

Evaluation of ResNet Network for Semantic Segmentation of Coral Reefs

Šejla Džakmić¹, Azra Yildiz¹, Ali Almisreb¹, Nursuriati Jamil²

¹Faculty of Engineering and Natural Sciences, International University of Sarajevo, 71210 Ilidža Sarajevo, Bosnia and Herzegovina, alimes96@yahoo.com

²Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, Malaysia, lizajamil@computer.org

ABSTRACT

The Deep Residual Network (ResNet) learning model is known to achieve better accuracy and requiring shorter training time compared to other pre-trained learning models for image classifications and recognition. In this paper, the use of ResNet networks for semantic segmentation of coral reefs images was explored. Three ResNet networks (ResNet-18, ResNet-50, and ResNet-101) were evaluated and compared using 900 images as training dataset and 38 images as test dataset. The last three layers of the pre-trained ResNets were replaced with a set of layers that classified each pixel of the images into four classes: 'dead', 'alive', 'sand' and 'unknown'. A Softmax layer was introduced to reduce the imbalanced defects. Then, DeepLabv3+ employed the Atrous convolution to extract the features computed by applied CNN and segment the pixels of the object. ResNet-101 was shown to achieve better results compared to ResNet-18 and ResNet-50. Further analysis of the results implied that the class weightage assignment needs to be improved and more a larger training dataset should be acquired.

Key words : Semantic segmentation, Deep Residual network, class balancing, coral reef.

1. INTRODUCTION

Coral reefs are among the most biologically diverse ecosystems on Earth. Besides providing humanity with fisheries, they also served as a coastal protection and contributed to carbon sequestration that reduces global warming [1,2]. During the last few decades, coral reef ecosystems are being exposed to massive and rapid decline caused locally by overfishing, pollution, deforestation, and sedimentation. Global issues were caused mostly by global climate changes, including rising sea level and seawater temperature, and ocean acidification leading to a state of marine environmental crisis. Statistical data provided in [3] showed that in 2011, 19% of coral reefs had disappeared, while 75% were endangered. Therefore, it is essential to determine how both local and global stressors affect them and how to reduce degradation by providing solutions to sustain coral reefs. Manual mapping of underwater environments had been done for field survey and coral reef monitoring of the

ecosystems. However, research and practical application results showed that there are many deficiencies in terms of relevance of the outcomes [4]. Besides, manual mapping of underwater environments is very time consuming, while aerial photography and satellite remote readings are inapplicable due to light absorbance properties of seawater [5]. With the advancement of autonomous underwater vehicles (AUV) that can capture high-resolutions images, underwater sea mapping is now done using video-based robotic surveys. Computer visions and image processing techniques were then used to understand the distribution of the coral taxa such as hard corals, soft corals, and algae. However, there are many difficulties in recognizing objects in underwater sea images due to color variations because of depth [6] dirt and sediments, and geometric variations of the objects. Therefore, semantic segmentation prior to recognition, annotation or detection is extremely critical for accurate results.

In this paper, a comparison of three coral reef semantic segmentation results employing the ResNet-18, ResNet-50 and ResNet-101 architectures was presented. All three architectures were applied on the same dataset, with hand-labeled images. To capture the contextual information at multiple scales, DeepLabv3+ model with atrous convolution with the encoder-decoder structure was applied, which allows us to effectively expand the field of view of the filters. DeepLabv3+ involved important semantic segmentation information from the encoder module. However, a detailed information of the object boundaries was missing because of the pooling or striding convolutions within the network. This effect can be mitigated by atrous convolution with a decoder module which can extract heavy feature maps and recover a detailed object boundary [7]. Encoder-decoder module [8] contributed the faster computation gradually recovering the boundaries of the sharp objects. The performance of the suggested methods was measured by pixel intersection-over-union averaged across four (4) classes, accuracy, and boundary F1.

2. RELATED WORK

Semantic segmentation is one of the computer vision methods applied to select semantic labels to each pixel in the image. A lot of work has been done in image segmentation, applying

different approaches and techniques [9] and [10]. The authors from [11] described two main problems of semantic segmentation, intra-class inconsistency and indistinct inter-class. To overcome these issues, they proposed a Discriminative Feature Network with two sub-networks, Smooth Network and Border Network. Atrous convolution for semantic segmentation was proposed by [9] [13]. In [12], Dense Atrous Spatial Pyramid Pooling (DenseASPP) was applied to generate multiscale features, covering larger scale range densely without significant impact on model size. Convolution of Atrous Spatial Pyramid Pooling and decoder modules was applied in [9] and [7], resulting in an encoder-decoder network with improved performances in terms of speed and strength. Besides the convolution of upsampled filters, ASPP was proposed to robustly segment objects at multiple scales and improve the localization of object boundaries, combining few methods of DCNNs and probabilistic graphical models.

Pixel-wise parsing for deep image segmentation of coral reef was evaluated across different classes of substrate in [13]. Support vector machine classifiers [14] applied on distinct datasets indicated the importance of choosing the right CNN architecture to perform better classification. A comparison of deep learning methods, patch-based CNN and fully convolutional neural network (FCNN) for semantic segmentation of coral reef survey images was described in [15]. It was shown that CNN methods performed better than Support Vector Machine-based classifiers, considering texture-based properties. Among five different deep learning CNN architectures, Resnet-152 performed the best on the labeled dataset of underwater coral reef images. The best results of FCNN were obtained with Deeplab v2 architecture. In a recent effort to automate coral identification in [16], the authors presented a model for accurate coral reef detection of underwater images and compared its performance to human abilities in terms of speed and accuracy. The accuracy of identification was about 95% higher than reached by humans. The proposed model was also able to identify coral reefs on blurry images or partially hidden reefs on clear images. Advanced deep learning tools for improvement of automatic analysis of AUV imagery was done in [17]. ResNet and VGG-net were applied to extract the features of corals and non-corals, followed by a classification process. The results indicated that the efficiency of automatic annotation of unlabeled sections may be improved when combined with AUV image analysis. One more method for training dense segmentation models was proposed in [18] based on ground truth labels allowing to effectively learn the coral segmentation.

The use of Deep Residual networks (ResNets) in computer vision [19, 20], speech recognition, bioinformatics, medical image analysis [21], natural language processing, object detection and recognition [22, 23, 24] were popular due to its ability to enhance training in terms of complexity and speed. These networks also attained better accuracy compared to

other neural networks [25, 26]. The authors from [27, 28] also stated that compared to the AlexNet and VGG16, ResNet-18 is deeper, having residual blocks that performed better than regular convolutional blocks. Additionally, AlexNet, VGG, ResNet family, DenseNet family, CorNet-S and CorNet-Z were applied as tested convolutional neural networks, with AlexNet as the baseline. The best layers of pre-trained networks were submitted to Algonauts challenge to check for Spearman correlation percentage. Accuracy and loss were evaluated by mean intersection-over-union (mIoU). In [29], end-to-end image segmentation with deep learning convolutional neural networks was applied for tongue segmentation. High accuracy was obtained by ResNet, simultaneously increasing the segmentation speed. A comparison of the performance of Inception-v3, ResNet-50 and ResNet-101 using multiple hardware platforms such as CPU and GPU was explained in [25]. Data training was obtained in TensorFlow indicating that the performance was significantly improved when the number of iterations increased, and the network became deeper. The shortest training time was required when ResNet-50 was applied as indicated by many literatures. A high error rate of 34.68% was achieved in [26], for 70,000 iterations of the training process. It was assumed that accuracy would be higher if the training data had higher resolution, since the size of the images were only 64x64 pixels. Apart from the image size, the size of the dataset also affected the training process and results. The impact of noisy labels on the learning process was described in [30]. Using the JFT-300M dataset with 375M noisy labels for 300M images, it was proved that the performance of computer vision methods increased logarithmically based on the volume of training data. Higher-scale dataset can boost the learning process and significantly improve the results.

3. METHODOLOGY

Transfer learning was done by re-tuning the pre-trained ResNet models. The last three layers of pre-trained ResNets were replaced with a set of layers that classified the coral reefs pixels into four classes. A Softmax layer was introduced to reduce the imbalanced defects. Then, DeepLabv3+ employed the atrous convolution to extract the features computed by applied CNN and perform the pixel-wise segmentation. A detailed methodology is described in the following sections.

3.1 Data Collections and Pre-Processing

In this paper, a collection of 128 coral images was captured from high-resolution videos of coral reefs located in the East Coast of Peninsular Malaysia. The images were then labelled. The labelled images served as the groundtruth dataset for the evaluations of semantic segmentation. Even though hand-labelling is challenging and time-consuming, it is more accurate as can be seen in Figure 1 where the results of hand-labelling and using automated labeling is compared. The pixels are labelled into four (4) classes as follows: green for alive, brown for dead, yellow for sand, and black for the unknown.

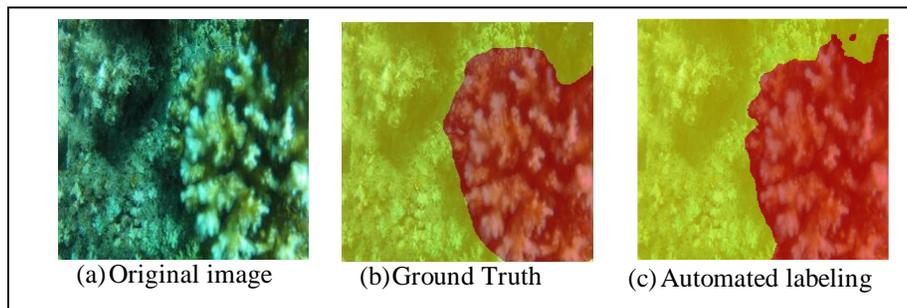


Figure 1: Comparisons of Ground Truth and Automated Labeling of the Coral Image

In ideal cases, all the classes should have approximately equal number of observations. Like many image datasets, it is very expensive in terms of time and labor to acquire balanced datasets. For the datasets, the ‘dead’ class has the highest number of observations, followed by ‘alive’ class. ‘Sand’ and ‘unknown’ classes have the least number of observations as can be seen in Figure 2. Therefore, a class weighting was applied by using the median pixel frequencies calculated by the labeled pixel counts. The class weights for each corresponding class is shown in Table 1. Class ‘unknown’ has the highest weightage to balance its low number of observations, while class ‘alive’ was assigned the lowest weightage.

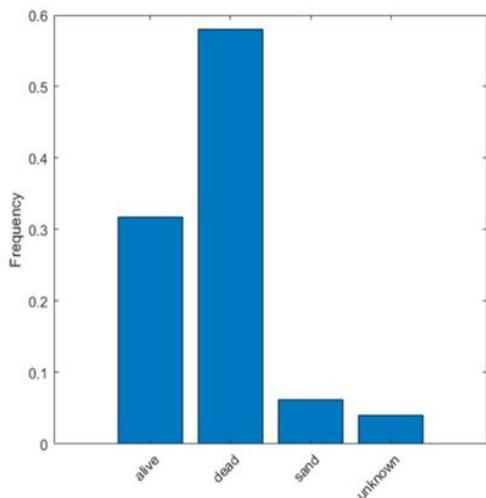


Figure 2: Class Observations

Table 1: Class Weightage Assignment

Class Name	‘alive’	‘dead’	‘sand’	‘unknown’
Class weight	0.5099	0.5774	3.7310	5.2552

The next step is to divide the images into training and testing datasets. Since 70:30 percent ratio for the division of dataset was used, 90 images were used as training, while 38 images were used for testing and datasets. Having a large dataset is crucial to train a deep learning model. Therefore, data

augmentation was done to increase the training dataset and improved the accuracy of the network. For each of the training images, 10 augmented images were created totaling to 900 images as training datasets. In this work, data augmentation was done by initially modifying the contrast, brightness and saturation of the images. Then, affine transformation was applied to the training images that were random scaling by a scale factor ranging from 0.8 to 1.5, horizontal reflection, random rotation using left/right reflection and X/Y translation of +/- 10 pixels. After augmentation, all the images’ size was reduced to 250x250. DeepLabv3+ was employed to create a network based on ResNet-18, 50 and 101 models.

3.2 ResNet Convolutional Neural Network

ResNet-18 consists of 17 convolutional layers and 1 fully-connected layer, all constructed into 5 convolution blocks. The first convolution block is composed of a single convolution layer with 64 filters of size 7x7. The following convolution blocks consist of 2 residual blocks, each consisting of two convolution layers with the same number of filters. Each filter is of 3x3 dimension. Each convolution block scaled down the output image size by 2 and doubled the number of filters (feature dimension). The fully-connected layer connects to 1000 classes and the rest of the network is considered as a feature extractor. For this work, the fully-connected layer to the 4 classes was reduced. A detailed architecture of ResNet models applied in this research can be found in [25].

Training process started with Feature Extraction Layer (FEL). The FEL is located at different layers of ResNet-18, ResNet-50 and ResNet-101. After locating the FEL which is prior to the classification layer, all the layers after the FEL were removed. the ResNet network was customized by removing three different layers. This step was repeated throughout the whole network, prior to each classification layer. In DeepLab v3+, the downsampling factor was typically reduced to either 16 or 8. Atrous convolution was applied to dilated convolutions to recover large receptive fields lost by removing downsampling layers. Atrous convolution was introduced in the network and described in the next section.

Then, skip layer was defined for each ResNet model. The decoder sub-network for DeepLabv3+ that restored the feature maps to their original resolution were added. Upsampling of the original resolution was done by a factor of 16. Finally, softmax and pixel classification layers were applied to classify each pixel of the images. The fully convolutional layer was replaced to classify four (4) classes instead of the pre-defined 1000 classes. Table 2 shows the modification to the layers' configurations.

Table 2: Configurations of Modified Layers

	ResNet018	ResNet-50	ResNet101
Feature Extraction layer	'res5b_relu'	'activation_49_relu'	'res5c_relu'
Removed layers	'pool5', 'fc1000', 'prob', 'Classification_Layer_prediction'	'avg_pool', 'fc1000', 'fc1000_softmax', 'Classification_Layer_fc1000'	'pool5', 'fc1000', 'prob', 'Classification_Layer_prediction'
Update image size	1000 × 1000	500 × 500	500 × 500
Layers with reduced samples	'res5a_branch1', 'res5b_branch2a'	'res5a_branch1', 'res5b_branch2a'	'res5a_branch1', 'res5b_branch2a'
Skip layer	'res2b_relu'	'activation_10_relu'	'res5b_relu'

3.3 Atrous Convolution and DeepLabv3+

Atrous convolution is a powerful tool that regulates the resolution of features computed by deep convolutional neural networks. It also adapts the field of view of the network filters to capture the multi-scale information [6]. Considering the case of two-dimensional signals, it can be described by Eq. 1.

$$y[i] = \sum_k x[i + r \cdot k] \omega[k] \quad (1)$$

where i is the location of the output feature map y and convolution filter ω , applied over input feature map x . Atrous rate, r determined the stride by which input signal was sampled. Changing the field of view of the network filter was done by modifying the rate value. DeepLabv3+ is a type of convolutional neural network used for semantic image segmentation. This paper introduced three ResNet models (ResNet-18, ResNet-50 and ResNet-101) for the training process, modified with respect to DeepLabv3+ as shown in Table 2 earlier.

3.4 Evaluation of Semantic Segmentation

The performance of semantic segmentation was evaluated using Intersection over Union (IoU). Generally, it measures

the difference between the original dataset with the groundtruth labels. It is also known as the Jaccard Index that measures the performance of the segmentation process, compared to the pixel-wise accuracy. Jaccard index is defined by the vectors: y^* vector of ground truth labels and \bar{y}^* vector of predicted labels, and class c ; given by Eq. 2 [31].

$$J_c(y^*, \bar{y}) = \frac{|\{y^* = c\} \cap \{\bar{y} = c\}|}{|\{y^* = c\} \cup \{\bar{y} = c\}|} \quad (2)$$

For multilabel dataset such as in this work, the Jaccard index was averaged across the classes, yielding the mean IoU [49]. Jaccard index was calculated pixel-wise over the evaluated segmentation dataset. Weighted IoU that is the average IoU of each class was also calculated because the dataset has disproportionate size classes. Other than IoU, other metrics such as accuracy and boundary F1 were also used in this work. Accuracy is the rate of correctly segmented pixels in a class according to the groundtruth. For this work, the Global Accuracy and Mean Accuracy for the datasets were measured. Global Accuracy is the estimate of the percentage of correctly segmented pixels, while Mean Accuracy is the average accuracy of all four classes. The last metric that was used is the boundary F1 (BF) that measures how well the predicted boundary of each class aligns with the actual boundary. The Mean BF score that is the average BF score of all classes in the dataset was calculated.

4. RESULTS AND DISCUSSION

Three Res-Net networks were compared: ResNet-18, ResNet-50 and ResNet-101. One example of the semantic segmentation result is presented in Figure 3. In this figure, the first row presented the original image to be segmented and its corresponding groundtruth image. Semantic segmentation of ResNet-18, ResNet-50 and ResNet-101 are subsequently demonstrated in the second, third and fourth row. The first column shows the result of color-coded semantic segmentation by the respective ResNet networks, where green are pixels classified as 'alive', brown for 'dead', yellow for 'sand' and black for 'unknown' class. In the second column, the segmented image is overlapped with the groundtruth image indicating oversegmentation and undersegmentation that are represented by light green and magenta colors, respectively. Visual inspection of the oversegmentation and undersegmentation suggested that ResNet-18 produced better results compared to ResNet-50 and ResNet-101. The 'alive' and 'dead' classes were also segmented better than 'sand' and unknown. Therefore, further evaluations were done by calculating objective measurements using accuracy, mean of intersection and boundary score as shown in Table 3.

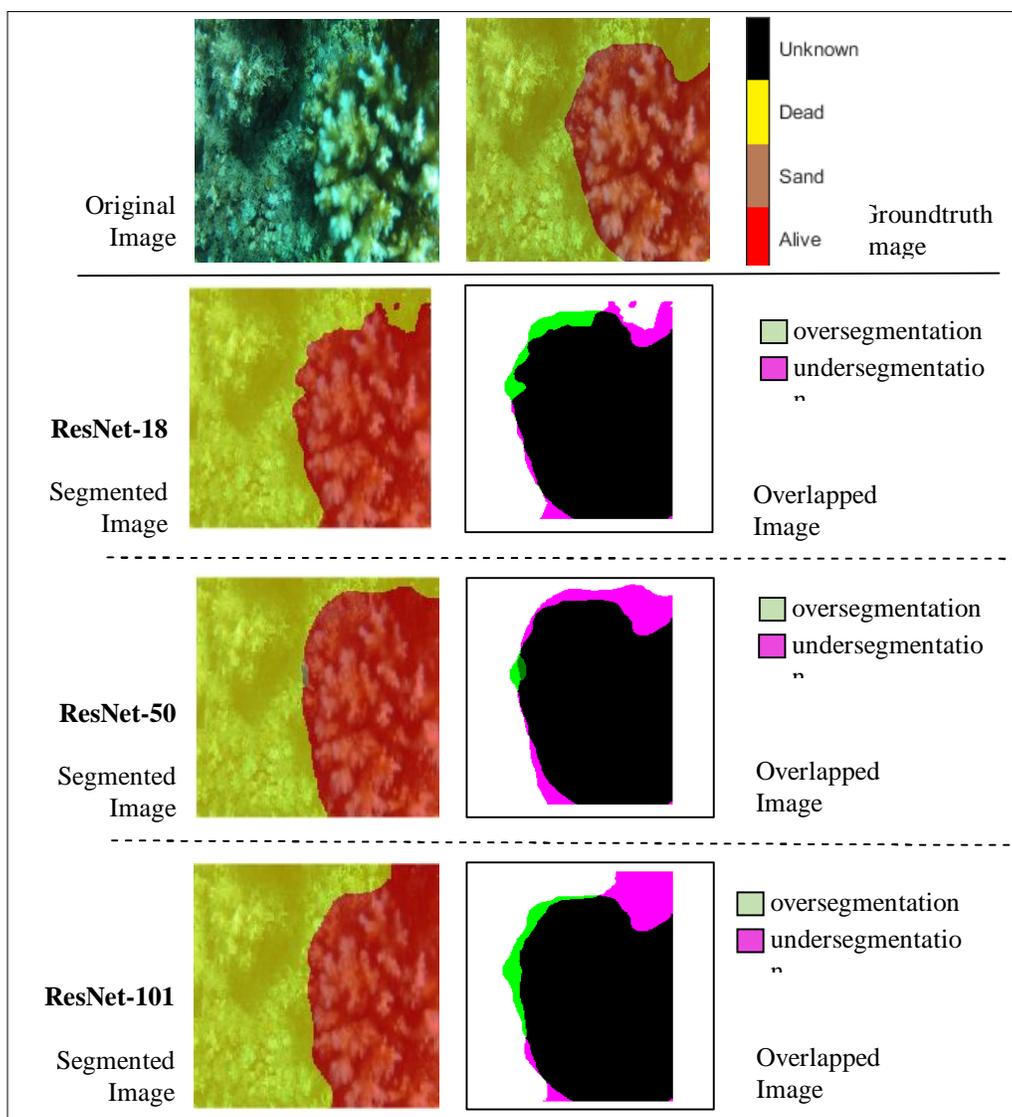


Figure 3: Sample of Semantic Segmentation Results

Based on Table 3, ResNet-18 network outperformed both ResNet-50 and ResNet-110 by achieving 84.22% accuracy, 69.49% IoU and 64.75% meanBFscore. ResNet-101, on the other hand, is slightly better at segmenting the coral images compared to ResNet-50.

Further evaluations of the semantic segmentation results was done by class, and this is presented in Table 4. It shows that class ‘dead’ achieved the highest score for IoU and MeanBFscore for all three ResNet networks. This implies that class ‘dead’ has a high percentage of overlapping pixels of the segmented image and groundtruth, and the predicted boundaries were also aligned with the actual boundaries of the class. While class ‘unknown’ achieved the lowest score for accuracy and IoU for all three networks. Since the number of observations were high for class ‘dead’ and very low for class ‘unknown’, these results were expected. A better class balancing should be done in the future.

Table 3: Objective Measurements of Semantic Segmentation

Models	Classes	IoU
ResNet-18	Alive	0.8912
	Sand	NaN
	Dead	0.9137
	Unknown	NaN
ResNet-50	Alive	0.8679
	Sand	0
	Dead	0.8881
	Unknown	NaN
ResNet-101	Alive	0.8467
	Sand	0
	Dead	0.8719
	Unknown	NaN

Table 4: Semantic Segmentation by Class

	Class	Accuracy	IoU	Mean BFScore
ResNet-18	Alive	0.7851	0.5456	0.2869
	Dead	0.5806	0.4481	0.3110
	Sand	0.3393	0.2656	0.1611
	Unknown	0.0955	0.0523	0.0673
ResNet-50	Alive	0.6539	0.3748	0.1800
	Dead	0.5435	0.4224	0.3054
	Sand	0.1852	0.1447	0.0751
	Unknown	NaN	0	NaN
ResNet-101	Alive	0.8062	0.5798	0.3220
	Dead	0.6330	0.5218	0.3109
	Sand	0.4393	0.5218	0.3109
	Unknown	0.0609	0.0299	0.0594

4.1 Testing Results

The ResNet networks were tested using 38 test images and the results are shown in Table 5. It should be emphasized that the tested images were the same for all the three ResNet networks. Based on Table 5, ResNet-101 network performed best compared to ResNet-18 and ResNet-50. The percentage of correctly segmented pixels for ResNet-101 was 68.32%, followed by ResNet-18 at 63.90% and ResNet-50 at 54.39%. The low accuracy attained by all three networks was due to the small size of the training dataset. The accuracy for each class was even lower as indicated by the mean accuracy at an average of 45% to 48% only. The number of images for each class was disproportionate, and the class weightage balancing using median may not be sufficient. The imbalanced size of each class also affected the weighted IoU, mean IoU and the mean boundary F1 measurements which performed rather poorly.

Table 5: Results of Model Evaluations

	ResNet-18	ResNet-50	ResNet-101
Global Accuracy	0.6390	0.5439	0.6832
Mean Accuracy	0.4501	0.4609	0.4848
Mean IoU	0.3729	0.2355	0.3641
Weighted IoU	0.4660	0.3745	0.5192
Mean BF Score	0.2556	0.1990	0.2816

5. CONCLUSION

In this paper, three Deep Residual networks were evaluated for semantic segmentation of coral reefs images. The aim of the evaluation is to investigate the potential of the ResNet model in semantic segmentation. Based on the results, ResNet-101 outperformed both ResNet-18 and ResNet-50 for pixel-wise segmentation of the coral reefs images. However, more investigations need to be done as the accuracy and Intersection-of-Union metric measurements can be further improved. Since the training datasets are still inadequate, more data should be acquired, and different augmentation processes can be applied. Since high-resolutions images of coral reefs are scarce, the potential of getting disproportionate classes are highly possible. Therefore, future work should also look at advanced techniques of class balancing.

ACKNOWLEDGEMENT

Due acknowledgement is accorded to the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA for the funding received for this publication.

REFERENCES

1. N. S. Foley, T. M. van Rensburg, and C. W. Armstrong. **The ecological and economic value of cold-water coral ecosystems**, *Ocean and Coastal Management*, Vol. 53, No. 7, pp. 313-326, 2010.
2. Conservation International. **Economic values of coral reefs, mangroves, and seagrasses: A global compilation**, Center for Applied Biodiversity Science, Conservation International, Arlington, Virginia, 2008.
3. L. M. Burke, K. Reytar, M. Spalding, and A. Perry. **Reefs at risk revisited**, World Resources Institute, Washington DC, 2017.
4. J. D. Hedley, C. M. Roelfsema, I. Chollett, A. R. Harborne, S. F. Heron, S. Weeks, W. J. Skirving, A. E. Strong, C. M. Eakin, and T. R. Christensen. **Remote sensing of coral reefs for monitoring and anagement: A review**, *Remote Sensing*, Vol. 8, No. 2, pp. 118, 2016.
5. A. King, S. M Bhandarkar, and B. M. Hopkinson. **Deep learning for semantic segmentation of coral reef images using multi-view information**, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, California, June 16-20, 2019, pp. 1-10.
6. M.K.S. Alsmadi, K. Omar, and S.A. Noah. **Fish recognition based on the combination between robust feature selection, image segmentation and geometrical parameter techniques using artificial neural network and decision tree**, *Journal of Computer Science*, Vol. 6, No. 10, pp. 1088-1094, 2009.
7. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schro, and H. Adam. **Encoder-decoder with atrous separable convolution for semantic image segmentation**, in *Proceedings of the European Conference on Computer*

- Vision (ECCV)*, Munich, Germany, September 8-14, 2018, pp. 801-818.
8. M. Chala, B. Nsiri, A. Soulaymani, A. Mokhtari, and B. Benaji. **Deep convolutional networks based on encoder-decode architecture for automatic optic disc segmentation in retina images.** *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 2, March-April 2020, pp. 2078-2084. <https://doi.org/10.30534/ijatcse/2020/181922020>
 9. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. **Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected Crfs,** *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 40, No. 4, pp. 834-848, 2017.
 10. Y. Li, H. Qi, J. Dai, X. Ji, and Y. Wei. **Fully convolutional instance-aware semantic segmentation,** in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, Hawaii, July 21-26, 2017, pp. 2359-2367.
 11. C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang. **Learning a discriminative feature network for semantic segmentation,** in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, June 18-22, 2018, pp. 1857-1866.
 12. M. Yang, K. Yu, C. Zhang, Z. Li, and K. Yang. **Denseaspp for semantic segmentation in street scenes,** in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, Utah, (2018), June 18-22, 2018, pp.3684-3692.
 13. A. Steffens, A. C. de A. Campello, J. Ravenscroft, A. Clark, and H. Hagraš. **Deep segmentation: Using deep convolutional networks for coral reef pixel-wise parsing,** in *Proceedings of the Conference and Labs of the Evaluation Forum*, Lugano, Switzerland, September 9-13, 2019, Working Notes.
 14. F. C. M. Rodrigues, N. S. Hirata, A. A. Abello, T. Leandro, D. La Cruz, R. M. Lopes, and R. Hirata Jr. **Evaluation of transfer learning scenarios in plankton image classification,** in *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, Funchal, Madeira, January 27-29, 2018, pp. 359-366 .
 15. AM. Alqudah, H. Alquraan, IA. Qasmieh, A. Alqudah, and W. Al-Sharu. **Brain tumor classification using deep learning technique-A comparison between cropped, uncropped, and segmented lesion images with different sizes.** *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 8, No. 6, Jan 23, 2020, pp. 3684-3691. <https://doi.org/10.30534/ijatcse/2019/155862019>
 16. A. King, S. M. Bhandarkar, and B. M. Hopkinson. **A comparison of deep learning methods for semantic segmentation of coral reef survey images,** in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, Utah, June 18-22, 2018, pp. 1394-1402.
 17. S. Villon, D. Mouillot, M. Chaumont, E. S. Darling, G. Subsol, T. Claverie, and S. Villéger. **A deep learning method for accurate and fast identification of coral reef fishes in underwater images,** *Ecological Informatics*, Vol. 48, pp. 238-244, 2018.
 18. S. Jaisakthi, P. Mirunalini, and C. Aravindan. **Coral reef annotation and localization using faster R-CNNs,** in *Proceedings of the Conference and Labs of the Evaluation Forum*, Lugano, Switzerland, September 9-13, 2019, Working Notes.
 19. I. Alonso, A. Cambra, A. Munoz, T. Treibitz, and A. C. Murillo. **Coral-segmentation: Training dense labeling models with sparse ground truth,** in *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, October 22-29, 2017, pp. 2874-2882.
 20. Z. Wu, C. Shen, and A. Van Den Hengel. **Wider or deeper: Revisiting the Resnet model for visual recognition,** *Pattern Recognition*, Vol. 90, pp. 119-133, 2019.
 21. A.S. Meghana, S. Sudhakar, G. Arumugam, P. Srinivasan, and KB. Prakash. **Age and gender prediction using Convolution, ResNet50 and Inception ResNetV2,** *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No.2, March-April 2020, pp. 1328-1334. <https://doi.org/10.30534/ijatcse/2020/65922020>
 22. A. Canziani, A. Paszk, and E. Culurciello. **An analysis of deep neural network models for practical applications,** arXiv preprint arXiv:1605.07678, pp. 1-7, 2016.
 23. N. F. P. Setyono, D. Chahyati, and M. I. Fanany. **Betawi traditional food image detection using Resnet and Densenet,** in *Proceedings of 2018 IEEE International Conference on Advanced Computer Science and Information Systems (ICACSIS)* Yogyakarta, Indonesia, October 27-28, 2018. pp. 441-445.
 24. J. Dai, Y. Li, K. He, and J. Sun. **R-FCN: Object detection via Region-Based Fully Convolutional Networks,** in *30th Conference on Neural Information Processing Systems (NIPS 2016)*, Barcelona, Spain, 2016, pp. 379-387.
 25. L. Sun. **Resnet on Tiny Imagenet,** Stanford University, pp. 1-7, 2016.
 26. X. Yu and S.-H. Wang. **Abnormality diagnosis in mammograms by transfer learning based on Resnet18,** *Fundamenta Informaticae*, Vol. 168, No. 2-4, pp. 219-230, 2019.
 27. A. Mahbod, G. Schaefer, I. Ellinger, R. Ecker, A. Pitiot, and C. Wang. **Fusing fine-tuned deep features for skin lesion classification,** *Computerized Medical Imaging and Graphics*, Vol. 71, pp. 19-29, 2019.
 28. H. Jung, M.-K. Choi, J. Jung, J.-H. Lee, S. Kwon, and W. Young Jung. **Resnet-Based vehicle classification and localization in traffic surveillance system in** *Proceedings of the IEEE Conference on Computer*

Vision and Pattern Recognition, Honolulu, Hawaii, July 21-26 2017, pp. 61-67.

29. B. Lin, J. Xie, C. Li, and Y. Qu. **Deeptongue: Tongue segmentation via Resnet**, in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Calgary, Alberta, April 15-20, 2018, pp. 1035-1039.
30. C. Sun, A. Shrivastava, S. Singh, and A. Gupta. **Revisiting unreasonable effectiveness of data in deep learning era**, in *Proceedings IEEE International Conference on Computer Vision*, Venice, Italy, October 22-29, 2017, pp. 843-852.
31. P. Jaccard. **Étude comparative de la distribution florale dans une portion des alpes et des jura**, *Bull Soc Vaudoise Sci Nat*, Vol. 37, pp. 547-579, 1901.