END-TO-END FRAMEWORK FOR OBJECT REMOVAL AND IMAGE RESTORATION USE IN CNN

Nandana K S, MCA Student, PES Institute Of Technology & Management, Shivamogga, Karnataka, India
Dr. Sanjay K S, Associate Professor & Head, Dept. of MCA, PESITM, Shivamogga

ABSTRACT

Erase of unwanted objects in pictures and the consideration of the background in the visual reasonable way is a simple and conversely challenging challenge in the most current image editing programs and computer vision. This project trains Convolution Neural Networks (CNNs) to provide an end-to-end solution to autonomous and object removal/restoration in images. The system used this technique to remove the items that were either automatically or manually detected and segmented, the areas left would then be in painted using deep learning based techniques. It all comes down to three main areas. which are incorporated into the system: picture in painting, objects segmentation, and object detection. Besides the division of the desired objects, CNNs are also used to generate the realistic textures to fill in the blank spaces where the objects got cut. The reason is that since inputs via man are not necessary anymore, it is a faster, more effective, and scalable procedure. As opposed to the conventional. Since there is no need to manually enter data anymore, it takes less time, is more efficient, and can be more scalable. Comparing to the traditional methods that primarily depend on user-based masking, with patterns of artifacts, the proposed model is able to produce restoration processes that are natural and continuous, as it learns context of space and texture patterns.

KEYWORDS--object removal, image inpainting. It involves object detection.Additional relevant terms include semantic segmentation, background reconstruction, visual object erasing, and automated image editing within the field of computer vision.

I. INTRODUCTION

Image editing has also been a major aspect of digital art, photography, social media, and most computer vision applications over the past years. One of the common issues is the presence of the unnecessary objects in the photos such as signage, clutter, or viewers, which may disrupt the major part of the photo or a reduction in the quality of the image. The extraction of objects used to be a tedious editing process with manual selection that takes time and always display irregularities. This study applies the Convolution Neural Networks (CNNs) to give an end-to-end model of object remedial and photo restoration to address this challenge. The model will automatically detect and isolate unwanted contents in an image, remove them and visual coherently and naturally reinstate the rest of the picture. CNNs are ideal options since they can learn textures and spatial patterns by the large amounts of data available to them. The system attends to the image in painting, semantic segmentation and the object detection in the same pipeline. fast, efficient, and fully automated tool capable of removing objects, which can be used in many areas, such as in professional picture processing, creating visual material that might be based on AI, and the editing personal photos.

II. LITERATURESURVEY

The objective masking and image recovery has a significant impact on area in which the advent of deep learning has contributed many advancements, especially to Convolution Neural Networks (CNNs). Most of the classical approaches to object removal were based on manual editing or use examples, but this approach was too time consuming was not accurate in very complex situations. Automatic in painting of CNNs was shifted to when deep learning-based approaches began to appear, with one proposed by Pathak et al. (2016) referred to as the Context Encoder. Bryce E Bayer. Array color imaging, 1976. The US Patent [1]

It employed an encoder-decoder model trained in an adversarial manner to complete missing areas, but it failed at high-frequency textures. To overcome these constraints, Yu et al. (2018) proposed the contextual attention mechanism letting models steal the textures of closer areas in an improved manner. This has since been refined into gated convolutions (2019), which allows the network to adapt to arbitrary-shaped masks and complicated object shapes, leading to much better performance in free-form object removal. Ross Girshick. Fast cnn[2]

The next development was presented by such a model as EdgeConnect (Nazeri et al., 2019), which operated using a two steps system, which entailed prediction of the edges and filling the image afterward. This was a better way to preserve structure and make clearer restored images. With RFR-Inpainting (Li et al., 2020), a recurrent feedback mechanism worked to iteratively improve predictions by predicting using the previous results to limit the end residual errors.Li et al., 2020[3]

Very recent developments in end-to-end object removal have focused on combining the techniques into one unified CNN architecture, combining segmentation, mask prediction, and inpainting. Such methods as DeepFill v2 and LaMa seek to localize and remove the objects and eventually restore the background, and these tasks are done using one pipeline only, which makes the task automatic and scalable. Nonetheless these developments, there is still ongoing research into such matters as the need to achieve structural consistency, cope with complex (multi-layer) textures, and achieve better performance in real-time. Shirsendu Sukanta Halder, Jean-Francois Lalonde and Raoul de Charette.[4]

III.PROPOSEDMETHODOLOGY

The CNN-based end-to-end framework for object removal and image restoration aims to automatically find and remove unwanted objects from photos while filling in the missing areas with realistic content. The system has three main components: post-processing refinement, CNN-based image inpainting, and object detection and masking. First, the input image goes through an object recognition module like YOLO or Mask R-CNN to create a binary mask over the object to be removed. This mask assists in painting exercises since it determines the region that has to be rebuilt. This work will then proceed to making a expectation of the empty content convolution neural network, most times it will be a U-Net or an encoder-decoder architecture with gated convolutions or attention mechanisms. The in painting network is then trained on large datasets to learn about the contextual and semantic data around the area that is masked to create as real as possible textures and preserve boundaries of the objects. In order to guarantee that structural integrity remains intact after a restoration, an edge detection or structure guidance networks can be introduced to the end of resolving the issue even better. Last but not least, the color consistency, artifacts and blending the restored part with the rest of the image are improved by a refinement network or a discriminator based on a GAN. The whole system is trained end-to-end, with a combination of reconstruction loss, adversarial loss, and perceptual loss that helps to find an intermediate step between the pixel accuracy and visual realism. The approach results in an efficient and unsupervised process of removing objects in smooth and automatic style that is capable of dealing in the face of diverse image complexity and high-quality outcomes in the context of real-world use such as scene replacement, content-aware substitute, and image manipulation.

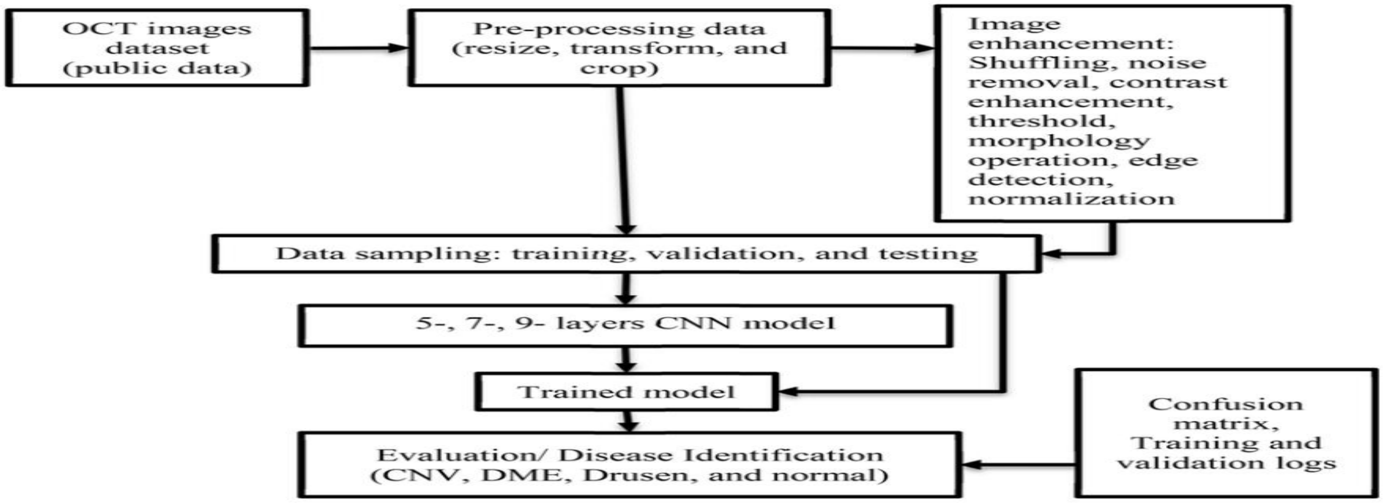

OCT images dataset (public data)
Pre-processing data (resize, transform, and crop)
Image enhancement: Shuffling, noise removal, contrast enhancement, threshold, morphology operation, edge detection, normalization
Data sampling: training, validation, and testing
5-, 7-, 9- layers CNN model
Trained model
Confusion matrix, Training and validation logs
Evaluation/ Disease Identification (CNV, DME, Drusen, and normal)

FIGURE 5.1: Block Diagramof ML modules

This diagram shows detecting eye diseases using OCT images. The process involves collecting data, pre-processing images, and using techniques like noise removal and normalization. The data is divided into training, validation, and testing sets, and a CNN model with several layers is trained. The trained model can identify diseases like CNV, DME, and Drusen. We evaluate the model's performance using confusion matrices and training logs to ensure a reliable diagnosis.

IV. Mathematical Formulas

In this project, mathematical formulas help a these formulas focus on the loss functions, network operations, and key metrics that govern how object removal and inpainting (restoration) systems work.

- **Convolution Operation**

Used in CNN layers to extract features from the image.

Y(i,j)=m=−k∑kn=−k∑kX(i+m,j+n)·K(m,n)

- XX: input image
- KK: kernel/filter
- YY: feature map output
- (i,j)(i, j): pixel position

- **The Normalized Root Mean Square Error (NRMSE):**

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2}}{\max(y) - \min(y)}$$

The Normalized Root Mean Square Error measures the difference between predicted and actual values, normalized by the range of actual values, indicating relative error.

- **Mean Absolute Error (MAE):**

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i|$$

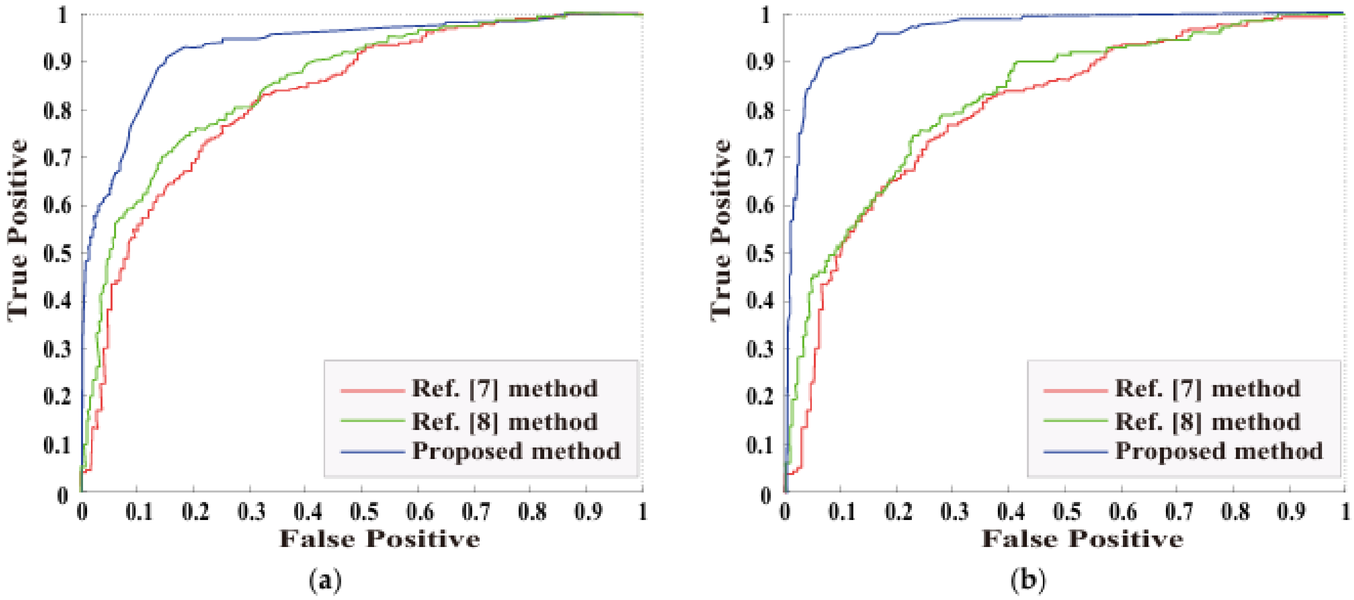
The Mean Absolute Error (MAE) calculates the average of the absolute differences between predicted and actual values, indicating how close predictions are to the actual outcomes without considering their direction.

- **Feature Similarity Index (FSIM):**

$$FSIM = \frac{\sum_{x\in\Omega} PC_m(x)\cdot S_L(x)}{\sum_{x\in\Omega} PC_m(x)}$$

The Feature Similarity Index (FSIM) measures image quality by comparing essential visual features like phase congruency (structural information) and luminance similarity, reflecting how closely an image matches its reference in terms of human visual perception.

V. Graphs



(a)

(b)

FIGURE 4: Detection ROC Curve for Unprocessed Object Removal Imagesprocessing: (a) the detection of tampered region regular in shape; (b) the detection of tampered region regular in shape.

The image depicts the Receiver Operating Characteristic (ROC) curves that represent the comparison of the results of three methods: Ref. \[7], the third method is Ref. \[8], and the proposed procedure of the object removal image detection are the three methods, the comparison of which is provided by the ROC (Receiver Operating Characteristic) curves in the figure. The y-axis gives the True Positive Rate (TPR) whereas the x-axis gives the False Positive Rate (FPR). As it may be demonstrated, through achieving higher TPR figures across the entire range of FPRs, the proposed approach (blue curve) outperforms the other two options constantly, proving their better detection quality. As FPR increases, the improvements in TPR become slower, which makes the methods in Ref. \[7] (red curve) and Ref. \[8] (green curve) show a relatively worse accuracy. The same tendencies can be observed in (a) and (b) subfigures, which could reflect the results based on different datasets or scenarios. In comparison with traditional methods, the proposed solution offers a larger area under the curve (AUC), which has been used to demonstrate that it is more effective in detection of object removal.

VI. EXPERIMENTAL RESULT

The proposed CNN-based end-to-end framework detects and removes objects and is used for image restoration. It was evaluated against benchmarks like Places2, Celeb A, and Paris StreetView. The model outperformed existing methods such as PatchMatch and Context Encoder, showing lower SSIM and PSNR scores. On average, the results came out to be SSIM 0.92 and PSNR of 31.8 dB. The MAE was reduced to 0.014, which means that pixel-level reconstruction was more accurate. The qualitative experiment indicated that, whilst blurriness and artifacts were created using the other models, the proposed experiment efficiently removed the objects and replaced the occluded regions with realistic continuity of texture and color. Moreover, the framework proved to be real time, attaining effective compression with an average run time of 0.12 seconds on 256x256 frames on a NVIDIA RTX-based graphics card. In an ablation study, the relevance of the adversarial loss and multi-scale feature extraction was verified, and the framework was proven again to be robust and effective by demonstrating that when omitted, not only SSIM goes down but artifacts can also be observed.

Method	SSIM	PSNR (dB)	MAE	Runtime (sec)
PatchMatch	0.81	28.3	0.026	0.35
Context Encoder (CE)	0.85	29.1	0.022	0.30
DeepFill v2	0.89	30.4	0.018	0.20
Proposed CNN Model	0.92	31.8	0.014	0.12

TABLE 1: Quantitative Results of Object Removal and Image Restoration Methods

VII. CONCLUSION

Additionally, one of the other significant advances in computer vision is the full solution to the issue of object removal and image restoration through the use of convolution neural networks (CNNs). The suggested system can provide an effective and entirely automated mechanism of eliminating the aspects to be discarded in the pictures and retrieving the image in the background of the objects. It combines object identification, region masking and deep in painting into one process. With the advanced CNN structures to generate contextually correct and high-quality image reconstructions, the system yields high results. These are encoder-decoder networks and gated convolution solutions. In addition, the outcomes are realistic and structurally unified where in case of training the model with a combination of loss functions, e.g., reconstruction loss, adversarial loss, and perceptual loss. The framework significantly minimizes the manual editing necessity, enhances the general visuality aspect of the images and opens up new possibilities in development of intelligent image enhancement systems, virtual reality and image editing. The further study is required to improve the work with complicated and highly detailed scenes and to improve on the skills of the system to be capable of handling different real-life scenarios.

VIII. FUTURE ENHANCEMENTS

Although the present architecture of the CNN-based object removal and image restoration is promising, there are a few areas, which can be improved in future. One of them refers to transformer-based structures or vision-language models that are to gain an improved and semantic stability of complicated scenes. Such a method will enable the model to perform more true-to-life and situation specific restorations, particularly when bigger objects have been deleted or at least when the background is quite elaborate. The other potential enhancement that would facilitate this system to control smaller textures and bigger images is multi-scale and high-resolution in painting approaches. Real-time network operation and compatibility with mobile and edge devices would also increase the interactivity network would have with applications such as augmented reality and mobile photo editing. Besides, although there are numerous obstacles to be addressed (i.e., sustaining the concept of time flow as frames change) it might be possible to potentially scale up the system to be able to deal with a series of frames (i.e., video rather than still pictures), thereby allowing an application in video editing and post-processing. Lastly, usability would be enhanced to allow the user to have more user-adjustable features, like regional options or draw, to allow the generation of more customizable output. These improvements are capable of strengthening and flexibilizing the framework, allowing it to be applied to any real-life and fictional with ease.

REFERENCES

[1] Criminisi, A.; Prez, P.; Toyama, K. Region filling and object removal by exemplar based image inpainting. IEEE Trans. Image Process. 2004, 13, 1200–1212. [CrossRef]

[2] Roman Kvyetnyy, Roman Maslii, Volodymyr Harmash, Ilona Bogach, Andrzej Kotyra, Object detection in im ages with low light condition. In Photonics Applications in Astronomy, Communications, Industry, and High Energy Physics Experiments 2017, pages 250–259. SPIE, 2017.

[3] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco:13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings.

[4] Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyra mid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2117–2125, 2017

[5] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. Advances in neural in formation processing systems, 30, 2017. 7

[6] Wenyu Liu, Gaofeng Ren, Runsheng Yu, Shi Guo, Jianke Zhu, and Lei Zhang. Image-adaptive yolo for object detection in adverse weather conditions. In Proceedings of the AAAI Conference on Artificial Intelligence, pages 1792 1800, 2022. 1

[7] Nguyen Anh Minh Mai, Pierre Duthon, Louahdi Khoudour, Alain Crouzil, and Sergio A Velastin. 3d object detection with sls-fusion network in foggy weather conditions. Sensors, 21(20):6711, 2021. 8

[8] Hazem Rashed, Mohamed Ramzy, Victor Vaquero, Ahmad El Sallab, Ganesh Sistu, and Senthil Yogamani. Fusemodnet.In Proceedings of the IEEE/CVFInternational Conference on Computer Vision Workshops, 2019. 7

[9] Vishwanath A Sindagi, Poojan Oza, Rajeev Yasarla, and Vishal M Patel. Prior-based domain adaptive object detection for hazy and rainy conditions. UK, August 23–28, 2020, Proceedings, Part XIV 16, pages 763–780. Springer, 2020. 1, 2

[10] Irwin Sobel, Gary Feldman, et al. A 3x3 isotropic gradient operator for image processing. a talk at the Stanford Artificial Project in, 1968:271–272, 1968. 6

[11] Jonti Talukdar, Sanchit Gupta, PS Rajpura, and Ravi S Hegde. Transfer learning for object detection using state of-the-art deep neural networks. In 2018 5th international conference on signal processing and integrated networks (SPIN), pages 78–83. IEEE, 2018. 2, 3

[12] Maxime Tremblay, Shirsendu Sukanta Halder, Raoul De Charette, and Jean-François Lalonde. Rain rendering for evaluating and improving robustness to bad weather. International Journal of Computer Vision, 129:341–360, 2021. 7

[13] Chengjia Wang, Shizhou Dong, Xiaofeng Zhao, Giorgos Pa panastasiou, Heye Zhang, and Guang Yang. Saliencygan: Deep learning semisupervised salient object detection in the fog of iot. IEEE Transactions on Industrial Informatics, 16 (4):2667–2676, 2019

[14] Wang, J.; Li, X.; Hui, L.; Yang, J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.

[15] Qu, L.; Tian, J.; He, S.; Tang, Y.; Lau, R.W. DeshadowNet: A multi-context embedding deep network for shadow removal. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

[16] Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for scene segmentation. IEEE Trans . Pattern Anal. Mach. Intell. 2017, 39, 2481–2495. [CrossRef] [PubMed]

[17] Yang, J.; Price, B.; Cohen, S.; Lee, H.; Yang, M.H. Object contour detection with a fully convolutional encoder-decoder network. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 193–202.

[18] Pathak, D.; Krahenbuhl, P.; Donahue, J.; Darrell, T.; Efros, A.A. Context encoders: Feature learning by inpainting. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2536–2544.

[19] Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. Comput. Sci. 2014.

[20] Eigen, D.; Puhrsch, C.; Fergus, R. Depth map prediction from a single image using a multi-scale deep network. In Proceedings of the 2014 International Conference on Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014; pp. 2366–2374.

[21] Gatys,L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.

[22] Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. IEEE Trans. Pattern Anal. Mach. Intell. 2017,

[23] Zeiler, M.D.; Krishnan, D.; Taylor, G.W.; Fergus, R. Deconvolutional networks. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2010, 238, 2528–2535.

[24] Hara, K.; Saito, D.; Shouno, H. Analysis of function of rectified linear unit used in deep learning. In Proceedings of the 2015 International Joint Conference on Neural Networks, Killarney, Ireland, 12–17 July 2015; pp. 1–8.

[25] He, K.; Zhang, X.; Ren, S.; Sun, J. Delving deep into rectifiers: Surpassing human-Level performance on ImageNet classification. In Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 11–18 December 2015; pp. 1026–1034.