Underwater Semantic Segmentation using Multi-scale Feature Extraction

Devika Singh (210101036)
Dr. Arijit Sur

Department of Computer Science and Engineering
Indian Institute of Technology Guwahati

26 November 2024

Contents

INTRODUCTION

SEMANTIC SEGMENTATION

- **Categorizes each image pixel into specific object classes**
- **Comparison:** More detailed than object detection; no bounding boxes, just pixel labels.
- **Purpose:** Provides detailed scene understanding by labeling every pixel.
- **Applications:** Essential for robotics, enabling detailed scene understanding and autonomous navigation.
- **Robot Capabilities:** Helps robots identify obstacles, infer spatial relationships, and interact effectively with their environment.
- **Usecases**
  - Underwater human-robot collaboration
  - Advanced Visual Tasks

Related Works

Cite Key	Work
(1)	End-to-end semantic segmentation on arbitrary-sized inputs
(2)	U-Net architecture for efficient biomedical segmentation
(3)	SegNet improves image segmentation accuracy
(4)	Large dataset for underwater image segmentation
(5)	Attention module enhances neural network features
(6)	Automates coral reef image annotation
(7)	Automates fish detection with CNNs
(8)	MobileNets for mobile and embedded vision applications
(9)	Few-shot learning for underwater image segmentation

Related Works (Contd.)

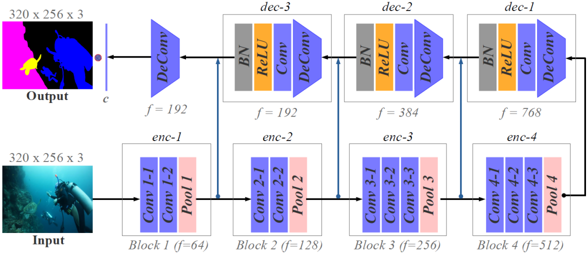

Figure: *SUIM-Net*$_{VGG}$ end-to-end architecture: first four blocks of a pre-trained VGG-16 model are used for encoding.

RELATED WORKS (CONTD.)

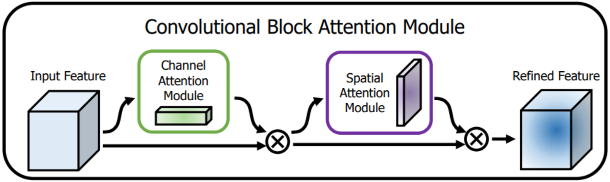

FIGURE: Illustration of a Convolutional Block Attention Module (CBAM). This module sequentially infers attention maps along both channel and spatial dimensions, which refines the feature maps to emphasize informative features and suppress less useful ones, enhancing the model's ability to focus on relevant aspects of the input data.

CHALLENGES

- **Lack of Feature Prioritization:**
  – SUIM-Net architecture uses residual skip connections, to recover spatial details lost during down-sampling in the encoder.
  – Treats all features equally, causing non-essential features to dilute the impact of crucial ones.
- **Handling Varying Object Sizes:**
  – SUIM-Net's fixed-scale feature extraction struggled with scale variability.
  – Smaller objects often missed, and larger objects lacked detailed segmentation.
  – Need to process features at multiple scales
- **Loss of Spatial Information:**
  – Underwater fine-grained textures and subtle differences that require higher resolution features.
  – Both max and average pooling reduce the spatial dimensions of feature maps.

Problem Statement and Proposed Approach

Problem Statement

- Enhance SUIM-Net by incorporating Parallel Dilation Convolution Encoder Blocks, improving its ability to handle varying object sizes while maintaining real-time efficiency.
- Used CBAM in Skip Connections to enhance segmentation quality by minimizing background noise and emphasizing salient objects.

Proposed Approach

Model

Builds upon the SUIM-Net architecture.

Integration with SUIM-Net

- Enhancements include three architectural modifications.
- Serial encoder blocks replaced by Parallel Dilation Convolution Block.
- Uses parallel processing with varying dilation rates.
- Pixel Shuffle Pooling is used in encoders.
- Skip connections augmented with Convolutional Block Attention Module (CBAM).

Proposed Approach (Contd.)

Parallel Dilation Convolution Block

- Configures parallel dilated convolutions and pixel shuffle pooling.
- Handles and processes multi-scale feature maps efficiently.
- Four serial encoders equipped with increasing filters: 64, 128, 256, 512.
- Components of the Block include:
    - Convolution 1: 3x3 kernel, dilation 1. Captures fine details.
    - Convolution 2: 3x3 kernel, dilation 2. Captures medium-scale features.
    - Convolution 3: 3x3 kernel, dilation 4. Captures large-scale features.
- Pixel Shuffle Pooling: Reduces spatial dimensions while aiming to preserve more textual and structural information than traditional pooling methods.

PROPOSED APPROACH (CONTD.)

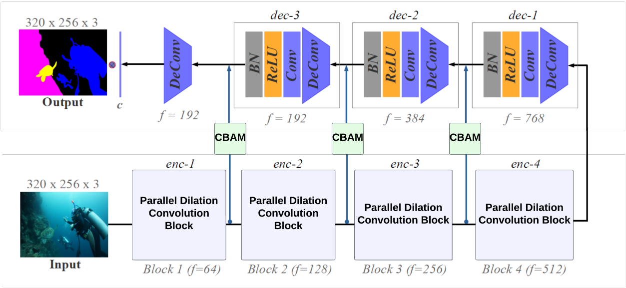

FIGURE: Enhanced *SUIM-Net*$_{VGG}$: Integrating Parallel Dilation Convolution Encoder Blocks with dilations rates = 1, 2, and 4, and CBAM in Skip Connections for Improved Underwater Segmentation.

Proposed Approach (Contd.)

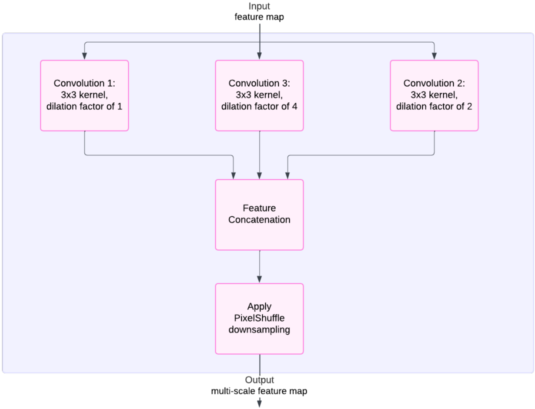

Figure: Parallel Dilation Convolution Block

Conclusion

Implementation:

– Implementation in Google Colab.
– Activation - Sigmoid
– Loss - Binary Cross Entropy
– Adam optimizer

CONCLUSION (CONTD.)

- **F-Score (Dice Coefficient):** The F-Score evaluates the balance between precision and recall for pixel-level segmentation. It is calculated as:

$$F = \frac{2 \cdot P \cdot R}{P + R}$$

where P is precision, and R is recall.

- **Mean Intersection over Union (mIoU):** This metric evaluates the extent of overlap between the predicted regions and the ground-truth regions. It is defined as:

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

The mIoU averages this score across all object categories.

FIGURE: Evaluation Metric

References

[1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2015.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical Image Computing and Computer-Assisted Intervention*, vol. 9351, pp. 234–241, 2015.

[3] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2015.

[4] M. J. Islam, C. Edge, and Y. Xiao, "Semantic understanding of marine imagery (suim): Large-scale dataset for underwater image segmentation," *IEEE Access*, vol. 8, pp. 194 536–194 551, 2020.

[5] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," *European Conference on Computer Vision*, pp. 3–19, 2018.

[6] O. Beijbom, P. J. Edmunds, C. Roelfsema, J. Smith, D. I. Kline, B. P. Neal, M. J. Dunlap, V. Moriarty, T.-Y. Fan, and C.-J. Tan, "Automated annotation of coral reef survey images," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1170–1177, 2012.

[7] M. Ravanbakhsh, H. Mousavi, H. Rastegari, and H. Naji, "Automated fish detection using deep convolutional neural networks," *International Journal of Environmental Science and Technology*, vol. 12, no. 5, pp. 1377–1384, 2015.

[8] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[9] M. Kabir, A. Karim, and M. A. Samad, "Few-shot learning for underwater image segmentation," *Journal of Ocean Engineering*, vol. 4, pp. 231–245, 2023.

Thank You