Underwater Object Recognition

Tommaso Cammelli

Student ID: 23215488

Faculty of Design and Creative Technologies

Auckland University of Technology

COMP 838: Deep Learning

Dr. Wei Qi Yan

5 April, 2024

Abstract

This project proposal investigates the efficacy of various deep learning architectures in the specialized field of underwater object recognition, a critical area of research with significant implications for marine biology, archaeological exploration, and autonomous underwater navigation. Recognizing the unique challenges posed by underwater environments such as low lighting, unclear images, and complex backgrounds. This project aims to perform a comparative analysis of various deep learning models and architecture like Convolutional Neural Networks, Residual Networks and Transformer models to find the most effective strategies for accurate and efficient object recognition in the underwater realm.

By using Python as the primary programming language, the project will use popular deep learning libraries like TensorFlow and PyTorch, for model implementation and training. The methodology focuses on a comprehensive evaluation of model performance, assessing accuracy, precision, recall, and computational efficiency under varying underwater conditions.

Expected outcomes include identifying optimal deep learning architectures that overcome the inherent challenges of underwater imaging.

Underwater Object Recognition

Underwater object recognition has emerged as a crucial field within marine research, underpinning advancements in ecological monitoring, archaeological exploration, and autonomous underwater vehicle navigation. The unique challenges of underwater environments, such as limited visibility, varying light conditions, and the presence of noise, demand robust and adaptive solutions. Deep learning, with its capacity to learn hierarchical representations, offers a promising solution for addressing these challenges.

Recent years have witnessed the development of various deep learning architectures, each with its strengths and limitations when applied to the underwater domain. Convolutional Neural Networks are well known for using spatial hierarchies for effective feature extraction. However, the introduction of architectures like Residual Networks (ResNets) and Transformer models have opened new possibilities for enhancing recognition accuracy and computational efficiency. This project aims to conduct a comparative analysis of these architectures, evaluating their performance in underwater object recognition tasks.

By analyzing the adaptability of these models to underwater conditions, their scalability, and their efficiency in recognizing a diverse range of objects, this project intends to identify optimal strategies for deep learning-based underwater object recognition. The comparison will not only focus on quantitative performance metrics but also consider factors such as model complexity, training data requirements, and real-world applicability.

Literature Review

Underwater object recognition has been subject of increasing interest in the scientific community, driven by the increasing number of potential application in areas such as marine biology, underwater archeology and autonomous underwater navigation.

Deep Learning in Underwater Object Recognition

Deep learning has significantly advanced the field of computer vision, surpassing the capabilities of traditional object detection methods that relied on hand-crafted features for image classification. Initial endeavors predominantly employed Convolutional Neural Networks (CNNs), used for their proficiency in automatically learning and adapting spatial hierarchies of features from images (Zhiqiang & Jun, 2017). For underwater detection CNN were used successfully to recognize objects with significant accuracy, managing to work against the challenges of low-quality images (F. Han et al., 2020)

Advancements and Architectural Innovations

Following the success of CNNs, more sophisticated architectures like Residual Networks (ResNets) have been introduced, allowing training on deeper networks by addressing the vanishing gradient problem (He et al., 2016). ResNets have demonstrated remarkable performance in general object recognition tasks, yet their efficiency and adaptability in underwater conditions remain an area of active research.

The advent of Transformer models, originally designed for natural language processing tasks, has been adapted for computer vision (K. Han et al., 2023). These models, which rely on self-attention mechanisms, offer a new approach to handling the spatial relationships in images, potentially offering advantages in complex underwater scenes where context and object relationships are important.

Challenges in Underwater environments

Recognizing objects underwater presents unique challenges not typically encountered in terrestrial environments. Factors such as variable lighting conditions, water turbidity, and the presence of particulates can significantly impact the performance of deep learning models (Li et al., 2020). Studies have begun to explore the robustness of different architectures under such conditions, emphasizing the need for models that can adapt to or correct for these environmental distortions.

Comparative Analyses in the Literature

Comparative studies specifically addressing the performance of deep learning architectures in underwater object recognition are sparse but growing. These studies are crucial for understanding the practical limitations and opportunities of applying these advanced computational models to underwater scenarios. For instance, a study by Teng

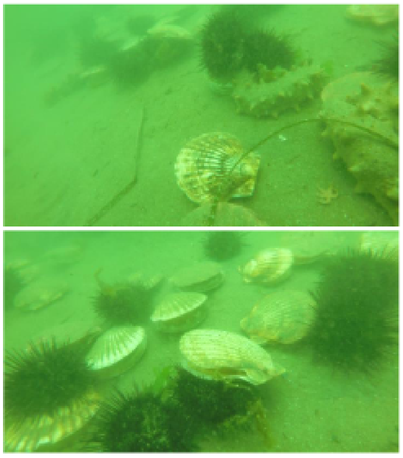

Figure 1

Input underwater Image

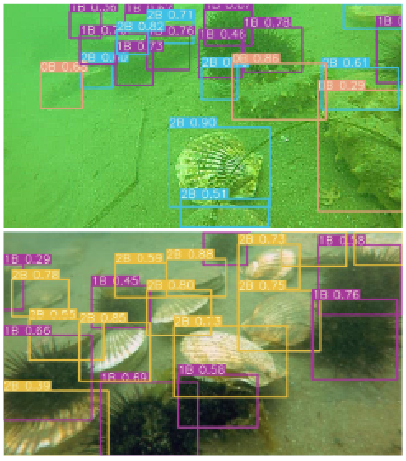

Figure 2

Enhanced Image with recognized objects

and Zhao (2020) compared the accuracy and computational efficiency of CNNs and ResNets in identifying underwater mines, highlighting the trade-offs between model complexity and recognition performance.

Methodology

Data Collection

In this project publicly available underwater image datasets will be used, such as the Underwater Object Detection Dataset (UODD) (Jiang et al., 2021) [1]. These datasets contain a diverse range of underwater scenes, including various flora, fauna, and man-made objects, providing a comprehensive basis for testing and comparison, along with annotations labeled in MS COCO format.

Given the inherent challenges of underwater imaging, such as varying light conditions and turbidity, images will undergo pre-processing to enhance quality and consistency. Techniques like color correction, contrast adjustment, and noise reduction will be applied to mitigate environmental effects on image quality, like shown in (Teng & Zhao, 2020) and (Peng et al., 2023).

[1] UODD dataset is available at https://github.com/LehiChiang/Underwater-object-detection-dataset

Model Selection and Development

- **CNNs**: Initial experiments will focus on conventional CNN architectures, which have proven effective in basic underwater object detection tasks. These models will serve as a benchmark for comparing more advanced architectures.
- **Residual Networks (ResNets)**: Given their ability to train deeper networks by mitigating the vanishing gradient problem, ResNets will be explored for their potential to improve recognition accuracy in complex underwater scenes.
- **Transformers**: The study will also incorporate Transformer models, which utilize self-attention mechanisms, to examine their effectiveness in capturing the spatial relationships of objects in underwater images.

Implementation

The project will employ Python as the primary programming language, leveraging its extensive ecosystem and the ease of finding pre-implemented models along with Jupyter notebook to help the visualization of data. For model implementation and training, I will utilize deep learning libraries such as TensorFlow and PyTorch. Experiments will be conducted on a laptop equipped with an Nvidia GPU to facilitate efficient model training and evaluation.

Evaluation Criteria

The evaluation of the various models will use common indicators like accuracy, precision, recall and F1 score along with the confusion matrix to determine the effectiveness of each model in correctly identifying and classifying underwater objects.

		Prediction outcome		
		Positive	Negative	Total
Actual values	Positive	TP	FN	P'
	Negative	FN	TN	N'
	Total	P	N	

Figure 3

Confusion Matrix used for model evaluation

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \tag{1}$$

Additionally, computational efficiency, measured in terms of training time and inference speed, will be considered to assess the practicality of deploying these models in real-world applications.

The performance of each architecture will be compared to establish their relative strengths and weaknesses in underwater object recognition. This analysis will also explore the impact of varying dataset complexities and environmental conditions on model performance.

Concluding Remarks

This project proposal outlines a comprehensive approach to exploring the potential of various deep learning architectures in the challenging domain of underwater object recognition. By using Python with the capabilities of TensorFlow, PyTorch, and the interactive environment of Jupyter Notebooks, this study aims to not only compare the efficacy of established models such as CNNs and ResNets but also investigate the emerging potential of Transformer models in this context.

This project could potentially contribute to the field of marine biology, archaeological exploration and autonomous underwater navigation by identifying optimal deep learning strategies that can help with the unique challenges posed by the underwater environments.

References

Han, F., Yao, J., Zhu, H., & Wang, C. (2020). Underwater Image Processing and Object Detection Based on Deep CNN Method. *Journal of Sensors*, *2020*, e6707328. https://doi.org/10.1155/2020/6707328

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tang, Y., Xiao, A., Xu, C., Xu, Y., Yang, Z., Zhang, Y., & Tao, D. (2023). A Survey on Vision Transformer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(1), 87–110. https://doi.org/10.1109/TPAMI.2022.3152247

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770–778. https://doi.org/10.1109/CVPR.2016.90

Jiang, L., Wang, Y., Jia, Q., Xu, S., Liu, y., Fan, X., Li, H., Liu, R., Xue, X., & Wang, R. (2021). Underwater Species Detection using Channel Sharpening Attention, 4259–4267. https://doi.org/10.1145/3474085.3475563

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., & Tao, D. (2020). An Underwater Image Enhancement Benchmark Dataset and Beyond. *IEEE Transactions on Image Processing*, *29*, 4376–4389. https://doi.org/10.1109/TIP.2019.2955241

Peng, L., Zhu, C., & Bian, L. (2023). U-shape Transformer for Underwater Image Enhancement. *IEEE Transactions on Image Processing*, *32*, 3066–3079. https://doi.org/10.1109/TIP.2023.3276332

Teng, B., & Zhao, H. (2020). Underwater target recognition methods based on the framework of deep learning: A survey. *International Journal of Advanced Robotic Systems*, *17*(6), 1729881420976307. https://doi.org/10.1177/1729881420976307

Zhiqiang, W., & Jun, L. (2017). A review of object detection based on convolutional neural network. *2017 36th Chinese Control Conference (CCC)*, 11104–11109. https://doi.org/10.23919/ChiCC.2017.8029130