EAI Endorsed Transactions
on Industrial Networks and Intelligent Systems

Research Article EAI.EU

A Fully Convolutional Network with Waterfall Atrous Spatial Pooling and Localized Active Contour Loss for Fish Segmentation

Le Thanh Viet[1], Vu Van Yem[1,*], Van-Truong Pham[1] and Thi-Thao Tran[1,*]

[1]School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Dai Co Viet, Hanoi, Vietnam

Abstract

Accurate measurements and statistics of fish data are important for sustainable development of aqua-enviroment and marine fisheries. For data measurements and statistics, automatic segmentation of fish is one of key tasks. The fish segmentation however is a challenging task due to arterfacts in underwater images. In this study, we introduce a deep-learning approach, namely FCN-WRN-WASP for automatic fish segmentation from the underwater images. In particular, we introduce a computational-efficient variation called Waterfall Atrous Spatial Pooling (WASP) module into a Fully convolutional network with Wide ResNet baseline. We also proposed a loss function inspired from active contour approach that can exploit the local intensity information from the input image. The approach has been validated on the DeepFish data and the SIUM data set. The results are promising for fish segmentation, with higher Intersection over Union (IoU) scores compared to state of the arts. The evaluation results showed that the incorporation of the image based active contour loss helps increase the segmentation performance. In addition, the use of the WASP in the architecture is effective especially for forground fish segmentation.

Keywords: Fully convolutional network, DeepFish, SIUM fish data, Waterfall Atrous Spatial Pooling.

Received on 19 December 2022, accepted on 05 February 2023, published on 20 April 2023

doi: 10.4108/eetinis.v10i1.2942

1. Introduction

It has been proved in much research that monitoring fish in their natural habitat is desperately crucial for sustainable fisheries [1-3]. Effective aqua-culture monitoring can provide information for fish development observation as well as for protection and restoration of fish to maintain healthy fish populations and environmental protection. Traditional fish monitoring tools such as rulers or echo-sounders and manual intervention is laborious and can lead to erroneous. Thus, it is essential to utilize an automatic monitoring system. In fact, the automatic analysis of underwater fish habitats or ecological monitoring system often requires a comprehensive, accurate computer vision system [4].

In a computer vision-based system for fish monitoring, the fish segmentation is a central task, that helps localize the fish for further steps such as fish counting and density estimations. Accurate foreground segmentation could help better analyzing fish counting and fish group behavior. Nevertheless, fish segmentation from underwater images is challenging due to diversity in fish, adversarial water conditions, high similarity of the appearance between fish and some elements in the background such as rocks, and occlusions between fish. Therefore, classical image segmentation methods such as region growing, fuzzy clustering [5], level set models [6] cannot handle the fish segmentation task from underwater images.

Recently, with the development of deep learning-based methods, automatic segmentation methods have shown excellent performance for image segmentation tasks in general and can be promising for aqua-culture and fish

[*]Corresponding authors. Email: yem.vuvan@hust.edu.vn
thao.tranthi@hust.edu.vn

segmentation. In deep learning-based approach for image segmentation, the forerunner approaches are relied on the well-known Fully Convolutional networks (FCN) [7]. Inspired by the FCN, many invariants such as the combination between the FCN variants with backbones original proposed for image classification tasks such as ResNet [8] have achieved dominant segmentation performance.

Although these above networks have shown their performance efficiency in various segmentation tasks, when applied for a challenging task like the fish segmentation from underwater images, it is necessary to be adapted with the task. In addition, for training the networks, generally the Cross Entropy is used to measure the dissimilarity between the masks of ground truths and the masks of the predictions, so it is lack of information about the length of the boundary and image intensity of object to be segmented. In this study, for the network architecture, inspired by the advances of deep learning based-approach, we propose a network for fish segmentation by incorporating into the FCN network with Wide ResNet (WRN) modules, and the Waterfall Atrous Spatial Pooling (WASP) [9]. The proposed network is hence named as FCN-WRN-WASP. In addition, along with the Cross Entropy loss, we propose an active contour loss that exploits the local intensity information from the input image to handle the intensity inhomogeneity in the images.

The proposed approach has been validated on two datasets including DeepFish [2] and SIUM [10]. The intensive evaluation and results show the promising performance of the proposed approach. The fundamental contributions of the study can be summarized as follows:

i. A new neural network namely FCN-WRN-WASP based on FCN has been proposed with the exploiting of Wide ResNet and Waterfall Atrous Spatial Pooling.
ii. A new loss function has been introduced based on localized based active contour model.
iii. A comprehensive evaluation of fish segmentation has been made to evaluate the performance of the proposed network and loss function.

In the remainder of this paper, we present the related work in Section 2; The methodology is given in Section 3. The experimental results are provided in Section 4. The conclusion is drawn in the last section of the paper.

2. Related work

2.1. Deep learning-based segmentation

The currently ubiquitous application of convolutional neural networks (CNNs) [11] has consequential utilization in semantic segmentation. By classifying object pixels into respective categories, semantic segmentation has been applied in numerous areas [7,12]. The pioneered approach, FCN proposed by Long et al. has been get much interest and has been inspiration for other approaches. FCNs [7] frequently generate pixel-wise labelling results with usage of encoder-decoder architecture for high-resolution images. For multi-scale feature fusion, feature pyramid plays as an efficient approach. Pyramid Pooling module has been introduced in PSP-Net [13], which encourages more representative context information extraction. Furthermore, as Atrous Spatial Pyramid Pooling [14] (ASPP) makes the use of atrous convolution filters at several dilation rates to capture small image information. DeepLab-based models have got consistent achievement on capturing objects at multiple scales due to the advantages of dilated convolutions and Atrous Spatial Pyramid Pooling (ASPP).

2.2. Wide Residual Neural Networks

Since published, the original ResNet by He et al. [15] has been a big deal in the world of deep learning. The use of residual blocks in ResNet models is very efficient for building deeper neural networks to scale to hundreds and even thousands of layers and get an improvement in terms of accuracy. Nevertheless, the idea of just stacking one residual block after the other has shown some shortcomings especially when training very deep residual networks [16]. To address these, in the work by Sergey Zagoruyko and Nikos Komodakis [16], they propose the 'Wide Residual Networks, (WRN for short). The structure of the Wide Residual Networks in comparison with the original ResNet module is shown in Fig.1. The WRN decreases the depth and increases the width of residual networks. By widening the ResNet, the WRN can be shallower with the same accuracy or improved accuracy.

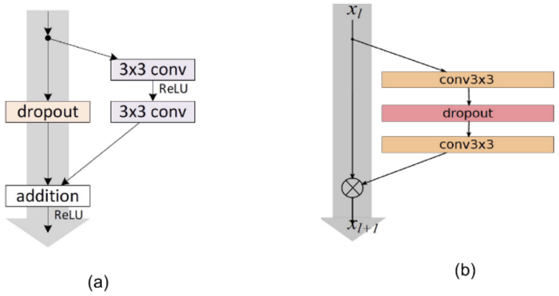

Figure 1. Comparison between (a) Residual Neural Network (ResNet) and (b) Wide Residual Neural Network (WRN) with dropout [16]

2.3. Waterfall Atrous Spatial Pooling (WASP)

The Waterfall Atrous Spatial Pooling (WASP) has been proposed by Bruno Artacho and Andreas E. Savakis in [9]. The WASP has shown efficient architecture for semantic segmentation. It leverages progressive filtering in a cascading architecture and maintains multiscale fields-of-view (FOV) comparable to spatial pyramid configurations. In [9], WASP when combined with the Resnet backbone will provide a powerful, efficient architecture and obtain potential results for segmentation problems. WASP is a computational-efficient variation, which is an Atrous Spatial Pooling (ASP) class variant in the DeepLabv3+ architecture. Sharma and

Shorya in [17] demonstrated the efficiency improvement of the WASP module in terms of computation time during training and parameter reduction. The WASP module is shown visually in Fig. 2. The WASP module consists of four branches of a large-FOV being fed forward shaped like a waterfall (Waterfall), and then combined together to give output. WASP is a novel architecture with Atrous Convolutions that is able to leverage both the larger FOV of the ASPP configuration and the reduced size of the cascade approach [9]. Unlike ASPP block, WASP shares sequential parameters to subsequent branches, thereby extracting more information and branch learning correlations with each other.

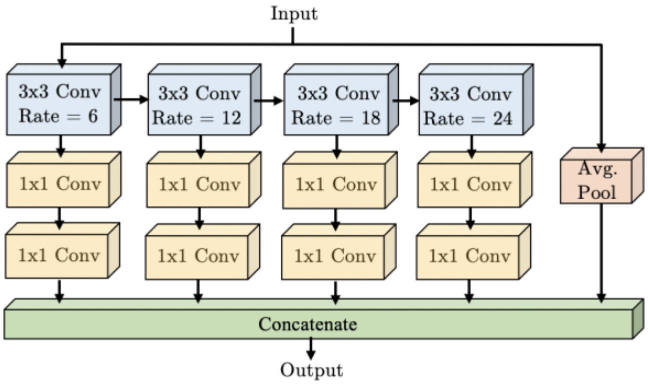

Figure 2. Waterfall Atrous Spatial Pooling (WASP) module [9]

3. Methodology

3.1. Model architecture

In this study, we propose a new architecture of deep learning for fish segmentation that combines the FCN network with Wide ResNet (WRN) modules, and the Waterfall Atrous Spatial Pooling (WASP). The proposed network architecture, namely FCN-WRN-WASP is shown in Fig.3. As in common Fully Convolutional Networks for image segmentation, the proposed architecture contains the two paths, encoder and the decoder. The encoding path is account for feature extraction and the decoder is used for upsampling to obtain the segmentation mask with the same resolution with the input image.

As shown in Fig.3, the encoder includes 4 WRN layers followed by convolutional layer and max pooling layers. In the WRN layers except the third layer, 3 consecutive convolution operations followed by activation operations are stacked each after the other, then fed into the pooling operation and go into the consequent layer. After the last WRN layer, the WASP module that combines the cascaded approach in [18] is added. The WASP helps to reduce the number of parameters and memory required, which leads to a reduction in the amount of computation, which is the main limitation of Atrous Convolutions [9].

After the input image has been forwarded through these layers, the outputs are upsampled and interpolated in the decoder path to get the segmentation map. In addition, to promote the segmentation outcome, in this study, we exploit the affinity-based approach for post-processing the final segmentation output [19]. To be more specific, the affinity-based approach used the affinity matrix to cluster the pixels into coherent classes that computes the correlation in brightness, color, and texture between image patches; thus, determining the gradient connecting two pixels in segmentation task. This postprocessing process is similar to the commonly used Conditional Random Fields (CRF) [20] for deep learning-based image segmentation methods. The total size of the proposed FCN-WRN-WASP model is 186.4M training parameters.

3.2. Proposed loss function

In this study, inspired from the localized based active contour models we proposed, a new loss function for fish segmentation as:

$$L_{CE\_AC} = \lambda L_{CE} + \alpha L_{AC} + \mu L_{Len} \qquad (1)$$

where λ,α, and μ are hyper-parameters; L_{CE}, L_{LAC}, L_{Len} are respectively the cross-entropy loss, the localized based active contour loss, and the length terms, defined as the following.

Let I be the image to be segmented, and **Ω** be the domain of the image I. Denote u_x as the pixel value located at the **x**-th pixel of the ground truth u, and v_x as the pixel value located at the **x-**th pixel of the v predicted from the deep learning model.

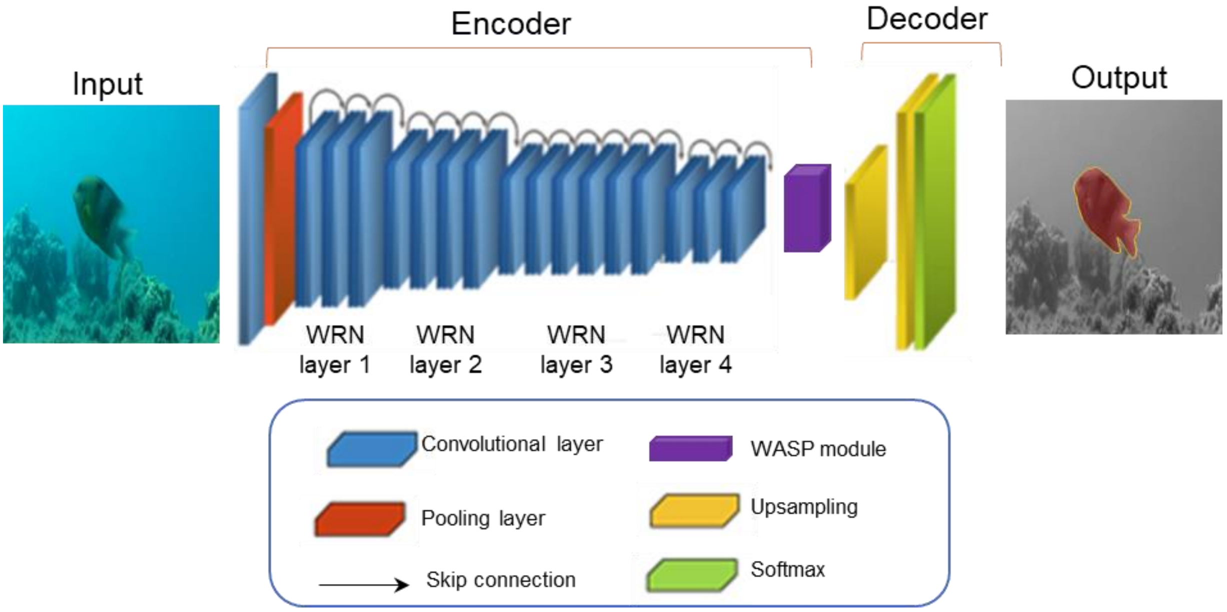

Figure 3. The proposed FCN-WRN-WASP network for fish segmentation

The cross-entropy loss is defined as:

$$L_{CE} = \frac{-1}{W \times H} \sum_{\mathbf{x}\in\Omega} \left(u_{\mathbf{x}} \log\left(v_{\mathbf{x}}\right) + \left(1-u_{\mathbf{x}}\right) \log\left(1-v_{\mathbf{x}}\right) \right) \qquad (2)$$

with W and H are respectively width and height of the image.

For the Localized active contour loss, inspired by the energy function in works in level set based active contour models [21], [22]. Let $I_{\mathbf{x}}$ the pixel value located at $\mathbf{x}$th of image I, and consider a circular neighborhood with a radius of σ cantered at each pixel $\mathbf{y} \in \Omega$, with $\Omega_{\mathbf{y}} \,\square\, \left\{ \mathbf{x} : \left|\mathbf{x}-\mathbf{y}\right| \le \sigma \right\}$. Consider a nonnegative window function $\mathrm{K}\left(\mathbf{y}-\mathbf{x}\right)$, with $\mathrm{K}\left(\mathbf{y}-\mathbf{x}\right)=0$ for $\mathbf{x} \notin \Omega_{\mathbf{y}}$ as in [21].

The Localized active contour loss is proposed as an unsupervised term that measures the dissimilarity between the image intensity values inside and outside the prediction v of the image I, is expressed as:

$$L_{AC} = \frac{1}{W \times H} \sum_{\mathbf{x}\in\Omega} \left(v_{\mathbf{x}} \left(I_{\mathbf{x}} - d_1\right)^2 + \left(1-v_{\mathbf{x}}\right)\left(I_{\mathbf{x}} - d_2\right)^2 \right) \qquad (3)$$

where

$$d_1 = \frac{\sum_{\mathbf{y}\in\Omega_{\mathbf{x}}} K\left(\mathbf{x}-\mathbf{y}\right) I_y v_y}{\sum_{\mathbf{y}\in\Omega_{\mathbf{x}}} K\left(\mathbf{x}-\mathbf{y}\right) v_y}; d_2 = \frac{\sum_{\mathbf{y}\in\Omega_{\mathbf{x}}} K\left(\mathbf{x}-\mathbf{y}\right) I_y \left(1-v_y\right)}{\sum_{\mathbf{y}\in\Omega_{\mathbf{x}}} K\left(\mathbf{x}-\mathbf{y}\right)\left(1-v_y\right)} \qquad (4)$$

The length term, adapted from [23] is defined as

$$L_{Len} = \frac{1}{W \times H} \sum_{\mathbf{x}\in\Omega} \left| v_{\mathbf{x}+1} - v_{\mathbf{x}} \right| \qquad (5)$$

The length term is just used to regulate the smoothness of the prediction.

It is noted that the cross-entropy loss is a supervised term that measures the dissimilarity between the binary masks of ground truth u and prediction v of the image I. The localized active contour is an unsupervised term that measures the dissimilarity between the local image intensity inside and outside the v. The local active contour term only considers the image intensity, while the cross entropy takes the ground truth into account.

4. Experimental results

To evaluate the performance of the proposed approach for fish segmentation, we assess and conduct the comparative experiments on two datasets, the DeepFish dataset [2] and the SUIM dataset [10]. The segmentation results are also given in comparison to those reported by previous works. In addition, ablation study is also made to evaluate the performance of the WASP module in the neural network architecture, and the role of the localized active contour loss term in the loss function.

4.1. Benchmarks

DeepFish Dataset

The DeepFish dataset [2], contains about 40000 images of 20 aqua-environments in Australia. This dataset is classified into 3 groups: FishClf consists of classification annotations for classification task; FishLoc consists of point-label for localization task and FishSeg consists of fish segmentation annotations for segmentation task. In this paper we have utilized the FishSeg set for the task of fish segmentation. The FishSeg data include 620 images along with their corresponding masks. The set is divided into 310 images for training, 124 images for validating and the remained 186 images are used for testing.

SUIM Dataset

The SUIM dataset [10] provides masks for multiple categories, and also separate the annotations for each object category in the test set. Thus, we can use the fish and other vertebrate in the SUIM dataset for fish segmentation task. Similar to the work of Zhang et al. [24] the fish and other vertebrate categories are assigned as the foreground while other categories as background for training and validation. The data for fish segmentation include 1525 image pairs for training and validation phase, and 110 images are used for testing.

4.2. Implementation details

Training

The neural network is trained with the Pytorch framework and conduct experiments with NVIDIA Tesla P100 16GB GPU using Nadam optimizer with a learning rate of 0.00001 through the training period with 200 epochs on the DeepFish and the SIUM benchmarks. The training time of our proposed network is approximately 2-3 hours. For hyperparameters of the loss, the λ controls the importance of the reference masks to the prediction masks, so it is set as 1. The α regulates the impact of the local image intensity and is set small as 0.01. The μ is used to make the contour smooth so and is typically set as 10^{-5}. The setting of the parameters are based on experiments and experience from previous research on the active contour- based models.

Evaluation Metrics

Intersection over Union index (IoU) is used to evaluate the performance of the segmentation by the neural network. IoU measures the similarity and diversity of sample pixel sets, which is determined by:

$$IoU = \frac{TP}{TP + FN + FP} \tag{6}$$

where TP, FN, and FP are respectively the number of true positive, false negative, and false positive predictions.

4.3. Results on the DeepFish data

Representative results on DeepFish

For observation, qualitative results on the test set of the DeepFish data by the proposed segmentation approach performance are shown in Fig. 4. As shown in this figure, proposed algorithm achieves quite proper contours of the fish boundaries even for small objects, and close to the desired object boundary. The approach can segment multiple fish in the existence of complex underwater background, as can be easily observed in the second and fourth row of Fig. 4.

Evaluation on DeepFish

To evaluate the performance of the proposed approach in terms of the IoU scores, we provided the comparative results with other state of the arts in Table 1. As shown in the last row of Table 1, our approach achieves highest/best scores for both background and foreground classes of the DeepFish segmentation dataset, with the average IoU of 94.88% on the test set of this benchmark.

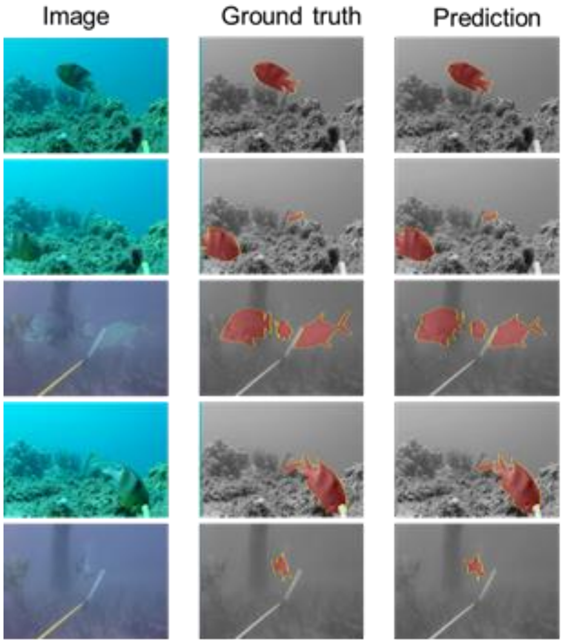

Figure 4. Representative fish segmentation results by the proposed approach on the DeepFish data. The Ground truths and Predictions are overlaid on the gray scale image for better visualization.

Table 1. Comparison between other methods for fish segmentation on the Deepfish data

Methods	IoU scores on the DeepFish test set		
	Background (%)	**Foreground** (%)	**Mean IoU** (%)
FCN [7]	99.21	66.30	82.75
SegNet [25]	98.89	68.94	83.91
DeepLabv3+[26]	99.11	71.35	85.23
SPSNet [13]	99.15	72.61	85.88
SIUM-Net [10]	99.03	78.40	88.71
DGCNet [27]	99.21	81.42	90.32
DRANet [28]	99.33	79.42	89.37
GFFNet [29]	99.20	81.49	90.35
DPANet [24]	99.31	82.86	91.08
FCN8-ResNet50 [30]	99.70	86.37	93.03

FCN8-VGG16 [30]	99.72	87.73	93.73
Proposed FCN-WRN-WASP	**99.78**	**89.98**	**94.88**

4.4. Results on the SIUM fish data

Representative results on the SIUM fish

The representative results on the test set of the SIUM data by the proposed segmentation approach are shown in Fig. 5. As shown in this figure, the results by the proposed approach are in good agreement with those in ground truth. The fish can be segmented even in the presence of intensity inhomogeneity and occlusion by the background, as obviously shown in the third and fifth rows of this figure.

Evaluation on the SIUM fish

For quantitative assessment on the SIUM data, the comparison by the approach and other comparative methods are provided in Table 2. As shown in the table, the proposed approach achieves the mean IoU of 86.05%, the best average score compared to other state of the arts. For the background segmentation, the IoU by the proposed approach is the second to best, lower than that by DPANet, but the score for the foreground is highest, 74.41% compared to 72.45% by the DPANet.

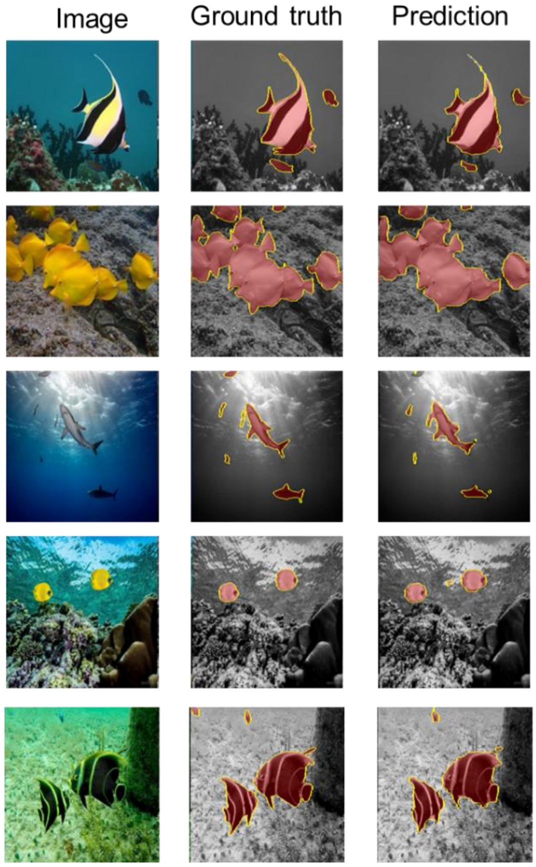

Figure 5. Representative fish segmentation results by the proposed approach on the SIUM data. The Ground truths and Predictions are overlaid on the gray scale image for better visualization.

Table 2. Comparison between other methods for fish segmentation on the Sium data

Methods	IoU scores on the SIUM test set		
	Background (%)	**Foreground** (%)	**Mean IoU** (%)
FCN [7]	94.17	68.25	81.21
SegNet [25]	96.49	69.23	82.86
DeepLabv3+[26]	95.90	62.72	79.31
SPSNet [13]	92.00	57.32	74.66
SIUM-Net [10]	97.43	70.13	83.78
DGCNet [27]	98.04	69.84	83.94
DRANet [28]	97.21	71.01	84.11
GFFNet [29]	97.30	70.41	83.85
DPANet [24]	**98.33**	72.45	85.39
FCN8-ResNet50 [ ][30]	96.78	65.69	81.23
FCN8-VGG16 [30]	96.50	63.87	80.18
Proposed FCN-WRN-WASP	97.69	**74.41**	**86.05**

4.5. Ablation study

To evaluate the performance of the network and loss function proposed in the current study, comprehensive evaluation of fish segmentation has been made. In the first experiment, the WASP is eliminated from the architecture in the Fig.3. For the second experiment, we compare the results when using the Cross-Entropy loss by setting λ=1, α=0, and μ=0 from Eq.1.

Table 3a and 3b show the mean IoU scores in the cases of without using WASP (w/o WASP column) and with WASP module (w/ WASP column) on the DeepFish and the SIUM fish data sets respectively. As can be seen from the table, by using the WASP, the mean IoU increases significantly for foreground segmentation task, with an increase of 1.5% for the DeepFish data, and about 2.5% for SIUM fish data. This proves the advantage of the WASP module in the proposed FCN-WRN-WASP architecture.

Table 3. Comparison between the fish segmentation performance when using WASP (w/ WASP) and without using WASP (w/o WASP)

	a) Performance on the DeepFish		
	Background (%)	**Foreground** (%)	**Mean IoU** (%)
w/o WASP	99.74	88.42	94.08
w/ WASP	**99.78**	**89.98**	**94.88**

	b) Performance on the SIUM		

	Background (%)	**Foreground** (%)	**Mean IoU** (%)
w/o WASP	97.33	71.99	84.66
w/ WASP	**97.69**	**74.41**	**86.05**

To evaluate the impact of using the local image based active contour loss term in the loss function, we provide the results while using the proposed loss with those by using the common losses in image segmentation including Dice loss, Tversky loss, and Cross Entropy (CE) loss

To further show the effectiveness of the proposed loss, we provided the representative segmentation results by the comparative losses and the proposed loss on the DeepFish and SIUM fish datasets in Figs.6 and 7. As can be observed in the figure, the segmented results by the proposed loss are in better agreement with the ground truths while compared to those by other losses.

For quantitative assessment, we provided the IoU scores by training the proposed model with different losses on the two datasets in Table 4. As can be seen from Table 4a, and 4b, the CE loss and the proposed loss give better scores compared to the Dice loss and Tversky loss. Nevertheless, while compared to the CE loss, using the proposed loss, the mean IoU for foreground segmentation increases about 2% for DeepFish data, and approximately 2.5% for SIUM fish data.

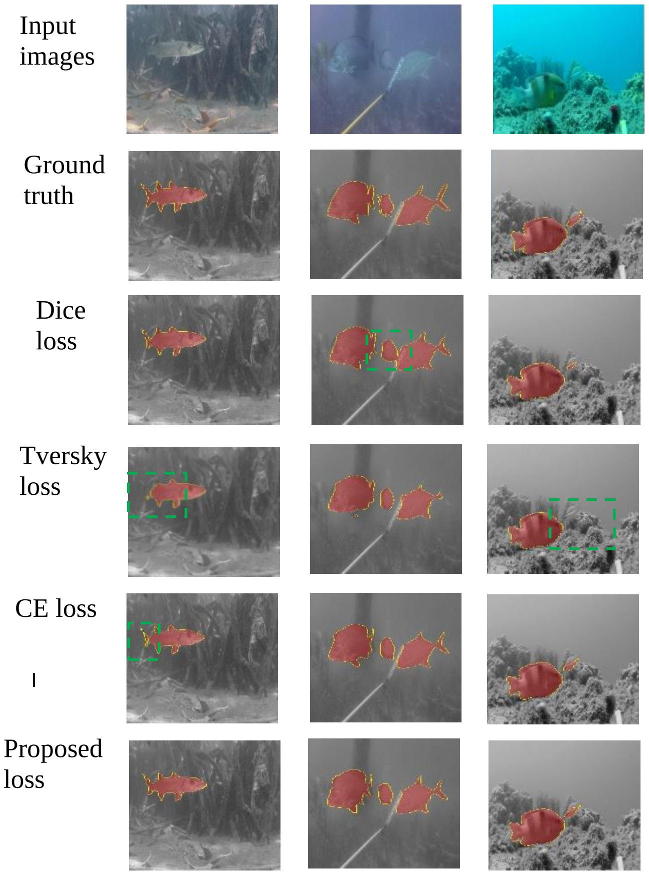

Figure 6. Representative fish segmentation results by the proposed model when trained with different losses on the (a) DeepFish data and. The Ground truths and Predictions are overlaid on the gray scale image for better visualization. The green dot rectangulars denote the undersegmented regions

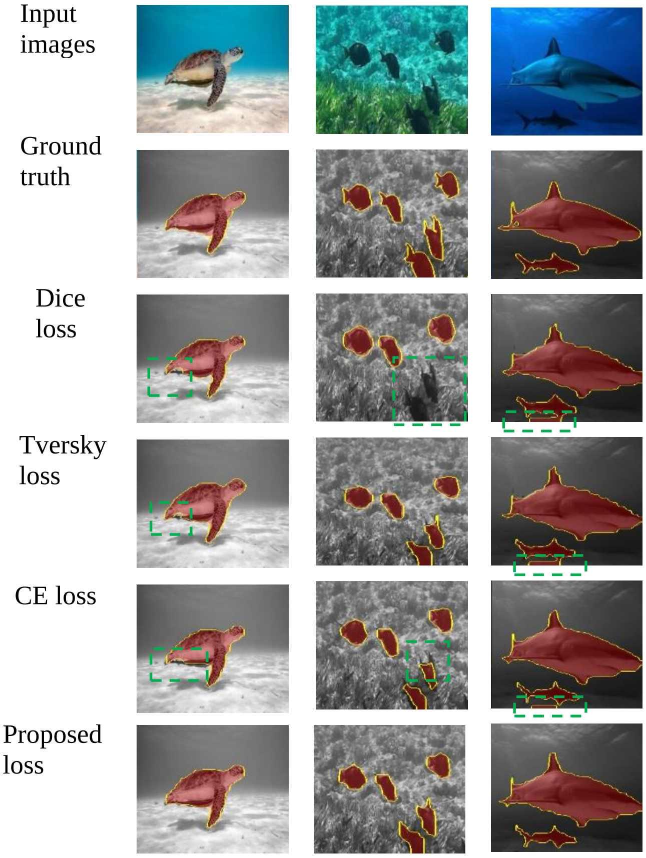

Figure 7. Representative fish segmentation results by the proposed model when trained with different losses on the SIUM fish data. The Ground truths and Predictions are overlaid on the gray scale image for better visualization. The green dot rectangulars denote the undersegmented regions

Table 4. Comparison between the fish segmentation performance when using the proposed loss and comparative losses

Loss	***a) Performance on the DeepFish data***		
	Background (%)	**Foreground** (%)	**Mean IoU** (%)
Dice loss	99.73	88.03	93.88
Tversky loss	99.72	87.15	93.43
Cross Entropy	99.74	88.04	93.89
Proposed	**99.78**	**89.98**	**94.88**
	b) Performance on the SIUM fish data		

	Background (%)	**Foreground** (%)	**Mean IoU** (%)
Dice loss	96.87	66.49	81.68
Tversky loss	96.97	67.63	82.30
Cross Entropy	97.45	71.95	84.70
Proposed	**97.69**	**74.41**	**86.05**

5. Conclusion

We have proposed a new DL based approach for fish segmentation. The FCN-based architecture utilizes the Wide ResNet and Waterfall Atrous Spatial Pooling that leverages the progressive extraction of larger fields-of-view from cascade methods for better segmentation. Besides, the approach introduces a localized based active contour loss for training the network that exploits the local intensity information of the segmented image. Through experiments on the DeepFish and SIUM datasets, the proposed approach shows dominant/promising results especially for foreground segmentation with higher IoU scored while compared with other state of the arts.

Acknowledgements.

This work is carried out in the framework of the project “Design and implementation of IoT system for aquacuture application in Kien Giang”. The authors would like to thank the Department of Science and Technology of Kien Giang, Vietnam for supporting this research.

References

1. Hussain, M.A., Saputra, T., Szabo, E.A., Nelan, B.: An overview of seafood supply, food safety and regulation in New South Wales, Australia. Foods **6**(7), 52 (2017). doi:https ://doi.org/10.3390/foods
2. Saleh, A., H. Laradji, I., A. Konovalov, D., Bradley, M., Vazquez, D., Sheaves , M.: A Realistic Fish Habitat Dataset to Evaluate Algorithms for Underwater Visual Analysis. Scientific Reports **10 Article number: 14671** (2020). doi:DOI: 10.1038/s41598-020-71639-x
3. Delgado, C., Wada, N., Rosegrant, M., Meijer, S., Ahmed, M.: Fish to 2020: Supply and demand in changing global markets. World Fish Center Technical Report **62** (2003).
4. Yang, L., Liu, Y., Yu, H., Fang, X., Song, L., Daoliang, L., Chen, Y.: Computer Vision Models in Intelligent Aquaculture with Emphasis on Fish Detection and Behavior Analysis: A Review. Archives of Computational Methods in Engineering **28**, 2785–2816 (2021).
5. Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag , A.A., Moriarty, T.: A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. IEEE Trans. Med. Imaging **21**(3), 193-199 (2002).
6. Tran, T.T., Pham, V.T., Shyu, K.K.: Image segmentation using fuzzy energy-based active contour with shape prior. J. Vis. Commun. Image Represent. **25**(7), 1732-1745 (2014).
7. J. Long, E. Shelhamer, T. Darrell: Fully convolutional networks for semantic segmentation. Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 3431–3440 (2015).
8. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016, pp. 770-778
9. Artacho, B., Savakis. A.: Waterfall atrous spatial pooling architecture for efficient semantic segmentation. Sensors **19**(4), 5661 (2019). doi:https://doi.org/10.3390/s19245361
10. Islam, M.J., Edge, C., Xiao, Y., Luo, P., Mehtaz, M., Morse, C., Enan, S., Sattar, J.: Semantic segmentation of underwater imagery: dataset and benchmark. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) 2020, pp. 1769–1776
11. O'Shea, K., Nash, R.: An Introduction to Convolutional Neural Networks. arXiv:1511.08458 (2015).
12. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. 2015, pp. 234-241
13. Zhao, H., Shi, J., Qi, X., Wang, X., Jia, J.: Pyramid Scene Parsing Network. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2017
14. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., L. Yuille, A.: DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(4), 834 - 848 (2018).
15. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016
16. Zagoruyko, S., Komodakis, N.: Wide Residual Networks. In: Proceedings of the British Machine Vision Conference (BMVC) 2016, pp. 87.81-87.12
17. Shorya, S.: Semantic Segmentation for Urban-Scene Images. arXiv:2110.13813 (2021). doi:https://doi.org/10.48550/arXiv.2110.13813
18. Baevski, A., Auli, M.: Adaptive input representations for neural language modeling. In: The International Conference on Learning Representations (ICLR) 2019
19. Fowlkes, C., Martin, D., Malik, J.: Learning affinity functions for image segmentation: combining patch-based and gradient-based approaches. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2003, pp. II-54
20. Krähenbühl, P., Koltun, V.: Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials. In: Advances in Neural Information Processing Systems 24 2012, pp. 109-117
21. Li, C., Kao, C.Y., C. Gore, J., Ding, Z.: Minimization of Region-Scalable Fitting Energy for Image Segmentation. IEEE Transactions on Image Processing **17**(10), 1940 - 1949 (2008).
22. Lankton, S., Tannenbaum, A.: Localizing Region-Based Active Contours. IEEE Transactions on Image Processing **17**(11), 2029 - 2039 (2008).
23. Chen, X., M. Williams, B., R. Vallabhaneni, S., Czanner, G., Williams, R., Zheng, Y.: Learning Active Contour Models for Medical Image Segmentation. Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 11623-11640 (2019).
24. Zhang, W., Wu, C., Bao, Z.: DPANet: Dual Pooling-aggregated Attention Network for fish segmentation. IET computer vision, 67-82 (2021).
25. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017).

26. Chen, L., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv:1706.05587 (2017). doi: http://arxiv.org/abs/1706.05587
27. Zhang, L., Li, X., Arnab, A., Yang, K., Tong, Y., Torr, P.: Dual graph convolutional network for semantic segmentation. arXiv:1909.06121 (2019). doi:http://arxiv.org/abs/1909.06121
28. Fu, J., Liu, J., Jiang, J., Li, Y., Bao, Y., Lu, H.: Scene segmentation with dual relation-aware attention network. IEEE Tran. Neural Netw. Learni. Syst. **32**(6), 2547-2560 (2020). doi:https://doi.org/10.1109/TNNLS.2020.3006524
29. Li, X., Zhao, H., Han, L., Tong, Y., Yang, K.: GFF: gated fully fusion for semantic segmentation. arXiv:1904.01803 (2019). doi:http://arxiv.org/abs/1904.01803
30. Yoo, I.: Sementic-segmentation-pytorch: Pytorch implementation of FCN, UNet, PSPNet and various encoder models. https://github.com/IanTaehoonYoo/semantic-segmentation-pytorch (2020). Accessed June 14 2020