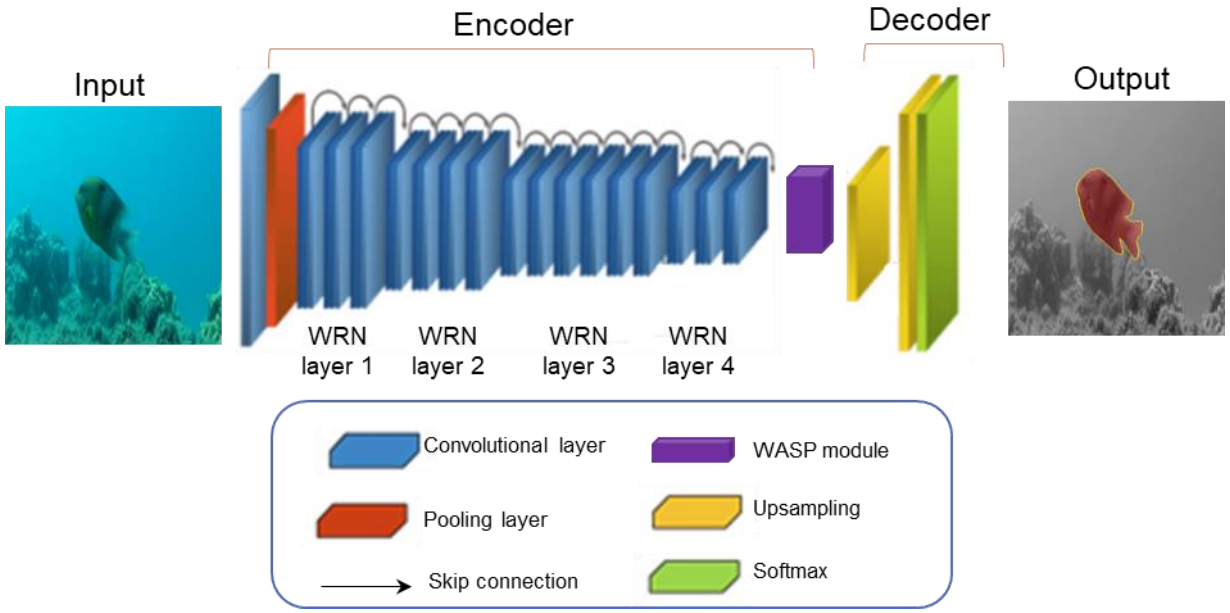

Figure 3. The proposed FCN-WRN-WASP network for fish segmentation

The cross-entropy loss is defined as:

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

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 $\Omega_{\mathbf{y}} = \left\{ \mathbf{x} : |\mathbf{x} - \mathbf{y}| \le \sigma \right\}$. 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:

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

where

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

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

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

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.

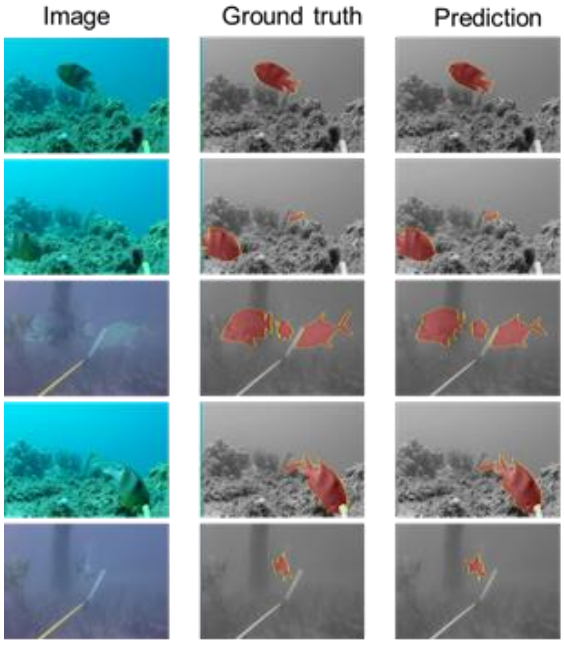

Figure 4. Representative fish segmentation results by the proposed approach on the DeepFish data. The Ground truths and Predictions are overlaid on the gray scale image for better visualization.

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.

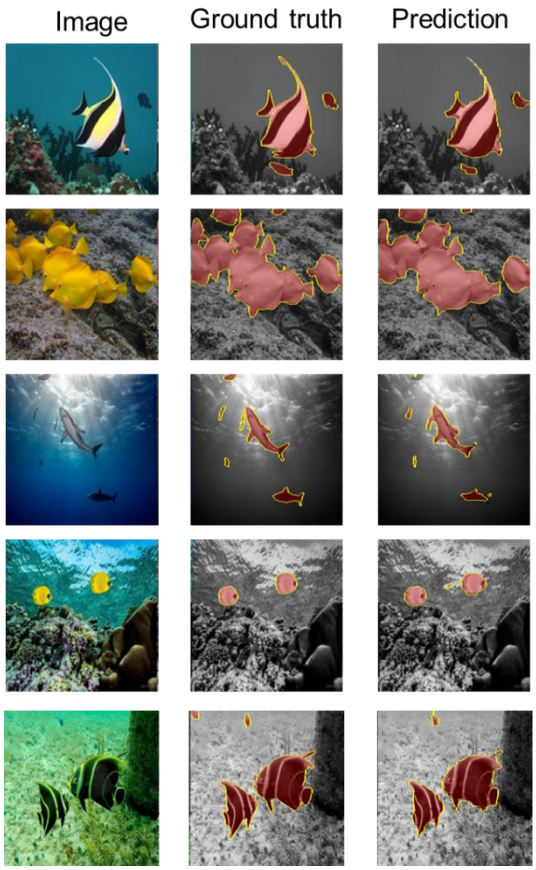

Figure 5. Representative fish segmentation results by the proposed approach on the SIUM data. The Ground truths and Predictions are overlaid on the gray scale image for better visualization.

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.

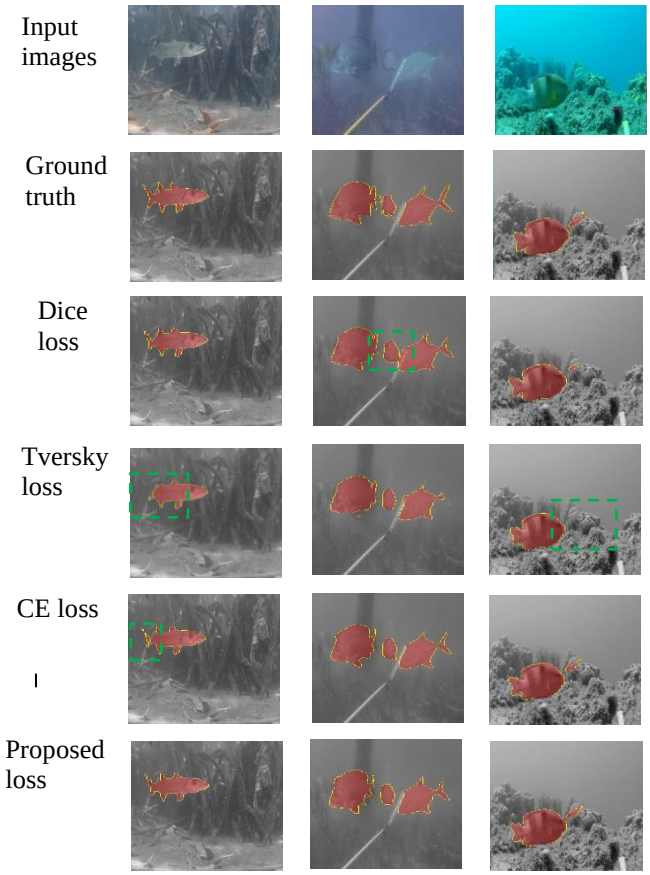

Figure 6. Representative fish segmentation results by the proposed model when trained with different losses on the (a) DeepFish data and. The Ground truths and Predictions are overlaid on the gray scale image for better visualization. The green dot rectangulars denote the undersegmented regions

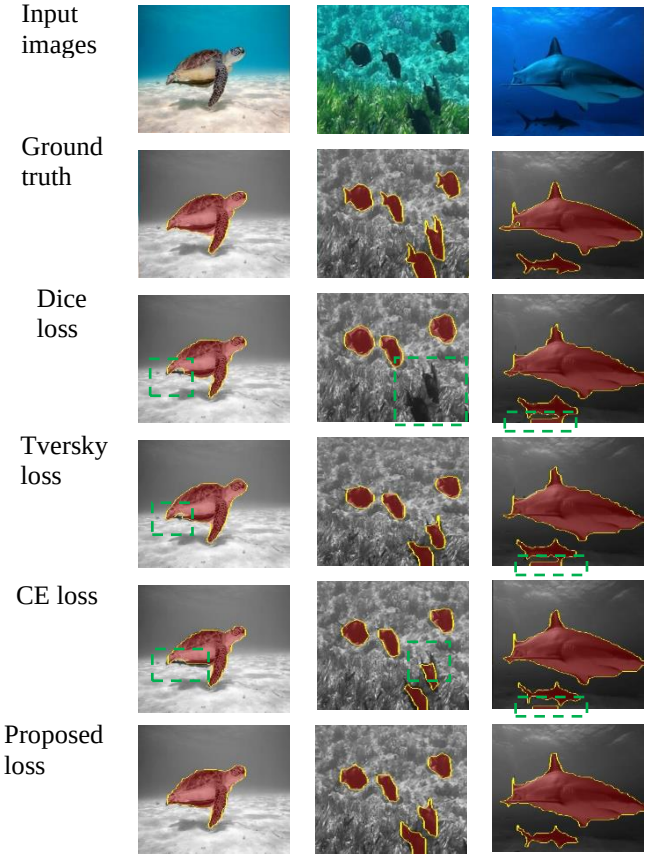

Figure 7. Representative fish segmentation results by the proposed model when trained with different losses on the SIUM fish data. The Ground truths and Predictions are overlaid on the gray scale image for better visualization. The green dot rectangulars denote the undersegmented regions

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