International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 16, No.1 (Jul-24)

http://dx.doi.org/10.12785/ijcds/160116

An Instance Segmentation Method for Nesting Green Sea Turtle's Carapace using Mask R-CNN

Mohamad Syahiran Soria[1], Khalif Amir Zakry[2], Irwandi Hipiny[3], Hamimah Ujir[4], Ruhana Hassan[5] and Alphonsus Ligori Jerry[6]

[1,2,3,4] *Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak, Sarawak, Malaysia*
[5] *Faculty of Resource Science and Technology, Universiti Malaysia Sarawak, Sarawak, Malaysia*
[6] *Sarawak Forestry Corporation, Sarawak, Malaysia*

Received 17 Oct. 2023, Revised 14 Apr. 2024, Accepted 16 Apr. 2024, Published 1 Jul. 2024

Abstract: This research presents an improved instance segmentation method using Mask Region-based Convolutional Neural Network (Mask R-CNN) on nesting green sea turtles' images. The goal is to achieve precise segmentation to produce a dataset fit for future re-identification tasks. Using this method, we can skip the labour-intensive and tedious task of manual segmentation by automatically extracting the carapace as the Region-of-Interest (RoI). The task is non-trivial as the image dataset contains noise, blurry edges, and low contrast between the target object and background. These image defects are due to several factors, including jittering footage due to camera motion, the nesting event occurring during a low-light environment, and the inherent limitation of the Complementary Metal-Oxide-Semiconductor (CMOS) sensor used in the camera during our data collection. The CMOS sensor produces a high level of noise, which can manifest as random variations in pixel brightness or colour, especially in low-light conditions. These factors contribute to the degradation of image quality, causing difficulties when performing RoI segmentation of the carapaces. To address these challenges, this research proposes including Contrast-Limited Adaptive Histogram Equalization (CLAHE) as the data pre-processing step to train the model. CLAHE enhances contrast and increases differentiation between the carapace structure and the background elements. Our research findings demonstrate the effectiveness of Mask R-CNN when combined with CLAHE as the data pre-processing step. With CLAHE technique, there is an average increase of 1.55% in Intersection over Union (IoU) value compared to using Mask R-CNN alone. The optimal configuration managed an IoU value of 93.35%.

Keywords: Computer Vision, Instance Segmentation, Mask R-CNN, CLAHE, Deep Learning

1. Introduction

Chelonia Mydas is one of the seven species of sea turtles in the world. Despite being a commonly-sighted species worldwide, the green sea turtles is classified as an endangered animal, with its population declining tremendously over the years [1]. As part of the conservation initiative, researchers perform biometric re-identification to identify an individual from a biometric sample of the carapace scute pattern using learned image descriptors [2]. This task is done practically using computer vision and deep learning techniques with scute pattern imagery as input. Nevertheless, obtaining an accurate classification requires a ton of training data. During the training phase, the neural network will learn all the features inside the image. As a result, this can impact longer training time and require expensive computing resources to train the model to differentiate between the target object vs. the unnecessary image background. Significantly, the model might be trained on unneeded features affecting the classification prediction. It has been proven that the locations or backgrounds during training will impact the classification system because the model cannot generalize well to new locations [3]. Moreover, manually cropping a target object is a cumbersome and laborious task [4]. Therefore, there is a need to segment the carapace, focusing on the scute pattern, to obtain an accurate segmentation of the target object in an automated manner.

Different methods and approaches have been done to leverage the performance of Mask R-CNN to tackle noisy and low contrast images. For the first time, we proposed the advantages of CLAHE as a data preprocessing technique to nesting green sea turtles' images. By using CLAHE, we could improve the local contrast and enhance the visibility of the subject against the background which is clinical to produce better image segmentation.

E-mail address: 22020014@siswa.unimas.my, 21020463@siswa.unimas.my, mhihipni@unimas.my

The dataset has a few inherent image challenges that need to be solved to enhance Mask R-CNN's performance in predicting the turtle's carapace segmentation. In this paper, we evaluated Mask R-CNN performance using images extracted from videos of sea turtles captured during the nesting season. As we wish to reduce interferences with the animals, we used a HIK Vision camera to capture videos in low-light conditions, as shown in Figure 1. The camera model is a closed-circuit television (CCTV) type camera equipped with a CMOS sensor. CMOS is an image sensor technology commonly used in consumer-level digital cameras and other imaging devices. This sensor technology exhibits higher levels of noise that can manifest as random variations in pixel brightness or colour, especially in low-light conditions [5]. Moreover, our dataset images also contain a loose boundary between objects due to various factors, including image noise, blurring, low image resolution, and similar intensity values between the carapace against the flipper of the turtle as shown in Figure 1. These factors introduce uncertainty and ambiguity, leading to challenges in image segmentation, object detection, and boundary extraction tasks [6] [7].

This research proposed an instance segmentation method of a sea turtle's carapace in an automated manner using a Mask R-CNN-based method. We implemented Contrast-Limited Adaptive Histogram Equalization (CLAHE) [8] method as the data pre-processing step. In many cases, CLAHE improves local contrast while enhancing and preserving details inside the image. Application examples include CT scans [9], retinal fundus [10] and animals [11]. In Section 3, We performed sensitivity analysis on a few internal parameters of CLAHE to show their impacts towards mask segmentation prediction by comparing the segmented results against our manually annotated ground truth.

Previous works have incorporated different image enhancement techniques for data preprocessing with Mask R-CNN. Jiangping et al. [12] addressed the problem of protein crystallization in macromolecular crystallography and the challenges posed by image quality and classification algorithms. In their work, CLAHE was used as a data preprocessing step to enhance the visibility of protein crystals in images. This inclusion improved the classification accuracy of the Mask R-CNN model, resulting in a 42% improvement in mean average precision. However, the network sometimes misidentifies non-crystal spots outside the droplet as crystals, highlighting the need for enhance segmentation accuracy. Naufal et al. [13] addressed the problem of face mask detection under low-light conditions during the Coronavirus disease (COVID-19) pandemic. The low-light condition can make image detection more difficult due to the presence of high-noise images, poor illumination, and reflectance. To overcome this issue, the images need to be processed using two different approaches, such as CLAHE and Gamma Correction, before training the model to provide visual quality by fine-tuning brightness and contrast levels. The methods improved the detection of face masks under low-light conditions by 98.13% accuracy using the Gamma correction method and 97.86% for the CLAHE method. The proposed method shows better result compared with the data trained using pre-trained models such as MobileNetV2, VGG16, and VGG19. However, the paper also suggests finding more alternative methods that can improve performance and classification for low-light images. Xu et al. [14] adopted CLAHE techniques for data preprocessing as a preliminary step in preparing the input data for their tuberculosis detection model. Given that chest radiograph images often exhibit high noise levels and low contrast, CLAHE was utilized to extract essential features for training the model to detect tuberculosis lesions within the images. Consequently, this method effectively enhanced the contrast between pulmonary tuberculosis regions and the background.

Figure 1. Hik Vision Camera used in the data collection and noisy turtle images captured

Unlike previous studies, we have developed an image segmentation model using Mask R-CNN combined with the CLAHE algorithm to segment biometric parts of wildlife animals in extremely low-light conditions at night. Thus, the CLAHE segmentation technique was utilized as a data preprocessing step to overcome the problem of low contrast in the images. This method improves the local contrast visibility of fine features and enhances the overall image clarity. This enhancement is crucial for achieving accurate segmentation with Mask R-CNN, particularly under challenging lighting conditions.

2. Method

To train our dataset with Mask R-CNN, we collected the dataset, annotated it, enhanced the images using CLAHE, and evaluated the mask performance based on Intersection over Union (IoU). This section provides a detailed discussion of the methodology followed.

A. Data Acquisition

For this research, we collected video data at Talang Besar Island, located in Sarawak, Malaysia. Each video data is around 2 - 3 minutes in duration and contains video footage of a slowly moving green sea turtles. The dataset was captured in a low-light environment with no artificial light. The data acquisition was conducted during active nesting periods, from around 10 p.m. until 4 a.m. The full dataset contains a total of 34 individual videos of green sea turtles captured using a downward-facing camera with a resolution of 2560 x 1920 pixels.

B. Data Preparation

The videos were stabilized using Vidstab [15] to reduce blurriness, jittering, and noise. Frames were extracted from each video and converted into JPEG image format. Even after stabilization, some frames remained excessively blurred and needed to be manually removed. LabelMe [16] annotation tools are adopted to label the target objects with the polygon shape to annotate the ground truth for distinguishing the carapace from the background. The region of interest of the target object is labelled as a "carapace" with its key points values which indicate the location of the carapace shape in the image. The annotation data is then exported into a JSON file and will be fed into training along with the images. The data then was split into the train, validation, and test subset that consist of 70, 30, and 30 images, respectively. Figure 2 shows examples of ground truth images with annotations. Rather than using the full dataset, we decided to extract frames from 15 selected videos only. We decided against using the full dataset because the ground truth annotation requires a lot of effort and time. The extracted frames were then used for training, validation, and testing. Table I shows the summary of the data utilized for training, validation, and testing, specifying the distribution across each set.

C. Data Preprocessing

CLAHE is a computer image processing technique used to enhance the contrast and visibility of details in digital images. Two hyperparameters of CLAHE, clip limit, and grid size, need to be tuned to find the optimal contrast of the image. The clip limit of a histogram is defined by the following mathematical formula [17],

$$\beta = \frac{M}{N}\left(1 + \frac{\alpha}{100}(S_{\max} - 1)\right) \tag{1}$$

where, the parameter M indicates the size of the region, N represents the grayscale value (256), and α signifies a clipping factor that determines the inclusion of a histogram, ranging from 0 to 100.

The grid size divides the image into different local tiles or regions for histogram equalization. Each tile undergoes histogram equalization with a predefined clip limit, involving steps such as histogram computation, excess handling, redistribution, and scaling using a cumulative distribution function as described in [8]. Each tile is processed independently to enhance the contrast of that specific region. Additionally, the clip limit refers to the degree of contrast enhancement applied to each local tile of the histogram during the equalization process. It controls the maximum number of allowable amplification factors for contrast in the tiles.

We applied a clip limit value of 5 and various grid sizes in our implementation. Utilizing different grid sizes to experiment with different settings to determine optimal setting when training with Mask R-CNN. Furthermore, we also aim to monitor the impact on model performance based on IoU scores and determine the most effective CLAHE parameters to handle turtle images characterized by intricate patterns or regions exhibiting varying contrast levels. Based on our observation, a clip limit of 5 strikes a good balance between contrast enhancement and preserving the overall spatial data in the image. This setting allows for a noticeable improvement in the visibility and detail of the turtle while avoiding excessive distortion. For visual representation, please refer to Figure 4, which illustrates the turtle images before and after image enhancement techniques with different grid sizes such as 8x8, 16x16, and 32x32. To mitigate bias in this research, we ensured consistency in our annotations across all models. Specifically, we applied the same annotations as those conducted for the original images. The workflow of our experiment is depicted in Figure 3.

D. Mask-RCNN Architecure

The Mask R-CNN [19] is an extended version of Faster R-CNN, which adds another branch for predicting and segmenting masks in its architecture. The mask head layer is an additional feature used to draw and predict the pixel-wise segmentation mask of the turtle carapace from its background. As illustrated in Figure 5, the first stage of Mask R-CNN focuses on object detection and proposes regions of interest. Next, the second stage refines these regions by performing pixel-wise instance segmentation, producing detailed segmentation masks for each detected object in the image. This two-stage architecture allows Mask R-CNN to achieve both precise object localization and accurate pixel-wise segmentation.

The Mask R-CNN used in this research is adopted from Matterport [20]. In this experiment, we tuned a few hyperparameters to leverage the performance of Mask R-CNN. Firstly, we set the value of the maximum ground truth instance and maximum detection instances to 1. This limits one ground truth instance during training and will only detect one instance during inference. This setting could be useful in our scenarios where there is only one object of interest in each image and can reduce the model training time. The confidence level threshold is set to 0.9 to ensure high accuracy detection and minimize false positives, ensuring that the model predicts only instances of the carapace with very high confidence. Moreover, selecting the backbone of Mask R-CNN between ResNet-101 and ResNet50 is crucial because it will influence the trade-off between training time and accuracy. Thus, while ResNet-101 excels in detection and segmentation accuracy, it has a slower training time due to its numerous layers [13]. Since

TABLE I. Summary of data utilized in this research

Dataset	Individual green sea turtles Video	Total Annotation Video
Training	10	70
Validation		30
Test	5	30

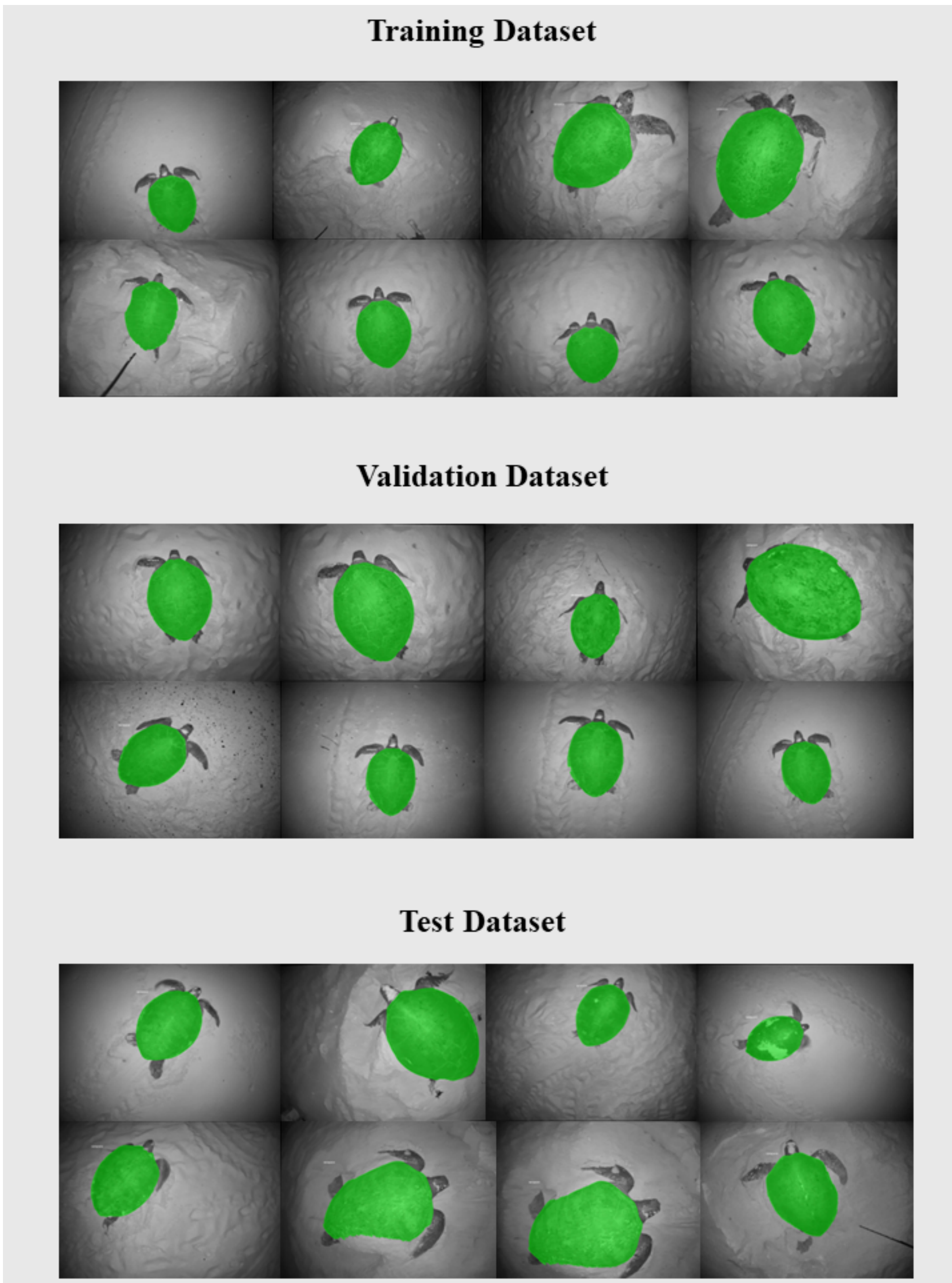


Figure 2. Sample images from training, validation, and test datasets. Best viewed in color

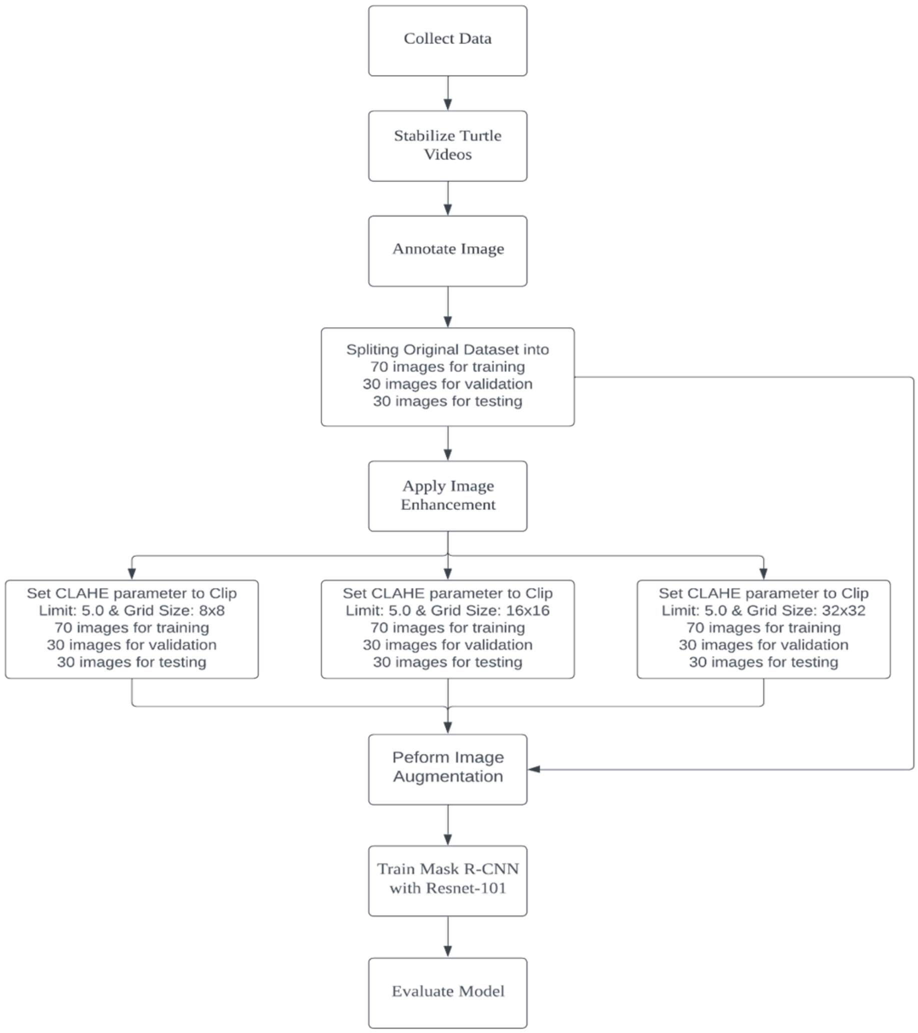


Figure 3. Our research workflow.

we have a small dataset, we employ data augmentation to overcome the limitations posed by the small number of labeled training samples, allowing the model to learn from a more diverse set of examples. Additionally, transfer learning is employed in this research using a pre-trained model with COCO weights trained on 80 classes [21]. The use of a pre-trained model provides advantages to our model, reducing training time, as demonstrated in [4]. Lastly, we also configure the training settings with a learning rate of 0.001, 50 epochs, 100 steps per epoch, and a weight decay of 0.0001.

E. Evaluation Metric

To evaluate the performance of the carapace segmentation, we use official COCO evaluation metrics [22] by calculating the area of overlap between the predicted segmentation and the ground truth. The calculation is as follows,

$$\mathrm{IoU}\,(\mathrm{mask_t}, \mathrm{mask_s}) = \frac{\mathrm{I}\,(\mathrm{mask_t}, \mathrm{mask_s})}{\mathrm{U}\,(\mathrm{mask_t}, \mathrm{mask_s})} \qquad (2)$$

where I ($mask_t$, $mask_s$) signifies the area where the prediction of the carapace intersects with the binary mask of the ground truth object. This value is then divided by U ($mask_t$, $mask_s$), representing the union area of the carapace with the ground truth. The resulting IoU indicates the percentage of overlap between the ground truth and the network prediction of the Mask R-CNN.

In this experiment, we purposely used a different set of videos for the training and validation set versus the test set. Using this approach, the trained Mask R-CNN models are validated using unseen samples. Each trained model underwent testing with test images produced using the same clip limit and grid size value as the training and validation set. This approach ensured a fair assessment of the model's

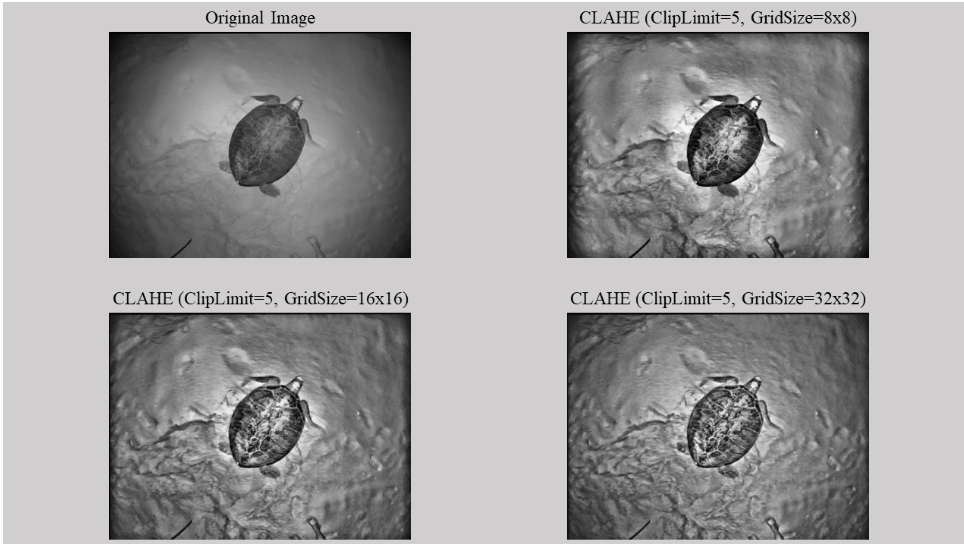


Figure 4. Image of green sea turtles before and after image enhancement with clip limit 5.0 and various grid size.

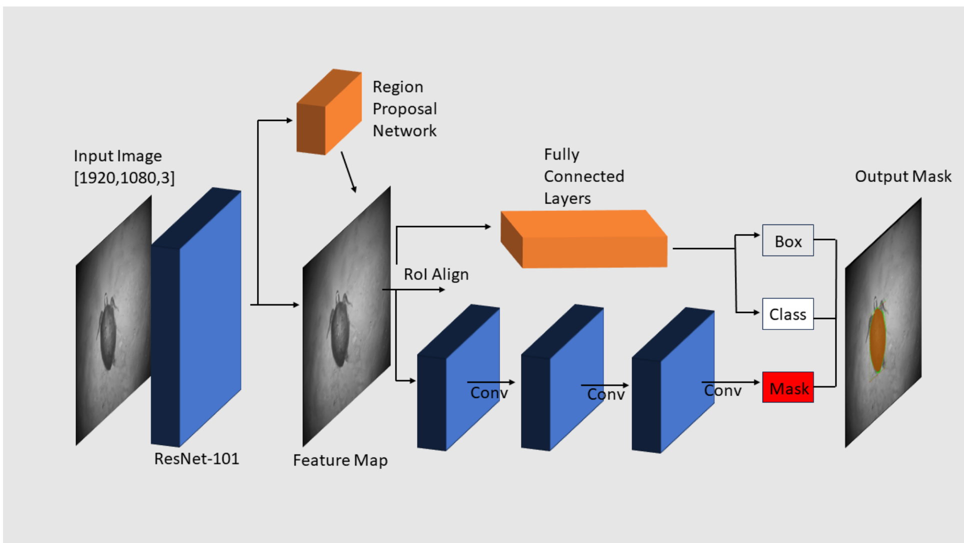


Figure 5. Overview of Mask R-CNN Architecture with ResNet-101 adapted from [18].

performance, as they were tested against images produced using parameters consistent with their training data.

3. Result and Discussion

This research used Mask R-CNN to detect turtles and segment their carapaces for creating a dataset suitable for turtle classification tasks. Consequently, we will evaluate the model's performance trained on four different datasets using IoU to compare the mask predictions to the ground truth. The test set comprises unseen data, which the network has never encountered during training. It consists of unique images of individual turtles recorded separately. We will conduct tests using 30 turtle images to assess the accuracy of turtle segmentation in various positions and conditions. The test dataset comprises two types of images: one featuring clear turtles with visible carapaces, and the other presenting a significant challenge where heavy occlusion by sand covers the top and sides of the carapace, distorting its intrinsic shape. Figure 6 illustrates these distinct characteristics of the turtle images used for testing.

A. *Quantitative Evaluation of Mask R-CNN Trained with 4 different datasets*

Table II presents the results of four different models based on Mask IoU over 30 images. This metric is used to match each predicted mask with the ground truth and assess the quality of segmentation. A higher IoU indicates better performance in detecting and capturing the entire area of the carapace scute pattern.

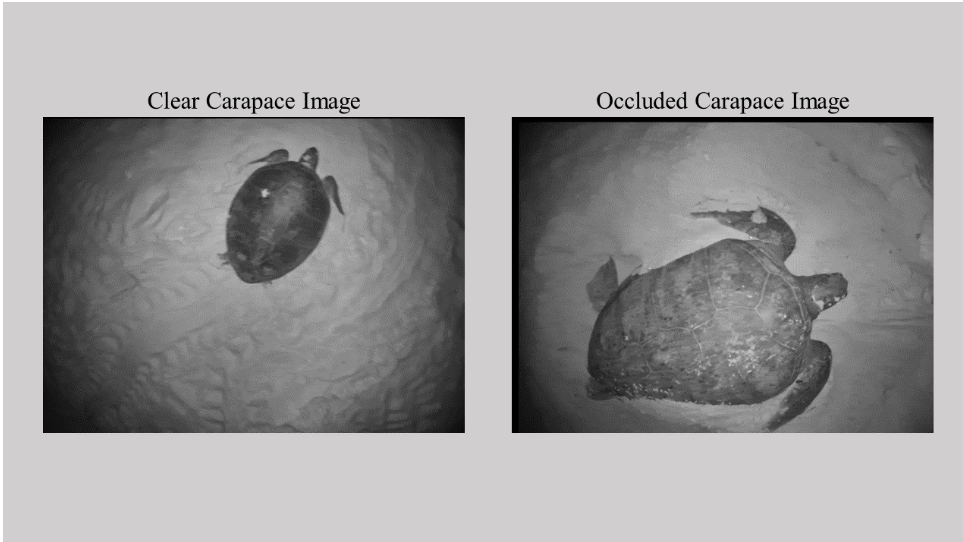


Figure 6. Sample of different characteristics of test images.

TABLE II. Quantitative Evaluation of Mask R-CNN Trained with Different Datasets

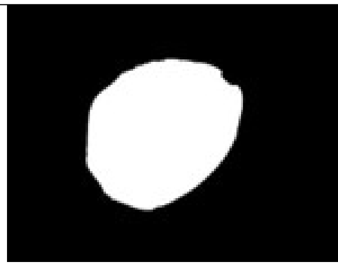
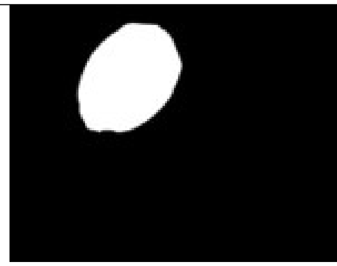
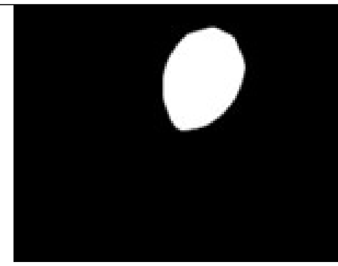
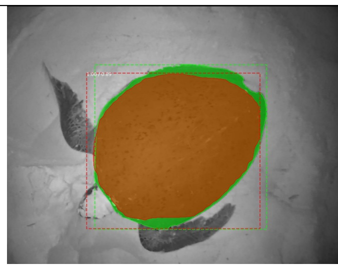
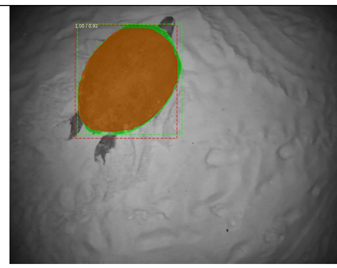
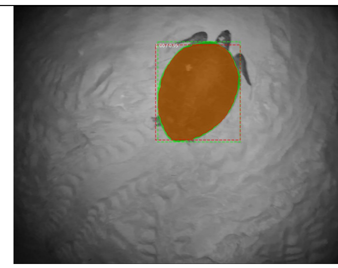
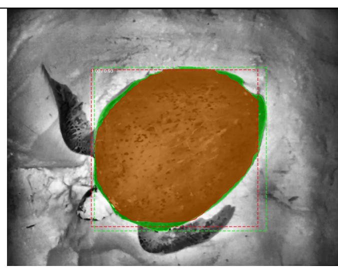
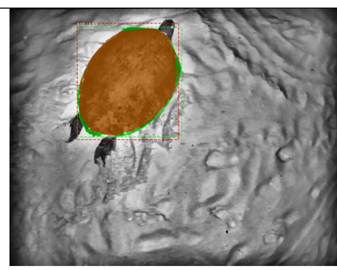
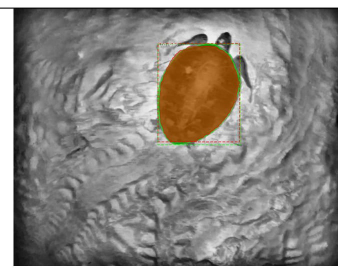
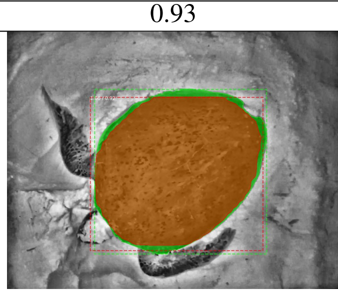
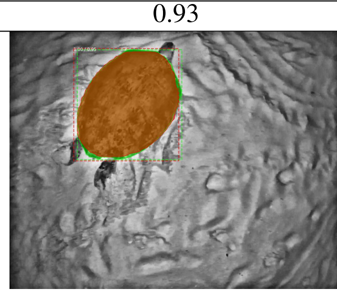
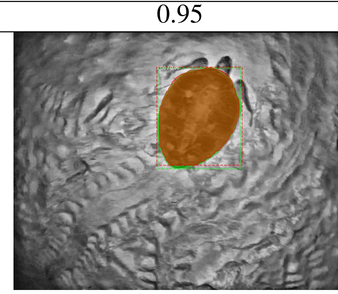
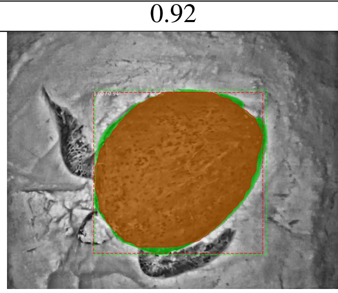
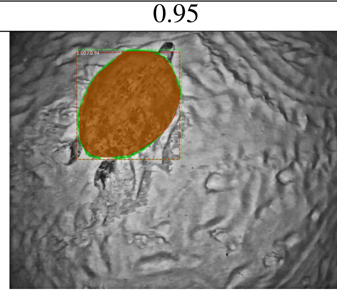
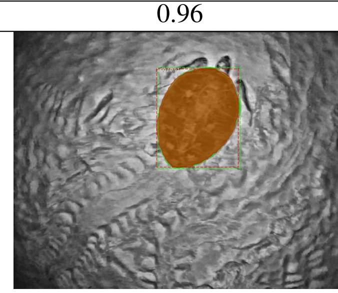
Model	Trained on	Number of Test Images	Mask IoU(%)
A	Original Dataset	30	91.80
B	CLAHE 8x8	30	92.69
C	CLAHE 16x16	30	93.10
D	CLAHE 32x32	30	93.35

Based on the results presented in Table II, the overall performance of Mask R-CNN with ResNet-101, trained using preprocess dataset with CLAHE, regardless of the grid size, demonstrates a positive impact on the Mask IoU scores compared to the original dataset. Model A achieves a commendable 91.80% IoU for segmentation, while the best-performing model for predicting Mask IoU scores is Model D, which achieves 93.35%. This indicates an improvement of 1.55% in overall performance. Model A may have struggled to predict the target object's boundary with the background because of the similar intensity of contrast between them. It shows that the image enhancement technique using CLAHE is capable of enhancing identification and localization [23]. Moreover, the enhanced images produced by CLAHE exhibit superior appearance in local contrast, which enhances the feature details. This allows the feature extractor to generate more detailed and accurate feature extractions for neural networks to learn, as stated in [24], and these features are lacking in the original dataset. Based on the results obtained, we can see that the grid size of 16x16 and 32x32 for CLAHE is more effective in enhancing the model's segmentation performance than 8x8 grid size. This is due to the higher grid size, which has shown some improvement in distinguishing features from noise effects and better object separation of carapace and the background, resulting in more distinct masks for individual objects, making the model easier to segment.

B. *Qualitative Evaluation of Mask R-CNN trained with 4 different datasets*

Table III shows qualitative examples of turtle carapace predictions made based on 4 models on different turtle images of different IDs. It is noted that the images of the turtles are taken from different videos, individuals, sizes of the carapace, and various positions in the frame. Based on the table, the green mask represents the ground truth of the carapace, while the orange mask represents the predicted pixel segmentation that overlaps with the ground truth. As shown in image ID Test 24, 17, and 30 below, we can observe that the green mask or ground truth can be identified more in the original dataset, indicating less overlapping of prediction, such as in image ID Test 24, compared to the dataset trained with CLAHE. Higher Mask IoU values will provide better segmentation quality. Therefore, all models demonstrate high segmentation Mask IoU scores, with the lowest prediction achieved by Model A at 0.89 for image ID Test 24, and the highest prediction achieved by Model D for image ID Test 30, with a score of 0.96, indicating significant overlap with the ground truth. Based on the obtained results, we can conclude that a higher-quality image of the target object, in terms of clear edges, pixel intensity, and boundaries, can provide more valuable information for mask segmentation head to accurately predict the carapace.

TABLE III. Visualization of image prediction on 4 models. Best viewed in color.

Image ID	Test 24	Test 17	Test 30
GT			
Model A			
IoU	0.89	0.92	0.95
Model B			
IoU	0.93	0.93	0.95
Model C			
IoU	0.92	0.95	0.96
Model D			
IoU	0.94	0.94	0.97

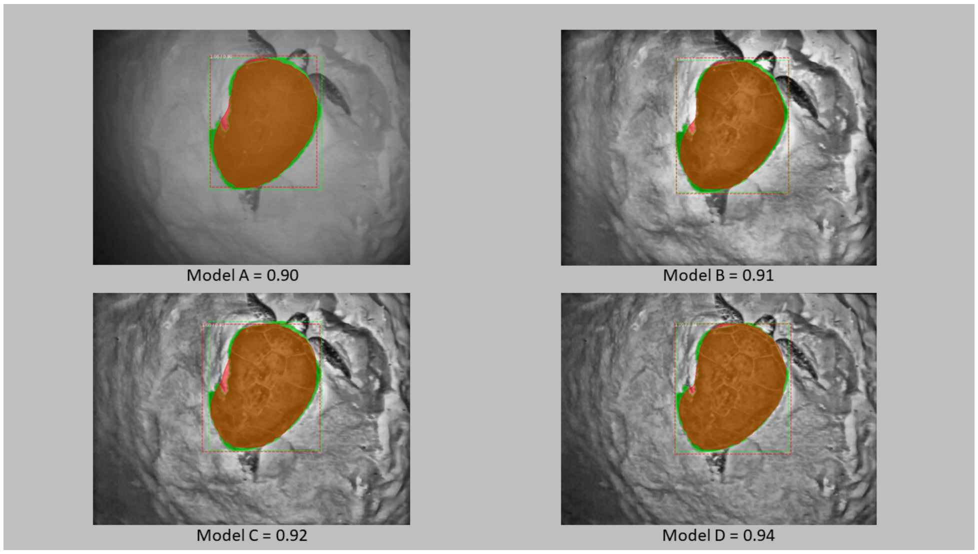


Figure 7. Prediction on failure cases. Best viewed in color.

C. Discussion on Failure Cases

Despite the high segmentation results achieved by the models, they encountered difficulties in specific image scenarios, as observed in Image ID Test 19, depicted in Figure 7. In this scenario, all models mistakenly predict the sand as part of the carapace. There are several reasons why a Mask R-CNN model may make false detection during predictions. Mask R-CNN relies on objects' visual appearance and context for making predictions. If the background contains elements or textures that closely resemble the target object, our model may erroneously detect and segment those background elements as part of the target object, which can result in false positive detection [25]. Another contributing factor could be the limitations of the training data used with these types of images [26]. The quality and diversity of the training dataset can significantly impact the model's ability to distinguish objects from the background. Therefore, we can conclude that if the training dataset lacks examples of objects in specific poses or under varying conditions, the model may encounter challenges in effectively segmenting objects.

4. Conclusion

In conclusion, employing CLAHE to enhance the contrast of green sea turtle images shows some improvement in terms of mask IoU, with an increase of 1.55% compared to using Mask R-CNN alone. This indicates that CLAHE can tackle the problems of noisy images, blurring, low image resolution, and similar intensity values between adjacent regions. Experimenting with different grid sizes has impacted the performance of Mask R-CNN in segmenting the target object. A higher grid size of 32x32 shows that the model can perform better with a 93.5% Mask IoU. However, all Mask R-CNN models trained with CLAHE datasets can surpass the original dataset. Although the models show excellent segmentation with higher mask IoU, we also identified some prediction errors, and the segmentation of all models does not adequately cover the edges of the carapace. The future work of this research involves enhancing the smoothness of the segmentation area and reducing false detection in challenging images. This can be achieved by employing image enhancement methods such as the Laplacian filter, to detect the edges of the green sea turtle's carapace and enhance the visibility of features along the edges.

5. Acknowledgement

The authors wish to express their gratitude for the financial support provided by the Ministry of Higher Education Malaysia through the Fundamental Research Grant Scheme (FRGS) under the reference FRGS/1/2021/ICT02/UNIMAS/02/1. The authors are sincerely grateful of the grant providers for their support in our research endeavours. Additionally, we extend our heartfelt thanks to the dedicated personnel and wardens at Sarawak Forestry Corporation, who not only granted us access but also offered physical and technical assistance for the wildlife fieldwork aspects of this research.Thank you as well to Universiti Malaysia Sarawak, UNIMAS, for the necessary support provided during this research.

References

[1] K. L. McNally, C. R. Mott, J. R. Guertin, and J. L. Bowen, "Microbial communities of wild-captured kemp's ridley (lepidochelys kempii) and green sea turtles (chelonia mydas)," *Endangered Species Research*, vol. 45, pp. 21–36, 2021.

[2] I. Hipiny, H. Ujir, A. Mujahid, and N. K. Yahya, "Towards automated biometric identification of sea turtles (chelonia mydas)," *arXiv preprint arXiv:1909.11277*, 2019.

[3] S. Beery, G. Van Horn, and P. Perona, "Recognition in terra incognita," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 456–473.

[4] C. Yu, X. Fan, Z. Hu, X. Xia, Y. Zhao, R. Li, and Y. Bai, "Segmentation and measurement scheme for fish morphological features based on mask r-cnn," *Information Processing in Agriculture*, vol. 7, no. 4, pp. 523–534, 2020.

[5] J. H. Chuah and D. Holburn, "Design of low-noise high-gain cmos transimpedance amplifier for intelligent sensing of secondary electrons," *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5997–6004, 2015.

[6] Y. Li, X. Xu, and C. Yuan, "Enhanced mask r-cnn for chinese food image detection," *Mathematical Problems in Engineering*, vol. 2020, pp. 1–8, 2020.

[7] S. Song, J. Zhu, X. Li, and Q. Huang, "Integrate msrcr and mask r-cnn to recognize underwater creatures on small sample datasets," *IEEE Access*, vol. 8, pp. 172 848–172 858, 2020.

[8] A. M. Reza, "Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 38, pp. 35–44, 2004.

[9] S. Sanagavarapu, S. Sridhar, and T. Gopal, "Covid-19 identification in clahe enhanced ct scans with class imbalance using ensembled resnets," in *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. IEEE, 2021, pp. 1–7.

[10] J. C. M. dos Santos, G. A. Carrijo, C. de Fátima dos Santos Cardoso, J. C. Ferreira, P. M. Sousa, and A. C. Patrocínio, "Fundus image quality enhancement for blood vessel detection via a neural network using clahe and wiener filter," *Research on biomedical engineering*, vol. 36, pp. 107–119, 2020.

[11] A. Abuduweili, X. Wu, and X. Tao, "Efficient method for categorize animals in the wild," *arXiv preprint arXiv:1907.13037*, 2019.

[12] J. Qin, Y. Zhang, H. Zhou, F. Yu, B. Sun, and Q. Wang, "Protein crystal instance segmentation based on mask r-cnn," *Crystals*, vol. 11, no. 2, p. 157, 2021.

[13] N. M. Athif, F. Sthevanie, and K. N. Ramadhan, "Face mask detection under low light condition using convolutional neural network (cnn)," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 8, no. 1, pp. 281–290, 2023.

[14] X. Xu, J. Zhang, X. Cheng, L. Lu, Y. Zhao, Z. Xu, and Z. Gu, "A tuberculosis detection method using attention and sparse r-cnn," *KSII Transactions on Internet and Information Systems*, vol. 16, no. 7, pp. 2131–2153, 2022.

[15] A. Spannbauer, "Vidstab," *Python. org*, 2020.

[16] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "Labelme: a database and web-based tool for image annotation," *International journal of computer vision*, vol. 77, pp. 157–173, 2008.

[17] F. Hana and I. Maulida, "Analysis of contrast limited adaptive histogram equalization (clahe) parameters on finger knuckle print identification," in *Journal of Physics: Conference Series*, vol. 1764, no. 1. IOP Publishing, 2021, p. 012049.

[18] S. Podder, S. Bhattacharjee, and A. Roy, "An efficient method of detection of covid-19 using mask r-cnn on chest x-ray images," *AIMS Biophys*, vol. 8, no. 3, pp. 281–290, 2021.

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.

[20] W. Abdulla, "Mask r-cnn for object detection and instance segmentation on keras and tensorflow," 2023 2017. [Online]. Available: https://github.com/matterport/Mask_RCNN

[21] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*. Springer, 2014, pp. 740–755.

[22] X. Chen, H. Fang, T.-Y. Lin, R. Vedantam, S. Gupta, P. Dollar, and C. L. Zitnick, "Microsoft coco captions: Data collection and evaluation server," *arXiv e-prints*, pp. arXiv–1504, 2015.

[23] A. Fatima, I. Shafi, H. Afzal, K. Mahmood, I. d. l. T. Díez, V. Lipari, J. B. Ballester, and I. Ashraf, "Deep learning-based multiclass instance segmentation for dental lesion detection," in *Healthcare*, vol. 11, no. 3. MDPI, 2023, p. 347.

[24] S. Nan, "Feature extraction and segmentation processing of images based on convolutional neural networks," *Optical Memory and Neural Networks*, vol. 30, pp. 67–73, 2021.

[25] P. M. Blok, F. K. van Evert, A. P. Tielen, E. J. van Henten, and G. Kootstra, "The effect of data augmentation and network simplification on the image-based detection of broccoli heads with mask r-cnn," *Journal of Field Robotics*, vol. 38, no. 1, pp. 85–104, 2021.

[26] Z. Huang, L. Huang, Y. Gong, C. Huang, and X. Wang, "Mask scoring r-cnn," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 6409–6418.

Mohamad Syahiran Soria holds a BSc (2021) in Computer Science (Software Engineering) from Universiti Malaysia Sarawak (UNIMAS) with his final year thesis titled "KamekTravel Application System using Hybrid Mobile App Development, Ionic Framework". He is currently pursuing his master education at the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak (UNIMAS). His research areas of interest include Artificial Intelligence and Computer Vision. He can be contacted at email: 22020014@siswa.unimas.my

Khalif Amir Zakry holds a BSc (2020) in Computer Science (Software Engineering) from Universiti Malaysia Sarawak (UNIMAS) with his final year thesis titled “Contract Distribution System Framework with Case Study on PC Building Market”. He was a system engineer for Sarawak Information Systems, working specifically in the field of audiovisual engineering. He is currently pursuing his master education at the Faculty of Computer Science and Information Technology, Universiti Malaysia Sarawak (UNIMAS). His research areas of interest include Artificial Intelligence and Computer Vision. He can be contacted at email: 21020463@siswa.unimas.my

Dr. Irwandi Hipni Mohamad Hipiny is currently a Senior Lecturer (DS52) at Universiti Malaysia Sarawak. He holds a BSc (2003) and an MSc (2007) in Computer Science from Universiti Teknologi Malaysia; later graduated with a PhD (2014) in Computer Vision from University of Bristol, UK. Prior to his current appointment, Irwandi was a recipient of Ekpress UTM-MARA scholarship (1999 – 2003), and a MOSTI’s National Science fellow (2003 - 2007). Irwandi’s current research interests are, but not limited to, Computer vision, Pattern recognition and Animal Re-ID. He can be contacted at email: mhihipni@unimas.my

Assoc. Prof. Ts. Dr. Hamimah Ujir is an Associate Professor at the Faculty of Computer Science and Information Technology (FCSIT), Universiti Malaysia Sarawak (UNIMAS). In 2003, she received a National Science Fellowship scholarship award from the Ministry of Science, Technology, and Innovation (MOSTI) and completed her master’s degree (by research) at the Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, Skudai, Johor. In March 2009, she pursued her Ph.D. in the School of Electronic, Electrical and Computer Engineering, University of Birmingham, United Kingdom. Her previous and current works include 3D physical simulation and 3D static and dynamic facial expression analysis. She also conducted research on academic quality in higher education. She can be contacted at email: uhamimah@unimas.my

Professor Dr. Ruhana Binti Hassan Graduated with a Bachelors Degree in Aquatic Biology from Universiti Sains Malaysia in 1993 and attained her Masters Degree of Science in Biology from the same university several years later. She received her Doctorate in Biological Sciences from the University of Birmingham in 2004. She has been a PhD supervisor for several students (5) with one student currently ongoing. She has won medals in various research expos such as UNIMAS INTEX, Malaysian Technology Expo (MTE) and Pre-University Malaysia Innovation Competition (PIITRAM). She can be contacted at email: hruhana@unimas.my

Alphonsus Ligori Jerry Former Park Warden of Talang Satang National Park (2012-2023) under Sarawak Forestry Corporation. His responsibilities as a Park Warden included managing Talang-Satang National Park in relation to sea turtle conservation in Sarawak and tourism based on conservation. With over 10 years of experience under Sarawak Forestry Corporation. He can be contacted at email: alphonsusl@sarawakforestry.com