Semantic Segmentation of the Underwater Images

Hari Mudipalli • Steven Fields • Khushboo Singh • Dinesh Reddy

Introduction:

Semantic Segmentation is a popular problem in the computer vision domain which is used for estimating scene geometry, inferring interactions and spatial relationships among objects, salient object identification, etc., [1]. On a broader level, segmentation involves three steps, classifying certain objects in an image, localizing them, and grouping the pixels in the localized image by creating a segmentation mask [8].

The recent introduction of transformer-based models and attention mechanisms in computer vision tasks has taken the spotlight. In this project, we are interested in applying the attention mechanism to a new segmentation model from [1] to underwater images. Semantic segmentation of underwater images is not well explored as in the case with terrestrial object segmentation. Authors of [1] have curated a dataset called Segmentation of Underwater IMagery (SUIM) and implemented SOTA segmentation models on it to compare with their model called SUIM-Net. We tweaked the SUIM-Net Residual Skip Block (RSB) by experimenting with the Efficient Channel Attention module [2] and Triplet Attention module [3] inside the SUIM-Net.

Background and related work:

With the advent of deep learning and large-scale annotated image and video datasets, learning-based semantic segmentation methodologies have made remarkable progress in the past few years. Models like VGG, and GoogLeNet were initial models which were based on fully convolutional networks (FCNs) for hierarchical feature extraction. Later encoder-decoder based models showed some promise after FCNs. More effective architectures which focused on global contextual information and instance awareness were proposed like SegNet UNet which was developed on top of the initial encoder-decoder architecture.

The major focus of semantic segmentation has been on images of terrestrial objects but very few on underwater objects. Though few works on fish detection and coral reef detection were explored, those are not so feasible for general underwater robotic use where many more objects are present other than fishes and reefs. This project is chosen to try, not so explored, SOTA models on underwater images using the SUIM dataset taken from [1].

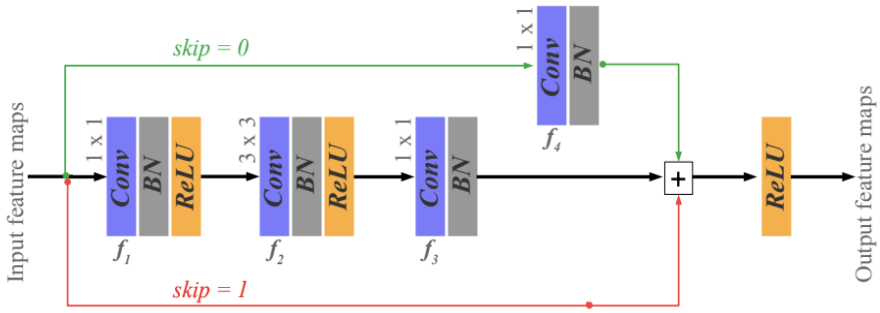
Methods:

The core of this project is to experiment with the Attention mechanism inside of the SUIM-Net. Attention mechanisms are inspired by human visual perception. The fundamental idea behind representation learning is that of finding or extracting discriminative features

from an input that differentiates a particular object from an object of a different type or class [6]. In computer vision models, there are primarily two types of attention mechanisms, spatial attention, and channel attention. In a nutshell, channel attention is used to weigh each feature map/channel in the tensor, while spatial attention provides context at each feature map level by weighing each pixel in a singular feature map [6]. The following sections will give a brief about the architectures involved in this project.

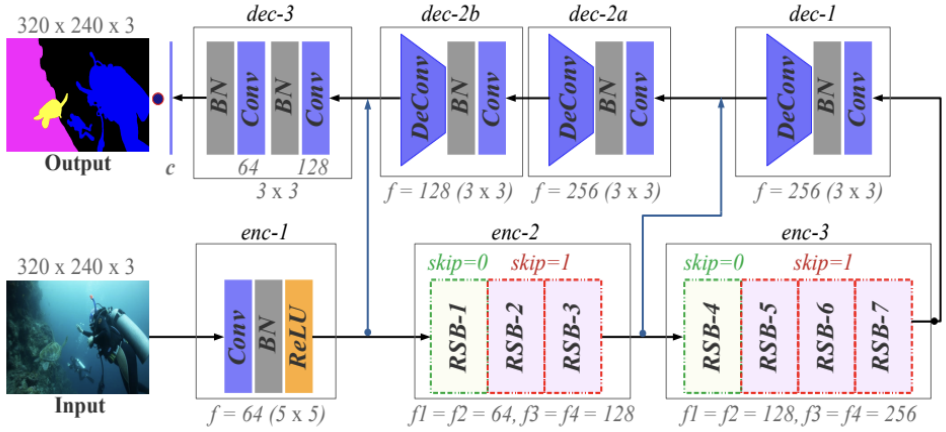
SUIM-Net:

SUIM-Net$_{RSB}$ is a fully convolutional encoder-decoder model with skip connections between mirrored layers [1]. Figure 1 shows the RSB block and Figure 2 shows the SUIM-Net architecture using RSB blocks as encoders.



(a) Architecture of an RSB: the skip-connection can be either fed from an intermediate `conv` layer (by setting *skip=0*) or from the input (by setting *skip=1*) for local residual learning.

Figure 1



(b) The end-to-end architecture of SUIM-Net$_{RSB}$: three composite layers of encoding is performed by a total of seven RSBs, followed by three decoder blocks with mirrored skip-connections.

Figure 2

Efficient Channel Attention:

ECA-Net is developed to overcome the limitations of Squeeze and Excitation net (SENet). Cross-Channel Interaction is diluted with dimensionality reduction (DR) in SENet but ECA-Net address this limitation using the concept of a local neighbourhood. It achieves an adaptive local neighbourhood size for each tensor and computes attention for each channel within that neighbourhood with respect to every other channel within that neighbourhood [5]. Figure 3 shows the ECA (Efficient Channel Attention) module.

ECA has 3 stages:

1. Global Feature Descriptor
2. Adaptive Neighbourhood Interaction
3. Broadcasted Scaling

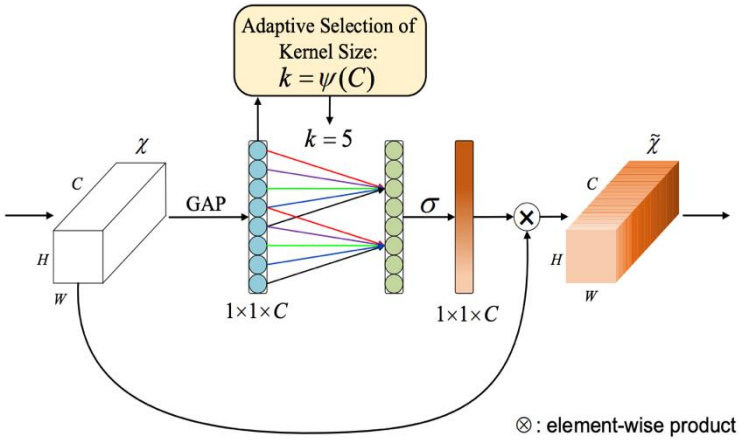


Figure 3

Triplet Attention:

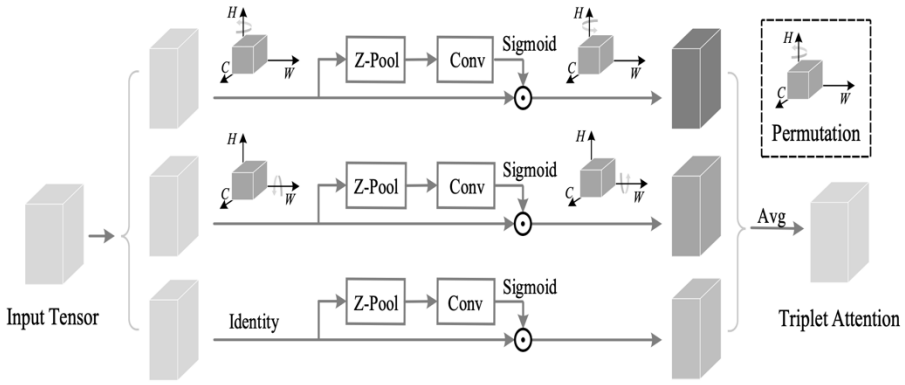


Figure 4

Triplet Attention is a three-branch structure, as shown in figure 4 and figure 5, where each branch is responsible for computing and applying the attention weights across two of the three dimensions of the input tensor [6]. Referencing figure 4, the top 2 branches compute the channel attention and the bottom branch computes the spatial attention. The triplet attention module aims to capture Cross Dimensionality Interaction while allowing some form of dimensionality reduction which is unnecessary for capturing cross-channel interaction and removing information bottleneck [6].

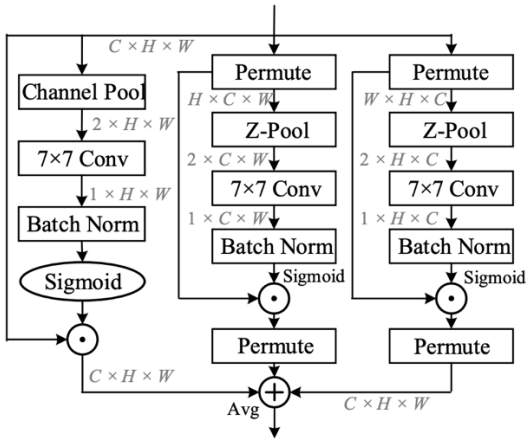


Figure 5

Experiments:

As discussed before, we are interested in checking the behavior of the SUIM-Net when the attention layer is added. For this purpose, we chose 2 attention modules, the Efficient Channel Attention module, and the Triplet attention module. These attention modules were introduced in encoder 1 of the SUIM-Net whose output has a skip connection to the decoder 3.

Three different experiments were run using the following setup,

- Batch size: 4
- Epochs: 40
- steps per epoch: 100
- learning rate: 1e-4
- optimizer: Adam
- loss = 'binary_crossentropy'
- GPU: Tesla T4 (Colab pro)

*SUIM-Net + ECA - 18 + 22 + 40 epochs: There were breaks in the training of SUIM-Net training. Though the plots of SUIM-Net + ECA look different, the results are identical to the other two experiments but on a different scale.

Experiment 1: Only SUIM-Net

Experiment 2: SUIM-Net + Efficient Channel Attention

Experiment 3: SUIM-Net + Triplet Attention

Results & Discussions:

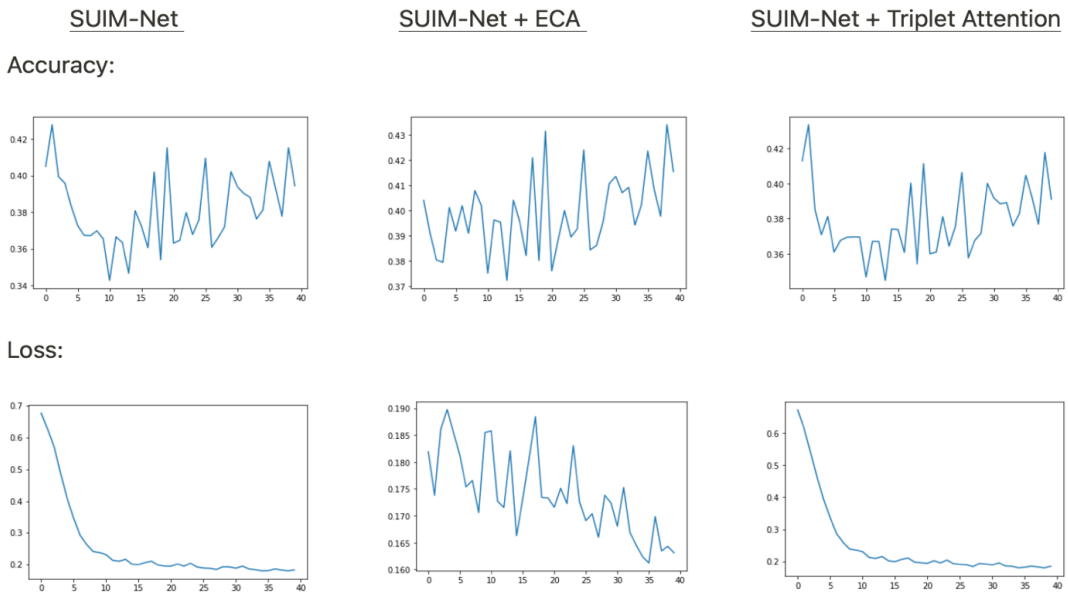


Figure 6

The above figure highlights the accuracy and loss of our models for SUIM-Net, SUIM-Net + ECA, and SUIM-Net + Triplet Attention. As you can see 3 experiments follow a near-identical loss and accuracy progression, which makes it difficult to truly tell how much impact the addition of the attention module had on the SUIM-Net model. If we had more time to train it might show a considerable difference, but at the time it does not present a big enough benefit to outweigh the cost of the additional training time the module would add to the model.

When comparing the results of all three models in the figure below, we see similar outcomes. While the results are not anything remarkable due to the limited training time, we do see that each model is making progress at a similar rate to being able to predict different portions of the image. Notably, it manages to predict the outline of the vessel in the photograph (figure 7), which is impressive considering the drastically lower amount of training we did compared to the original model.

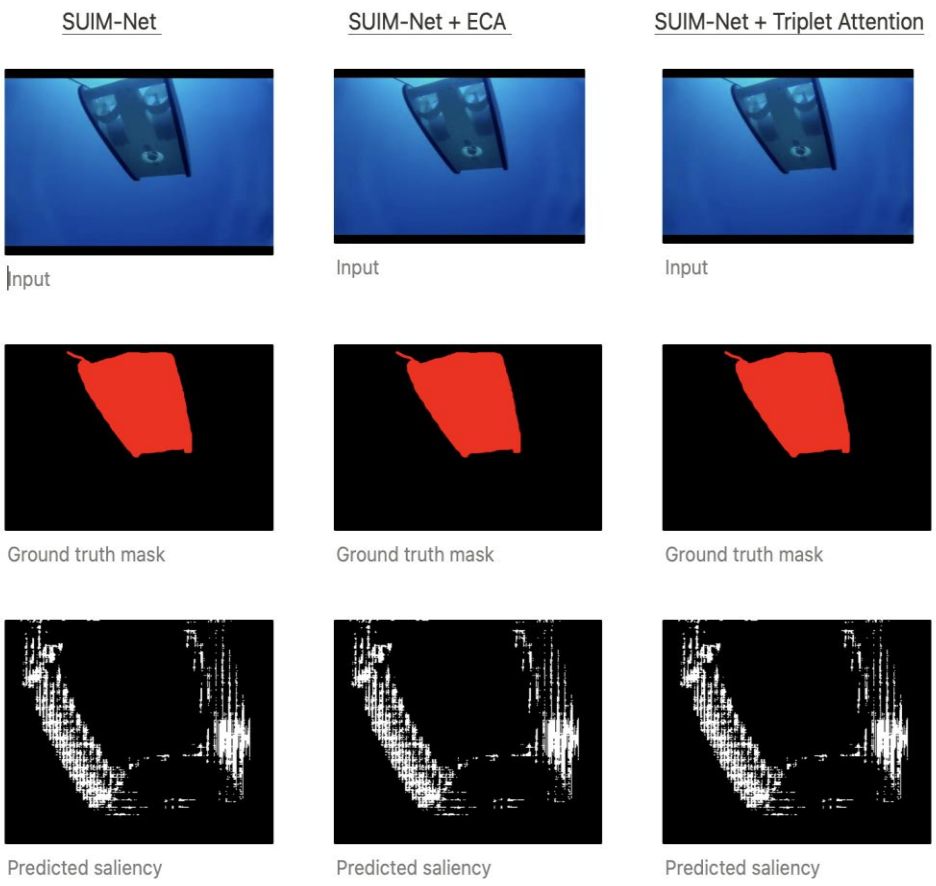


Figure 7

Conclusion:

In this project, we focused on observing the effects of the attention mechanism on SUIM-Net. Due to the computational constraints, we had to reduce a few training parameters significantly, which we think is the reason for not producing any significant difference in the performances of our experiments. In future, we intend to re-run these experiments with sufficient compute power. We would also like to perform additional experiments using transformers.

References:

[1] Semantic Segmentation of Underwater Imagery: Dataset and Benchmark

[2] ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks

[3] Rotate to Attend: Convolutional Triplet Attention Module

[4] An Overview of the Attention Mechanisms in Computer Vision

[5] ECA-Net blog(code used for this project)

[6] Triplet Attention Explained (WACV 2021)

[7] Triplet Attention GitHub(used for this project)

[8] Computer Vision Tutorial: A Step-by-Step Introduction to Image Segmentation Techniques (Part 1)

[9] Dataset and Code of SUIM-Net used by authors in their paper(Used for this project)