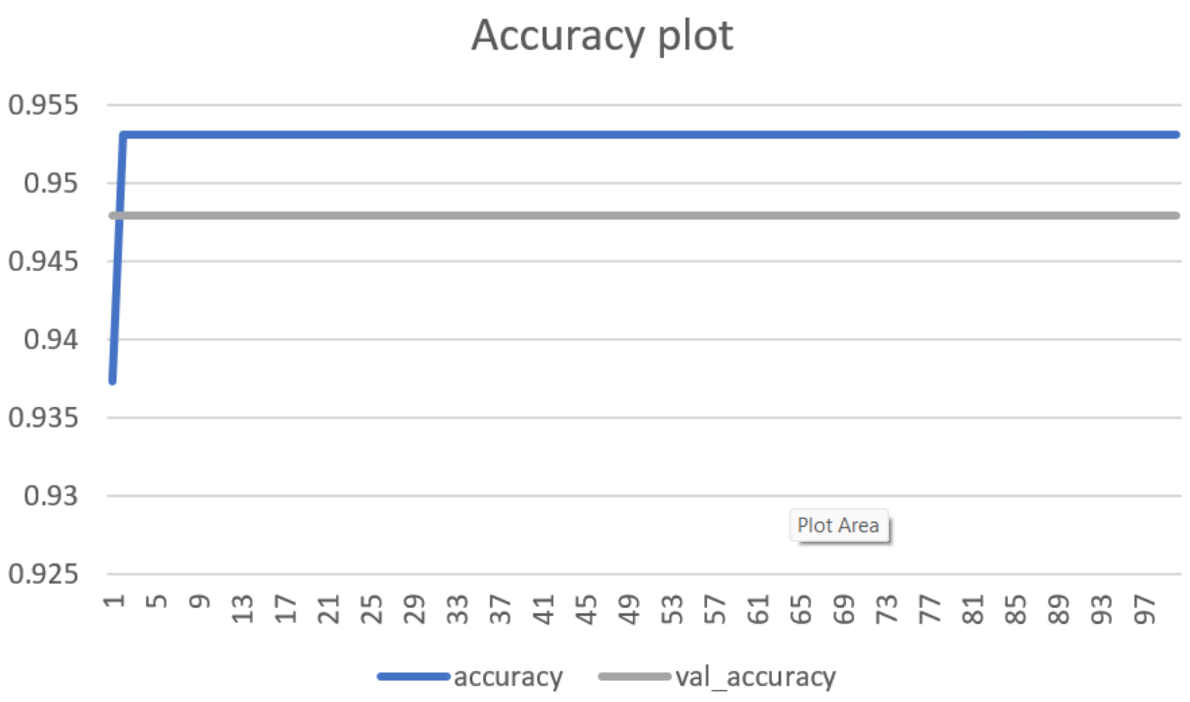


Fig 5.2: Plot between Accuracy and validation accuracy of the trained model

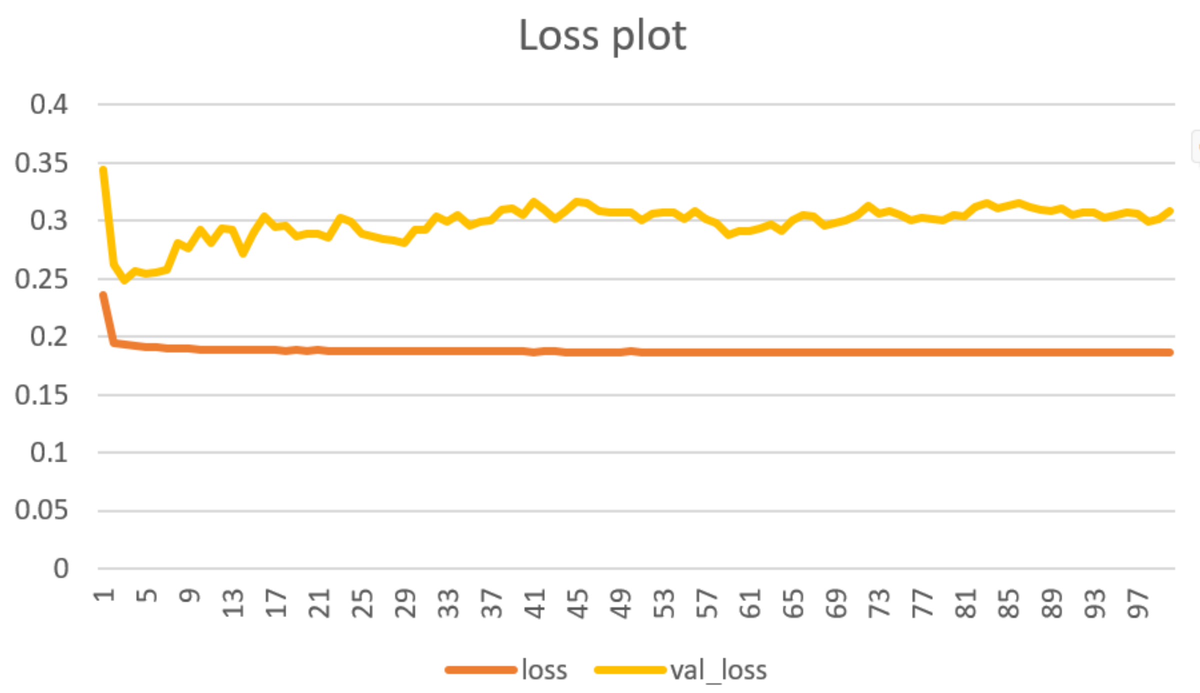
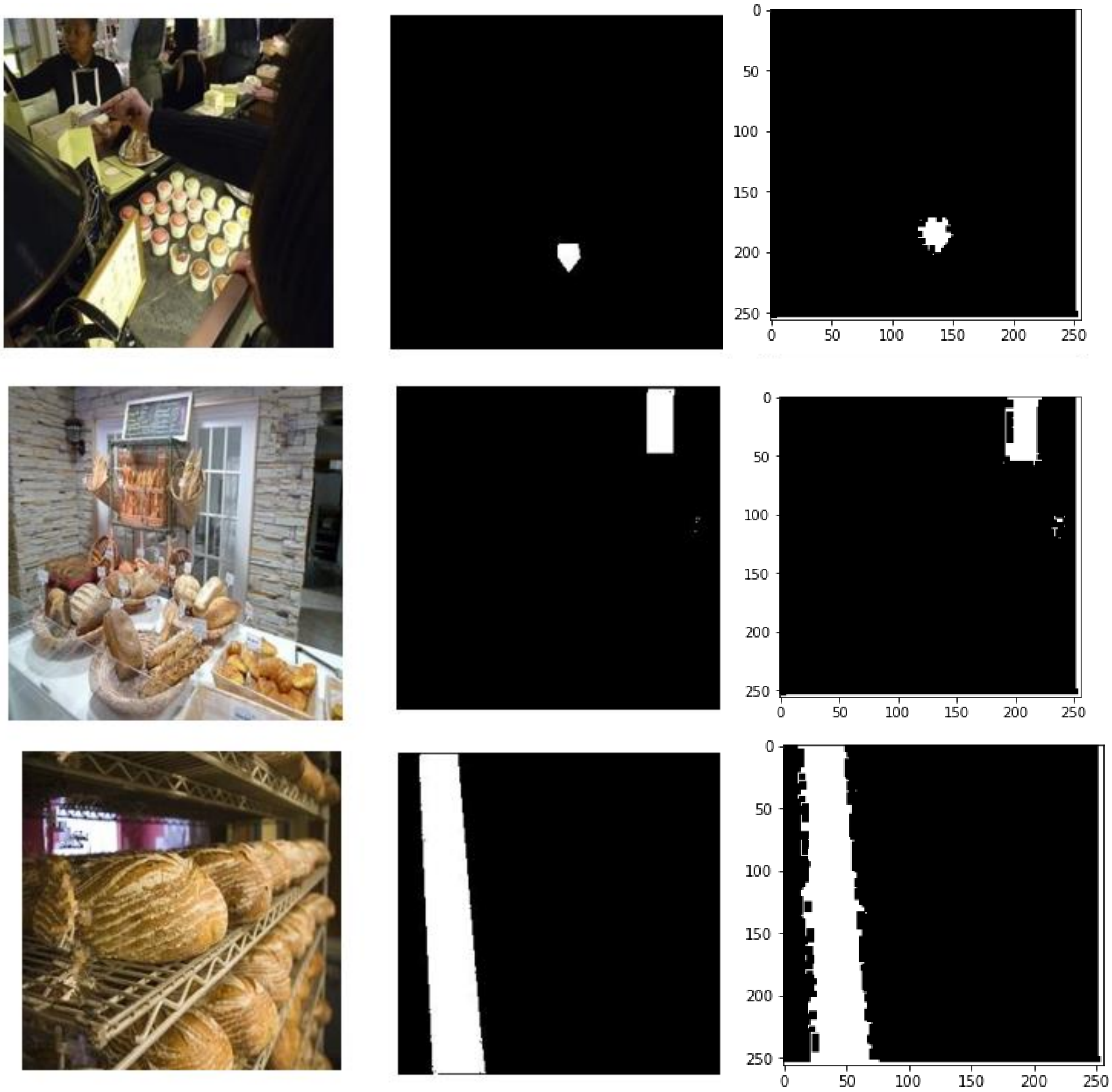


Fig 5.3: Plot between Loss and validation loss of the trained model

5.2 <u>Testing and Results</u>

To check the validity of our suggested architecture, we randomly selected 500 colour photos from the database. These pictures weren't taken into consideration for training. We examined our findings on 400 photos with various levels of tampering and found that, in terms of localization, we outperformed the previous detection approach in terms of irregular shape detection results. The first section in Fig depicts the Inpainted images, the second section the reference mask, and the third section the detected Inpainted region, demonstrating the detection performance of our suggested model. Despite the size of the inpainted region, we were able to get improved detection results.



(a) (b) (C)

Fig 5.4: a) represents the Inpainted images b) represents the reference mask c) represents the detected Inpainted region

Intersection over union (IoU): The percentage of overlap between the target mask and the projected output is calculated using the intersection over union (IoU) measure. It is calculated by intersecting a large number of pixels from the target mask and the predicted mask with a large number of pixels from each mask. The test result has a meanIoU plot, as seen in fig.5.5 below. It shows the average of all IoU scores of the tested result.

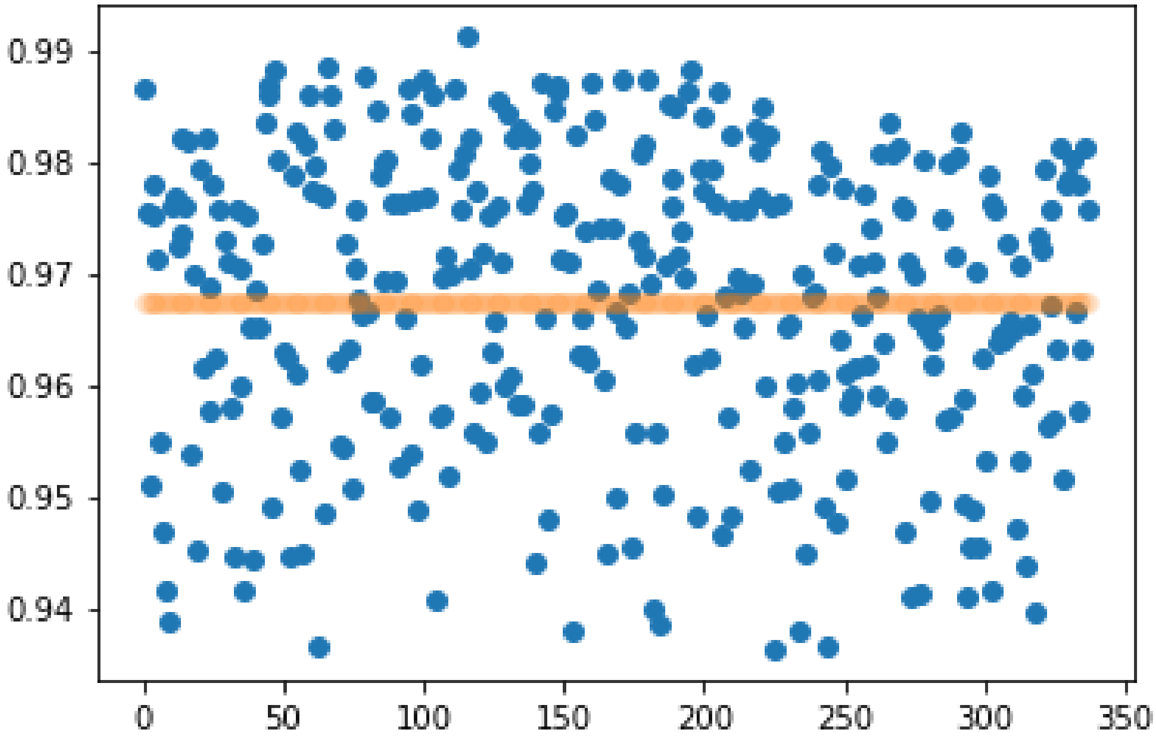


Fig 5.5: Plot of Mean IoU of the tested results

6. Conclusion

Image inpainting is one method used to forge images. This study introduced a brand-new technique for identifying picture inpainting. We proposed a hybrid CNN model for detecting targets in Inpainted regions. To undertake training and validation on the suggested model, we built an Inpainted picture library and evaluated it using an Inpainted region with an uneven shape. By using a common statistic like 96% of the mean average intersection over the union, we have attained remarkably high detection accuracy. Future training can make this model resilient to any inpainting process as well as other modification methods.

7. References

1. Zhu, X., Qian, Y., Zhao, X., Sun, B. and Sun, Y., 2018. A deep learning approach to patch-based image inpainting forensics. *Signal Processing: Image Communication*, *67*, pp.90-99.
2. Zheng, L., Zhang, Y. and Thing, V.L., 2019. A survey on image tampering and its detection in real-world photos. *Journal of Visual Communication and Image Representation*, *58*, pp.380-399.
3. Wang, X., Niu, S. and Wang, H., 2021. Image inpainting detection based on a multi-task deep learning network. *IETE Technical Review*, *38*(1), pp.149-157.
4. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S. and Liu, J., 2019. Ce-net: Context encoder network for 2d medical image segmentation. *IEEE transactions on medical imaging*, *38*(10), pp.2281-2292.
5. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.
6. Ronneberger, O., Fischer, P. and Brox, T., 2015, October. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
7. Idelson, A.I. and Severini, L., 2018. Inpainting. The Encyclopedia of Archaeological Sciences, pp.1-4.
8. Kumar, N. and Meenpal, T., 2021. Semantic Segmentation-Based Image Inpainting Detection. In Innovations in Electrical and Electronic Engineering (pp. 665-677). Springer, Singapore.
9. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T. and Efros, A.A., 2016. Context encoders: Feature learning by inpainting. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2536-2544).
10. Alharbi, A., Alhakami, W., Bourouis, S., Najar, F. and Bouguila, N., 2020. Inpainting forgery detection using hybrid generative/discriminative approach based on bounded generalized Gaussian mixture model. Applied Computing and Informatics.
11. Wang, X., Niu, S. and Wang, H., 2021. Image inpainting detection based on a multi-task deep learning network. IETE Technical Review, 38(1), pp.149-157.
12. Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P., 2017. Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).