Deep Learning Based Hand Wrist Segmentation using Mask R-CNN

GokulaKrishnan Elumalai
School of Computer Science and Engineering,
Vellore Institute of Technology, Chennai, India
Gokul.krish55@gmail.com

Malathi Ganesan
School of Computer Science and Engineering,
Vellore Institute of Technology, Chennai, India
malathi.g@vit.ac.in

Abstract: *Deep learning is one of the trending technologies in computer vision to identify and classify objects. Deep learning is a subset of Machine Learning and Artificial Intelligence. Detecting and classifying the object was a challenging task in traditional computer vision techniques, and now there are numerous deep learning techniques scaled up to achieve this. The primary purpose of the research is to detect and segment the human hand wrist region using deep learning methods. This research is widespread to deep learning enthusiasts who needs to segment custom objects using instance segmentation. We demonstrated a segmented hand wrist using the Mask Regional Convolutional Neural Network (R-CNN) technique with an average accuracy of 99.73%. This work also compares the performance evaluation of baseline and custom Hand Wrist Mask R-CNN. The achieved validation class loss is 0.00866 training and 0.02736 validation; both the values are comparatively deficient compared with baseline Mask R-CNN.*

Keywords: *Hand wrist, segmentation, mask R-CNN, fast R-CNN, faster R-CNN, object detection.*

Received June 27, 2020; accepted October 13, 2021
https://doi.org/10.34028/iajit/19/5/10

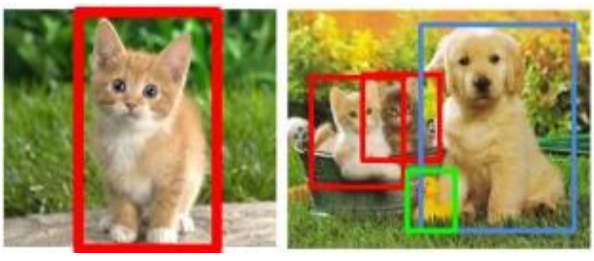
1. Introduction

These days image segmentation is one of the sensitive topics in the computer vision world. The traditional image segmentation technique partitions the image into multiple chunks. The main objective of image segmentation is to instantiate the significant image to be easier to analyze. Image segmentation is the process of grouping a similar set of attributes like edges, color, intensity, or textures. In most situations, interest lies in the extraction of the region of interest of the semantic image. Over the past years, many researchers have effectively developed the image segmentation algorithm to solve domain-specific problems such as medical imaging, video surveillance, automated driving system, and machine vision [15]. Several techniques have been proposed, such as thresholding, cluster-based segmentation, and graph partitioning; these methods are suitable in some circumstances, such as calculation difficulties and fewer parameters. Still, these traditional segmentation techniques need to be fine-tuned on performing the segmentation task without any explicit natural information. The stepping of deep learning into the computer vision world, the traditional performance issues faced in segmentation, has been achieved with the help of the Convolution Neural Network (CNN).

Image Segmentation can be laid down into a classification problem [20]; if there is one object in an image (cat), we can quickly build up the (cat-dog) classifier model and predict which object is available in an image. Leading to questions such as what happens if more than one object is present in a single image? Does it pose a need to train a multi-label classifier to predict each instance? Another issue is that the exact location of each object is not known. Generally, we rely on two major components:

1. Image localization assists us in recognizing the single object in each image.
2. Object detection helps us to identify the class of the individual object in the image.

Below given, Figure 1-a) shows the cat and Figure 1-b) dog classifier using localization and detecting each object.

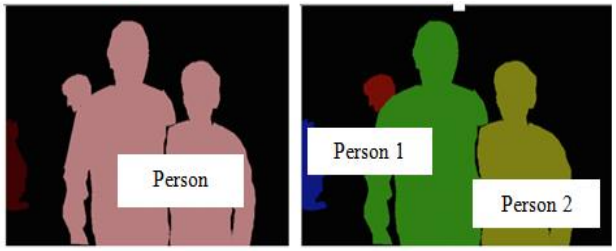


a) Classification. b) Localization.

Figure 1. Classification localization and object detection.

Before classifying or detecting the objects in the image, it requires to find what the image comprises. So, this is how segmentation comes into a clear picture of partitioning similar objects. Semantic segmentation fences off foreground and background in an image; for example, every pixel corresponds to the class. In instance segmentation, fencing off each different object

is of the same class with different fencing colors. In Figure 2-a) and Figure 2-b), visual representation of types of segmentation without explicit programming, deep-learning allows us to complete a more complex task using machines. Researchers are more focused on deep learning techniques to solve real-world problems. These are the basic deep learning Techniques that are getting to know, such as Fully Convolution Neural Networks (FCNN), CNN, Recurrent Neural Networks (RCNN), Generative Adversarial Networks (GAN), and Deep Reinforcement Techniques used to solve a more complex task in the field of computer vision. In GokulaKrishnan and Malathi [10] did a complete survey on Hand Biometrics which gives a clear perspective on types of hand biometrics.



a) Semantic. b) Instance.

Figure 2. Semantic and instance segmentation.

This paper mainly focuses on hand wrist segmentation using Mask-RCNN techniques using these techniques to achieve instance segmentation. Segmentation of the hand wrist is a challenging task in traditional computer vision techniques. With the advent of deep learning, it is possible to achieve more accuracy and performance in segmentation. The core principle of hand wrist segmentation is to use it as a biometric identification of unique individuals. This paper elaborates in detail on how the hand wrist is segmented using mask-RCNN. In section 2, we will deep dive into different CNN technique which has been applied in research works. Section 3 elaborates our proposed works of hand wrist segmentation. Section 4 tells about our proposed work results.

2. Related Works

Malathi and Shanthi [19], the traditional segmentation techniques were employed to develop the prototype and classify the ultrasound placenta image as normal or abnormal based on the watershed approach to attain the statistical measurement of the stereo-mapped placenta image. Based on the statistical measurement, we can predict the gestational diabetics during pregnancy so that the growth of the foetus is monitored closely. Region-based image redemption using automatic segmentation [2], based on the query images geodesic-based segmentation. Using Euclidean distance, relevant images are fetched. The experimental results show high efficiency. In the past few years, CNN's are evolved drastically to classify the images with reasonable accuracy. Classifying the single object using CNNs is easy, but in the real world, the requirement is to classify the multiple objects in a single image, which is a tedious task.

In Deep learning few challenges posed are detecting the multiple overlapping objects in a single image. The Regional Convolutional Neural Network (R-CNN) approach [9] comprises three primary tasks:

1. Using CNNs to identify the region proposal with the help of localization and segmenting the object.
2. Supervised training the images with the help of creating a bounding box and labeling each object in the image.
3. At the final step, extracted features are loaded into SVM to arrange the occurrences of the object.

It is a high-level selective search for the corresponding class to recognize the objects in the image. The team participated in the ImageNet Large scale visual recognition challenge (ILSVRC-2013) R-CNN achieves mean average precision (mAP) of 31.4% over the 200-classes of classifying the objects. R-CNN Techniques has few drawbacks for training the networks it takes enormous time since it leads to classifying the 2000 region proposals, and more testing time is required to test the image (47 sec). To overcome the few drawbacks of RCNN, [8] Fast R-CNN is proposed. This approach is more similar to R-CNN; instead of provisioning the region proposal to the CNN, the convolutional feature map is initiated, which identify the region proposals and wrap them into squares and finally reshape the images into fixed-size using ROI pooling layer and to predict the class SoftMax layer is used to offset the bounding boxes.

The primary motivation of the Fast R-CNN is to reduce the training and testing time. Even though Fast R-CNN has improved to reduce the training time still there is a bottleneck on this technique because, at the initial stage, Fast R-CNN uses a region proposer with a bunch of bounding boxes to detect the location of the object, and also it uses the selective search which has a slow process which affects the overall performance. Faster R-CNN [16] arrived to solve this current performance issue; instead of selective search, they used the same CNN results for region proposals because region proposals depend on features of the image, which were already calculated using the forward pass CNN. Using this logic, Faster RCNN achieved more performance while training and testing the image with reasonable accuracy.

Mask R-CNN [12] has been developed underlying with Faster/Fast R-CNN technique using this approach they extended from object detection to carry out pixel-level segmentation. The central intuition is that providing the input image generates the binary mask with fencing based upon which it is easy to classify the object's class with segmentation. In recent years many researchers have shown interest in Mask R-CNN for

identifying and segmenting the objects and license plate [18] recognition system using Mask R-CNN with different angles. The study conducted the experiments with different tilts angles (0-60 degrees) and achieved the mAP rates of 90% accuracy compared to the YOLOV2 method. Review [7], Semantic segmentation with different Deep learning techniques pinpoints the difficulties in semantic segmentation for creating the datasets. The CT lung segmentation using Mask R-CNN [13] with the help of machine learning technique both supervised and unsupervised learning (K-means kernel), an automated lung region has been mapped successfully with reasonable accuracy. It was experimented [22] with cattle segmentation based on Mask R-CNN in the cattle feed farming based on cattle instance segmentation and critical frame extraction of detecting the cattle segmentation. The study outperforms contours extraction with MPA 92% and an Average Distance Error (ADE) of 33.56 pixels.

Performance of U-Net and Mask R-CNN segmenting the pomegranate Tree canopy and found Mask R-CNN achieves more performance [26] mAP 98.5% over 61.2 U-net and concludes Mask R-CNN provides better performance. A framework [6] for detecting and segmenting the orange fruit using Mask R-CNN is based on the segmentation of pixel-wise RGB and RGB+HSV images. The study experimented and concluded that the score obtained from RGB+HSV is 0.89. Robots are evolved more robustly in the AI era. Harvesting the strawberry with the help of robots [25] in a non-structural environment using Mask R-CNN using visual localizing method has been used to pick the ripe strawberry fruits and performed well with an accuracy of 95.41%. Recognizing the Arab script based on Deep Learning (DL) [1] used a complex KHATT data set to recognize each character. The proposed technique is based on Multi-Dimensional Long Short-Term Memory (MDLSTM) networks which scan the Arabic text lines and provide better recognizing accuracy of 80.02%, which sets the benchmark. A prototype [24] that uses DL methods to detect intrusion and achieved the 99.97% to detect the intrusion. Deep Neural Network (DNN) [4] was used to predict the gene expression using the GEO dataset, and prediction accuracy is 49.23% compared with D-GEX. DL and Artificial Neural Network (ANN) Techniques [14, 23] are used to segment and detect breast lesions, preventing breast cancer. As per the literature, it is found that Mask R-CNN has been used in various aspect domains and concludes the Mask R-CNN outperforms well in instance segmentation to extract the required regions.

3. Methods

The proposed system for segmenting the hand wrist is based on the Mask R-CNN technique; before entering hand wrist segmentation, this section will include a deep dive into mask R-CNN's internal architecture.

3.1. Mask R-CNN Architecture

The main objective of the mask R-CNN is to solve the instance segmentation problem in computer vision. When an input image is loaded to mask R-CNN, it separates the objects based on bounding boxes, classes, and mask. Since the Faster R-CNN performs well for object detection adding to the top of that, the image is segmented. The sum of Faster R-CNN and Mask is termed as Mask R-CNN.

Mask R-CNN = Faster R-CNN + Mask

To implement Mask R-CNN, there are two stages. In the first stage, loading an input image will generate the Region Proposal Network (RPN) based on the objects. So, while scanning the feature map, Anchors will mark the boxes based on the ground-truth classes according to the Intersection over Union (IoU) value. At the second stage, we need to propose another neural network to generate the regions mask using Region of Interest (ROI) Align to locate the objects of the feature map. The below Figure 3 elaborates the two stages of implementing the mask R-CNN which we explained crisply.

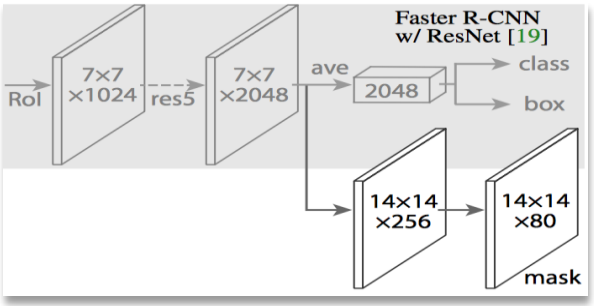


Figure 3. Mask R-CNN architecture.

3.2. Dataset Preparation

In this experiment, images of around 500 people (Male and Female), including left-Hand and right-Hand Wrist images, were collected and stored in the Hand Wrist dataset. These images are captured in a mobile camera with a 5 to 10 cm distance with a natural light source. The sample images comprised different age groups between 20-40 years old and used the standard naming convention of the individual Hand Wrist images [11]. Hand Wrist dataset naming conventions followed as below.

HW-Hand Wrist
M/F-Male/Female
L/R-Left/Right

For example, HWML000001 indicates the person's left-hand male hand wrist.

3.3. Pre-Processing

Since the image is captured using a mobile camera, there is always a possibility of blurred images while capturing

hand wrist images. A variance of the Laplacian technique is applied to detect the image is blurred or not. There are many techniques to assess image quality. The comparative study [21] on autofocusing methods found that a variance of Laplacian methods produced better results.

Algorithm: Hand Wrist Blur Image Detection using Variance of Laplacian (HWBID)

1. Load the input image 1…N
*2. Apply Laplacian kernel 3*3 for all the pixel x, y*

$$\begin{matrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{matrix}$$

3. Take mean-variance

$$\emptyset_{i,j} = \sum_{(i,j) \in \Omega(x,y)} (\Delta I(i,j) - \overline{\Delta I})^2 \quad (1)$$

Where $\overline{\Delta I}$ *= Mean value of image Laplacian within* $\Omega(x, y)$.

4. Predict the image is blurred or unblurred

$$\emptyset_{i,j} = \begin{cases} if\ \emptyset_{i,j} < threshold = blurred\ image \\ else\ \emptyset_{i,j} > threshold = \ not\ a\ blurred\ image \end{cases}$$

Where $\emptyset_{i,j}$ *= Laplacian individual image mean-variance.*

5. Iterate all the images

$$for\ iter = 1\ to\ N_{iter}\ do$$
$$for\ each\ images\ in\ trainingDir$$
$$repeat\ steps\ 1-4$$
$$end\ for$$
$$end\ for$$
$$where\ N_{iter} = Total\ number\ of\ images$$

The above algorithm describes the flow of blurred image detection using the mean-variance method. This method is an upfront technique where we will be implementing the single-channel image using the Laplacian kernel. After generating the mean value of Laplacian, by applying the threshold values, we will able to predict image is blurred or not. If an image is blurred, we will be removing the images in the training directory and this technique improves the accuracy for segmenting the Hand Wrist image.

3.4. Training

Before training, the environment needs to be set up. Google Colab is used for training and testing the custom Hand Wrist Mask R-CNN. Colab notebook codes are executed in a google server that has high-power computing. The dataset is prepared in such a way that 90% is used for training and 10% testing. The hand wrist region images placed in the training folder are annotated using the Visual Geometry Group (VGG) image annotation tool, generating JavaScript Object Notation (JSON) data. The data set is loaded into google drive and extracted for training. Resnet 101 has been used as backbone architecture. The configuration of Mask R-CNN is customized by providing necessary configuration like learning rate, classes, epochs, etc., and the Mask R-CNN is initialized for the training process with TensorFlow.

3.5. Internal workflow of Mask R-CNN using Hand Wrist

The below Figure 4 visually demonstrates the internal workflow of hand wrist segmentation using mask R-CNN. After training the mask R-CNN, the trained weight is used to predict and segment the object of the class along with instance segmentation. During the inference of hand wrist mask, R-CNN clearly identifies the person hand wrist alone. The mask R-CNN architecture with custom trained configuration along with custom inputs is used to segment the hand wrist region alone.

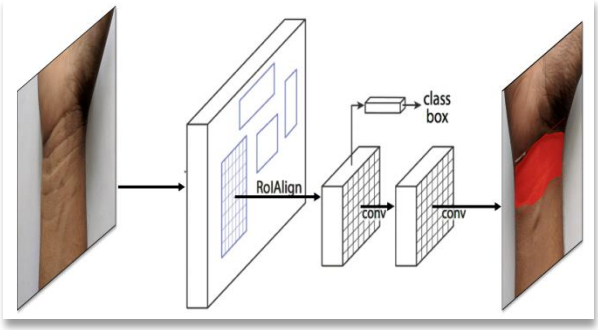


Figure 4. Hand wrist: mask R-CNN internal workflow.

3.6. Overall Architecture

Figure 5 depicts the overall flow of hand wrist segmentation using mask R-CNN. We segregated the two flows, training and testing. In training, collected hand wrist has been used for pre-processing. Using variance of Laplacian, we removed the blurred images in the folder, and the remaining images are considered for training the custom mask R-CNN and generating the trained weight. h5 files. During testing, the pre-trained weights are used to predict the hand wrist class region and fencing the instance segmentation. The hand wrist region is identified, and the segmented region alone is extracted, retaining the extracted region alone and remaining pixel values set to 0.

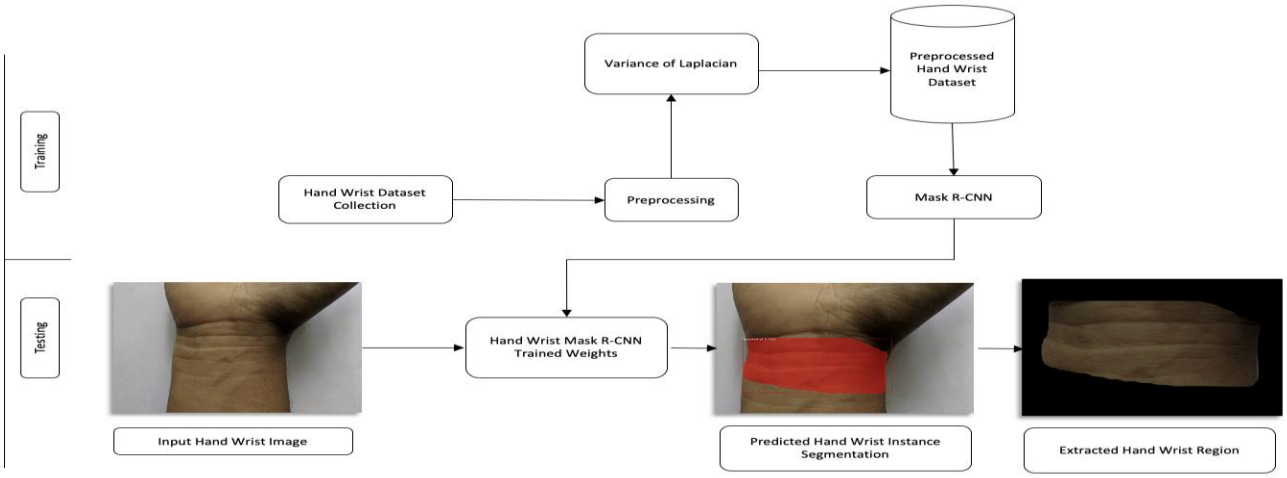


Figure 5. Overall architecture of custom mask R-CNN.

4. Results

Using Mask R-CNN, Hand wrist datasets have been trained in google colab with 30 epochs; each epoch carries 100 steps 3000 steps the training has taken place, and the duration of the training time to complete 30 epochs taken around 2.21 hours.

4.1. Performance Evaluation

- *Intersection over Union (IoU)*

The ratio between intersection and union of the predicted and ground truth boxes is called IoU [3].

$$\text{IoU} = \frac{\text{Area of Intersection}}{\text{Area of Union}} \quad (2)$$

If IoU > 0.9, then it is a true positive
If IoU < 0.9, then it is a false positive

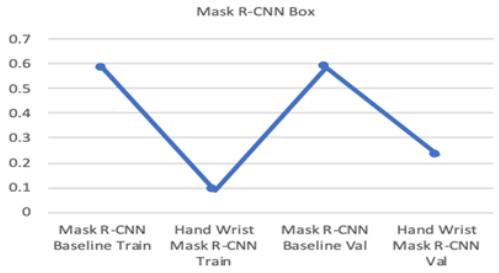
Mean Average Precision (mAP) and Mean Average Recall (mAR) [17] has been calculated using the below formula represented in Equations (3), and (4).

$$\text{Precison} = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \quad (3)$$

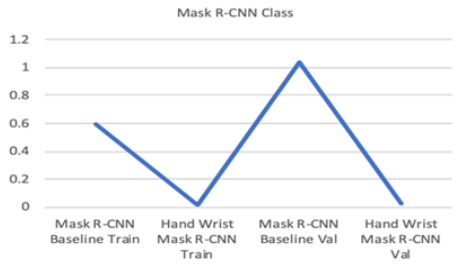
$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (4)$$

Table 1. Loss performance calculation of baseline and mask R-CNN.

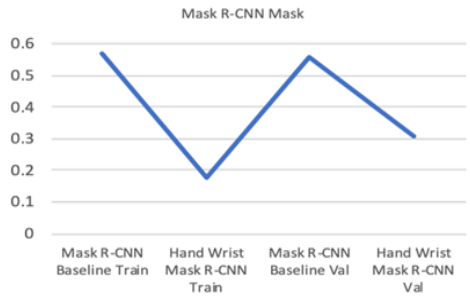
Loss Comparison	Dataset	Mask R-CNN Box	Mask R-CNN Class	Mask R-CNN Mask	Overall Loss
Mask R-CNN Baseline Train	COCO	0.5817	0.5912	0.568	0.65
Mask R-CNN Baseline Val	COCO	0.5869	1.039	0.5583	
Hand Wrist Mask R-CNN Train	Hand Wrist	0.089663	0.00866	0.17466	**0.39**
Hand Wrist Mask R-CNN Val	Hand Wrist	0.23294	0.02736	0.3055	
Pulmonary Nodule Mask R-CNN Val	LUNA16	0.516	0.680	0.801	0.79



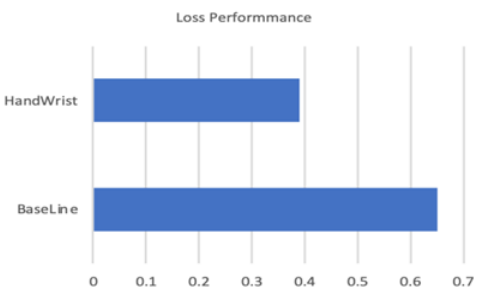
a) Mask R-CNN box-baseline VS hand wrist.



b) Mask R-CNN class-baseline VS hand wrist.



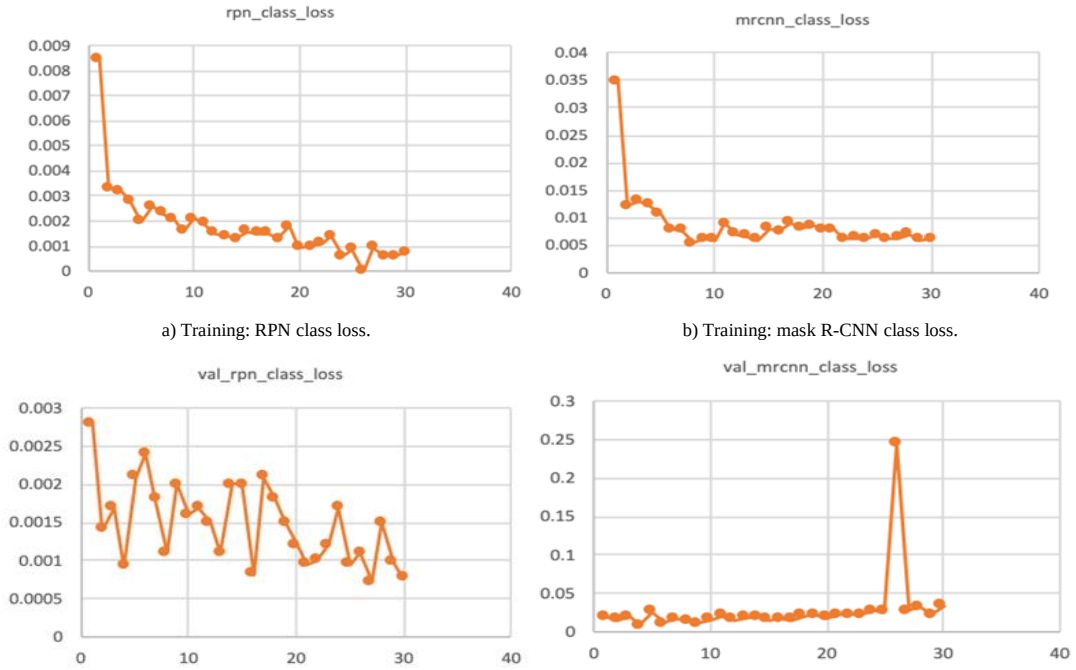
c) Mask R-CNN mask-baseline VS hand wrist.



d) Loss performance-baseline VS hand wrist.

Figure 6. Hand wrist mask R-CNN vs Mask R-CNN baseline loss performance graph.

The results are evaluated based on research community standards for object detection like mean average precision and mean average recall. In Figure 7-a), Figure 7-b), Figure 7-c) and Figure 7-d), indicates training vs validation of RPN and Mask R-CNN loss is plotted in graph. This model is based on Facebook Research Artificial Intelligence of Mask R-CNN with customized fine-tuned parameters that are used to detect the Hand wrist mask region alone. The IoU threshold is set as 0.9 to detect the more accuracy of Hand Wrist objects. Table 1 calculated the loss performance evaluation of the existing Commo Object in context (COCO) dataset and Hand Wrist and found the excellent performance of training class loss 0.008 when compared to baseline train 0.5912 of Mask R-CNN. The same validation loss also performs well 0.02736 of Hand Wrist and 1.039 for validation baseline. We also compared our results with [5] Pulmonary Nodule segmentation using Mask R-CNN where overall loss performance is 0.79. In Table 1 clearly shows that the loss performance is very low for Handwrits 0.39 when compared to all other Mask R-CNN methods. In Figure 6-a), Figure 6-b) and Figure 6-c) denotes graphical representation of baseline VS hand wrist Mask R-CNN box, class and mask. In Figure 6-d), we also plotted loss performance graph for the Hand Wrist Mask R-CNN and Baseline Mask R-CNN. Graphical chart clearly shows that Hand wrist loss performance is very low when compared to Baseline. In Table 2, the confidence percentage of Hand wrist object detection is evaluated. The test dataset of Hand wrist is validated and found that the average object detection accuracy is 0.987.



a) Training: RPN class loss.

b) Training: mask R-CNN class loss.

c) Validation: RPN class loss.

d) Validation: mask-RCNN class loss.

Figure 7. Hand wrist: training and validation graph for Region Proposal Network (RPN) class loss and mask R-CNN class loss.

Table 2. Sample hand wrist object detection validation confidence percentage using mask R-CNN.

Image Name	Object Detection Confidence%
HWML000401	99.70 %
HWMR000401	99.80%
HWML000402	99.10%
HWMR000402	99.90%
HWML000403	99.80%
HWMR000403	99.70%
HWML000404	99.70%
HWMR000404	99.90%
HWML000405	99.90%
HWMR000405	99.80%

5. Conclusions

The primary focus of this research paper is to provide in-depth knowledge on deep learning using mask R-CNN in various domains. We also demonstrated novel custom hand wrist instance segmentation with a detailed explanation of architecture and internals. The comparative loss percentage of existing baseline, Pulmonary Nodule VS hand wrist mask R-CNN in Table1 is calculated and found that Hand Wrist mask R-CNN class loss and mask loss are comparatively low. In the future, this research can be extended to detect and extract the hand wrist features for comparing the individual identification. There are a few challenges faced in this research. If the size of the datasets increases enormously, training time gets increased to generate the model, which is found to be a significant drawback. In the future, this paper may be extended and fine-tuned to improve this custom mask R-CNN class and mask loss.

References

[1] Ahmad R., Naz S., Afzal M., Rashid S., Liwicki M., and Dengel A., "A Deep Learning based Arabic Script Recognition System: Benchmark on KHAT" *The International Arab Journal of Information Technology*, vol. 17, no. 3, pp. 299-305, 2020.

[2] Amin A. and Qureshi M., "A Novel Image Retrieval Technique using Automatic and Interactive Segmentation," *The International Arab Journal of Information Technology*, vol. 17, no. 3, pp. 404-410, 2020.

[3] Badruswamy S., "Evaluating Mask R-CNN Performance for Indoor Scene Understanding," 2018.

[4] Bhukya R. and Ashok A., "Gene Expression Prediction Using Deep Neural Networks," *The International Arab Journal of Information Technology*, vol. 17, no. 3, pp. 422-431, 2020.

[5] Cai L., Long T., Dai Y., and Huang Y., "Mask R-CNN-Based Detection and Segmentation for Pulmonary Nodule 3D Visualization Diagnosis," *IEEE Access*, vol. 8, pp. 44400-44409, 2020,

[6] Ganesh P., Volle K., Burks T., and Mehta S., "Deep Orange: Mask R-CNN based Orange Detection and Segmentation," *ELSEVIER IFAC International Federation of Automatic Control*, vol. 52, no. 30, pp. 70-75, 2019.

[7] Garcia-Garcia A., Orts-Escolano S., Oprea S., Villena-Martinez V., and Garcia-Rodriguez J., "A Review on Deep Learning Techniques Applied to Semantic Segmentation," *arXiv:1704.06857v1*, 2017.

[8] Girshick R., "Fast R-CNN," *arXiv:1504.08083v2*, 2015.

[9] Girshick R., Donahue J., Darrell T., and Malik J., "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," *arXiv:1311.2524v5*, 2014.

[10] GokulaKrishnan E. and Malathi G., "A Survey on Multi-feature Hand Biometrics Recognition," *in Proceeding of Computational Vision and Bio Inspired Computing*, pp. 1061-1071, 2018.

[11] GokulaKrishnan E. and Malathi G., "Contactless Novel Hand Wrist Biometrics Feature Extraction using SURF," *International Journal of Civil Engineering and Technology*, vol. 9, no. 11, pp. 1102-1114, 2018.

[12] He K., Gkioxari G., Dollar P., and Girshick R., "Mask R-CNN," *arXiv:1703.06870v3*, 2018.

[13] Hu Q., Souza L., and Holanda G., Alves S., Silva F., Han T., and Filho P., "An Effective Approach for CT Lung Segmentation using Mask Region-based Convolutional Neural Networks," *Artificial Intelligence in Medicine*, vol. 103, pp. 101792, 2020.

[14] Inoue K., "Semantic Segmentation of Breast Lesion Using Deep Learning," *Ultrasound in Medicine and Biology*, vol. 45, pp. S52, 2019.

[15] Issac A., Manohar H., and Jain V., "Segmentation for Complex Background Images using Deep Learning Techniques," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 1746-1750, 2019.

[16] Jian S., Kaiming H., Ross G., and Shaoqing R., "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *arXiv:1506.01497v3*, 2016.

[17] Kopelowitz E. and Engelhard G., "Lung Nodules Detection and Segmentation Using 3D Mask-RCNN," *arXiv:1907.07676v6*, 2019.

[18] Lin C. and Li Y., "A License Plate Recognition System for Severe Tilt Angles Using Mask R-CNN," *in Proceeding of International Conference on Advanced Mechantronic Systems*, Kusatsu, pp. 229-234, 2019.

[19] Malathi G. and Shanthi V., "Statistical Measurement of Ultrasound Placenta Images Complicated by Gestational Diabetes Mellitus Using Segmented Approach," *International Conference on Signal and Image Processing*, vol. 2, no. 4, pp. 332-343, 2011.

[20] Minaee S., and Boykov Y., Porikli F., Plaza A., Kehtarnavaz N., and Terzopoulos D., "Image Segmentation Using Deep Learning: A Survey," *arXiv:2001.05566v4*, 2020.

[21] Pech-Pacheco J., Cristobal G., Chamorro-Martinez J., and Fernandez-Valdivia J., "Diatom Autofocusing in Brightfield Microscopy: A Comparative Study," *in Proceedings of 15th International Conference on Pattern Recognition*, Barcelona, pp. 314-317, 2000.

[22] Qiao y., Truman M., and Sukkarieh S., "Cattle Segmentation and Contour Extraction Based on Mask R-CNN for Precision Livestock Farming," *Computers and Electronics in Agriculture*, vol. 165, pp. 104958, 2019.

[23] Sen B. and Venugopal V., "Efficient Classification of Breast Lesion based on Deep Learning Technique," *Bonfring International Journal of Advances in Image Processing*, vol. 6, no. 1, pp. 1-6, 2016.

[24] Tabash B., Abd-Allah M., and Tawfik B., "Intrusion Detection Model Using Naive Bayes and Deep Learning Technique," *The International Arab Journal of Information Technology*, vol. 17, no. 2, pp. 215- 224, 2020.

[25] Yu Y., Zhang K., Yang L., Zhang D., and Kailiang Z., "Fruit Detection for Strawberry Harvesting Robot in Non-Structural Environment based on Mask-RCNN," *Computers and Electronics in Agriculture*, vol. 163, pp. 104846, 2019.

[26] Zhao T., Yang Y., Niu H., Wang D., and Chen Y., "Comparing U-Net Convolutional Network with Mask R-CNN in the Performances of Pomegranate Tree Canopy Segmentation," *in Proceeding of SPIE, Multispectral, Hyperspectral and Ultraspectral Remote Sensing Technology, Techniques and Application*, Honolulu, 2018.

GokulaKrishnan Elumalai is currently pursuing Ph.D. research scholar in Computer Science and Engineering from Vellore Institute of Technology, Chennai, Tamilnadu, India. His research interest includes Computer Vision, Deep Learning, Biometrics, and Artificial Intelligence. He filed a patent on his research areas on novel Biometrics.

Malathi Ganesan is working as a Professor in the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, Tamil Nadu, India. She has over 20 years of experience in teaching and research. Her area of specialization is Image Processing and Healthcare Analytics. She has been a resource person in FDPs, workshops, and Seminars. Currently, she is guiding 3 Ph.D. scholars and has several publications in reputed International Journals. She has authored a few book chapters. She has filed a patent in novel Biometrics. She received Best Outstanding Faculty Award in Computer Science 'by VENUS Foundation in July 2018.