

## Article

# Urban Aquatic Scene Expansion for Semantic Segmentation in Cityscapes

Zongcheng Yue <sup>1,\*</sup>, Chun-Yan Lo <sup>1</sup>, Ran Wu <sup>2</sup>, Longyu Ma <sup>1</sup> and Chiu-Wing Sham <sup>1</sup><sup>1</sup> School of Computer Science, The University of Auckland, Auckland Central, Auckland 1010, New Zealand; b.sham@auckland.ac.nz (C.-W.S.)<sup>2</sup> School of Computer Science, Harbin Institute of Technology, Harbin 150001, China

\* Correspondence: zyue743@aucklanduni.ac.nz

**Abstract:** In urban environments, semantic segmentation using computer vision plays a pivotal role in understanding and interpreting the diverse elements within urban imagery. The Cityscapes dataset, widely used for semantic segmentation in urban scenes, predominantly features urban elements like buildings and vehicles but lacks aquatic elements. Recognizing this limitation, our study introduces a method to enhance the Cityscapes dataset by incorporating aquatic classes, crucial for a comprehensive understanding of coastal urban environments. To achieve this, we employ a dual-model approach using two advanced neural networks. The first network is trained on the standard Cityscapes dataset, while the second focuses on aquatic scenes. We adeptly integrate aquatic features from the marine-focused model into the Cityscapes imagery. This integration is carefully executed to ensure a seamless blend of urban and aquatic elements, thereby creating an enriched dataset that reflects the realities of coastal cities more accurately. Our method is evaluated by comparing the enhanced Cityscapes model with the original on a set of diverse urban images, including aquatic views. The results demonstrate that our approach effectively maintains the high segmentation accuracy of the original Cityscapes dataset for urban elements while successfully integrating marine features. Importantly, this is achieved without necessitating additional training, which is a significant advantage in terms of resource efficiency.

**Keywords:** deep learning; semantic segmentation; Cityscapes; aquatic images



**Citation:** Yue, Z.; Lo, C.-Y.; Wu, R.; Ma, L.; Sham, C.-W. Urban Aquatic Scene Expansion for Semantic Segmentation in Cityscapes. *Urban Sci.* **2024**, *8*, 23. <https://doi.org/10.3390/urbansci8020023>

Academic Editor: Luis Hernández-Callejo

Received: 7 December 2023

Revised: 8 February 2024

Accepted: 13 March 2024

Published: 22 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Machine learning plays a pivotal role in urban planning and development [1–3]. Particularly, semantic segmentation [4,5] can serve as a foundational technology in applications ranging from smart city design to environmental monitoring. This process involves classifying each pixel in an image to demarcate distinct regions with semantic relevance, thus facilitating a detailed understanding of urban landscapes [6,7]. While the Cityscapes dataset [8] is a benchmark for assessing semantic segmentation algorithms in urban settings, it primarily focuses on terrestrial urban environments and lacks representation of aquatic regions [9,10]. This gap is significant in urban science, particularly for coastal cities and waterway management [11].

Retraining the Cityscapes dataset to include aquatic environments poses significant challenges, including high computational demands and extensive time requirements [12,13]. Simplification techniques, such as fixed-point calculations [14] or binary neural networks [15–17], offer some relief, but their impact is marginal given the scale and complexity of the data [18].

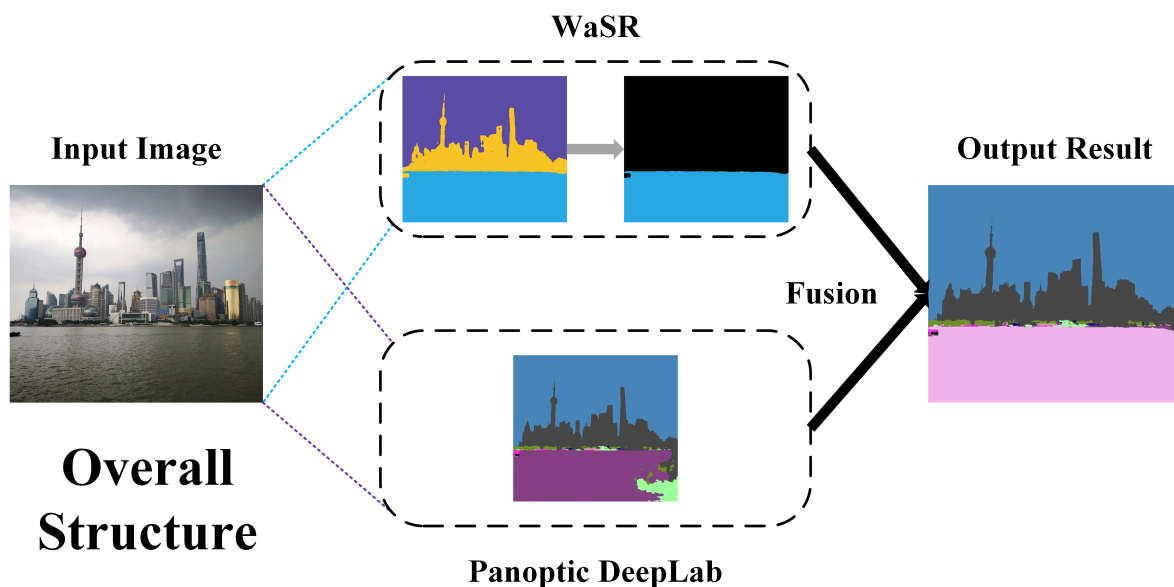
Addressing this limitation, our study introduces a novel fusion model that synergizes the capabilities of two advanced neural networks: Panoptic DeepLab [19] and Water Segmentation and Refinement (WaSR) [20]. Panoptic DeepLab, trained on the Cityscapes dataset, excels in urban landscape segmentation; however, as demonstrated by our experiments in this paper, it falls short in identifying aquatic regions due to dataset limitations.

Conversely, WaSR specializes in aquatic environment segmentation but lacks the diversity to classify varied terrestrial urban elements.

A fusion model [21] is proposed to combine the strengths of both networks, enabling comprehensive segmentation that includes aquatic areas without the need for retraining. The model is enhanced and the approach is illustrated in our model architecture diagram (Figure 1). The model first employs WaSR to segment aquatic regions in the input image and then integrates these results with the output from Panoptic DeepLab. This hybrid approach allows for accurate and holistic labeling of urban scenes, including both terrestrial and aquatic elements, which is essential for urban environmental studies and sustainable city planning. In general, our contributions are as follows:

- **Efficient and resource-saving methodology:** Our fusion model efficiently labels images with mixed urban and aquatic environments using the pretrained models Panoptic DeepLab and WaSR. This approach eliminates the need for retraining new networks, conserving significant time and computational resources.
- **Versatile application across various environments:** The model's adaptability extends its application beyond the Cityscapes dataset, enhancing semantic segmentation performance in diverse fields such as environmental monitoring and underwater exploration.
- **Comprehensive analysis of complex urban ecosystems:** By integrating urban and aquatic segmentation, the model offers a more nuanced understanding of urban landscapes, particularly beneficial in coastal or riverine cities where water bodies are integral.

The organization of this paper is as follows: Section 2 presents a review of the latest segmentation works. In Section 3, we provide a comprehensive overview of two neural network models: Panoptic Deeplab and WaSR. Following that, we will delve into the fusion process that combines these two models. The results of our experiments are presented in a visualized manner in Section 4. Finally, we summarize our findings and conclusions in Section 5.



**Figure 1.** Overall structure.

## 2. Literature Review

The existing literature reflects a diverse array of segmentation algorithms, each tailored for specific urban and environmental applications. Some research studies [22,23] focused on sea-sky scene perception, crucial for coastal urban monitoring, handling dual subtasks in a single inference process. In urban context understanding, some researchers [24,25] introduced two hierarchical frameworks that efficiently incorporate contextual information for enhanced semantic segmentation, crucial for complex urban environments.

Another work [26] contributed significantly to maritime urban science by creating MaSTr1325, a marine semantic segmentation training dataset, alongside the MODS benchmark [27]. MODS focuses on maritime object detection and obstacle segmentation, key aspects in developing safe and efficient marine navigation systems in urban coastal areas.

Previous research studies [28,29] proposed two methods for detecting the sea–sky line under complex backgrounds, a tool essential for urban coastal surveillance. Enhancing this, Bovcon et al. [20] innovated a method that merges inertial data with visual features for improved water edge detection, vital for urban planning in areas where land meets water.

Later, some researchers [30,31] extended the urban landscape segmentation by designing a network that classifies pixels into water, land, or sky using a CRF method. This is particularly relevant for urban areas with diverse landscapes, from water bodies to green spaces [32].

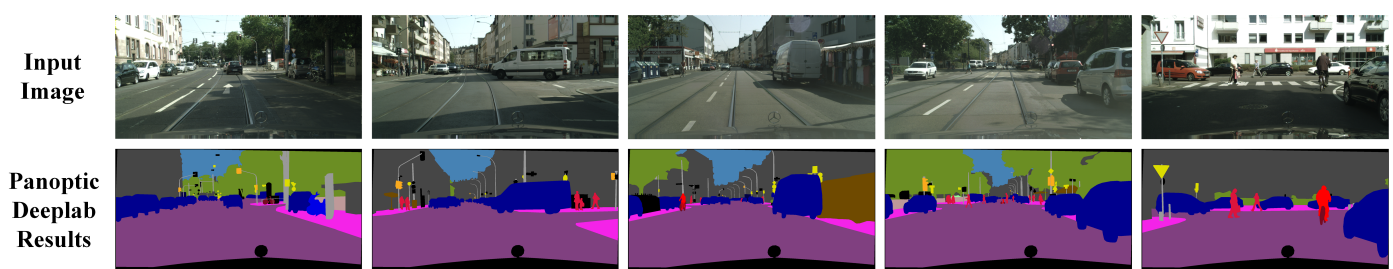
In the realm of detailed urban scene analysis, the researchers [33,34] focused on differentiating foreground and background elements, enhancing object detection in densely populated urban areas. Some researchers [35,36] introduced two methods to refine panoptic segmentation, a step forward in creating coherent urban maps. Further, researchers [37,38] developed two models optimized with point-based supervision, improving accuracy in urban feature detection. Gasperini et al. [39] presented Panoster, a segmentation method for urban LiDAR data, vital for 3D urban modeling. The researchers also [40,41] proposed a methodology based on a dual-encoder network to process RGB and depth data, enhancing 3D urban scene perception.

These advancements demonstrate the growing versatility and applicability of semantic segmentation algorithms in urban science, from coastal urban monitoring to detailed 3D urban modeling. The incorporation of environmental elements like water and sky into urban scene understanding, as seen in these works, is crucial for comprehensive urban planning and management.

### 3. Methodology

#### 3.1. Panoptic DeepLab Trained on Cityscapes

Figure 2 presents the results of Panoptic Deeplab applied to the Cityscapes dataset. Cityscapes comprises a substantial variety of distinct labels. However, Panoptic Deeplab demonstrates its ability to successfully recognize and classify all of these labels, due to its efficient computational architecture, following thorough training.



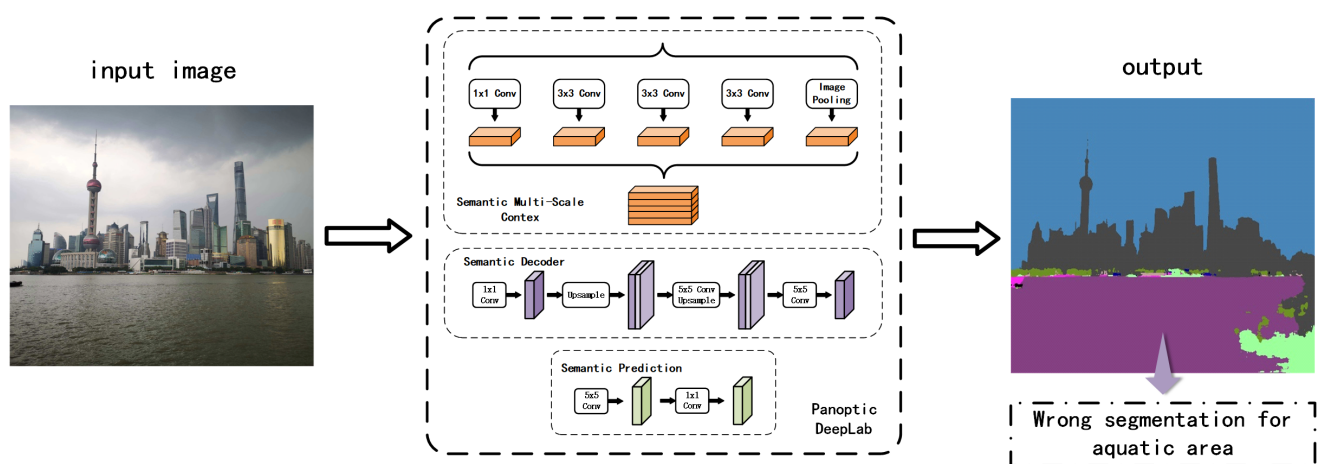
**Figure 2.** Results of Panoptic Deeplab.

Figure 3 illustrates the structure of Panoptic DeepLab. Panoptic-DeepLab stands out in the realm of image segmentation with its innovative dual-atrious spatial pyramid pooling (ASPP) and dual-decoder modules [42,43]; this structure is specifically designed to tackle the intricacies of both semantic and instance segmentation tasks within the broader scope of panoptic segmentation. This advanced architecture is bolstered by a shared backbone [44], which serves both segmentation tasks simultaneously. This not only optimizes computational resource usage but also ensures the extraction of rich feature representations that are equally beneficial for semantic and instance segmentation. A distinctive feature of Panoptic-DeepLab is its approach to instance segmentation through class-agnostic instance center regression. This method deviates from traditional top-down approaches [45] that

typically rely on region proposal networks. Instead, it directly predicts the center of each object instance and accurately computes the offset for each pixel within that instance, thus pinpointing its precise location [46]. The semantic segmentation branch of Panoptic-DeepLab aligns with conventional semantic segmentation models, focusing on classifying each pixel into various categories, including “things” and “stuff”. Complementing these features is the model’s efficient merging operation. In summary, the main functionality of Panoptic-DeepLab is as follows:

- The dual-ASPP and dual-decoder modules enable the network to handle the intricacies of both semantic segmentation (categorizing areas into broad classes) and instance segmentation (identifying individual object instances).
- The shared backbone allows for feature extraction that is beneficial for both segmentation tasks, maximizing the use of learned features.
- The instance center regression facilitates the bottom-up approach for instance segmentation, identifying individual objects without needing region proposals.
- The efficient merging operation combines the outputs of both segmentation tasks to create a cohesive panoptic segmentation map, integrating both “thing” (individual objects) and “stuff” (amorphous regions like grass or sky) categories.

In addition, Panoptic-DeepLab distinguishes itself with its simplicity, speed, efficiency, and state-of-the-art performance, setting a new benchmark in the field of panoptic segmentation. The architecture of the model is ingeniously crafted to be less complex yet robust, offering a simpler alternative to the more intricate two-stage methods commonly found in image segmentation. This simplicity not only facilitates ease of implementation and modification but also enhances its appeal for a broader range of applications. A key strength of Panoptic-DeepLab lies in its speed and efficiency, attributes that stem from its streamlined structure and the use of a shared backbone. This makes it particularly well suited for real-time applications, where quick processing is essential. In terms of performance, Panoptic-DeepLab excels, consistently achieving competitive or leading results across various renowned benchmarks, including Cityscapes, Mapillary Vistas, and COCO. This high level of performance is a testament to the model’s effectiveness in handling diverse segmentation tasks. Furthermore, its bottom-up approach in panoptic segmentation simplifies the process while maintaining high-quality output.



**Figure 3.** DeepLab based on Cityscapes. In our analysis, we found that while the DeepLab neural network performs well in detecting various environmental instances in the Cityscapes dataset, it struggles with identifying aquatic areas. Our experiments showed that the model incorrectly segmented some aquatic areas as roads, streets, or other incorrect classes. This highlights the need for a more specialized neural network to accurately identify and segment aquatic regions in an image.

The performance of neural networks, particularly in the domain of computer vision, is fundamentally tied to the caliber and diversity of the training data they are exposed to.



This principle is exemplified in the case of Panoptic DeepLab, a cutting-edge model in the field of panoptic segmentation. A key contributor to its success is the Cityscapes dataset, an extensive collection of stereo video sequences meticulously captured in street scenes from 50 different urban environments. This dataset is not just voluminous but rich in quality and variety, comprising high-quality pixel-level annotations of 5000 frames along with a substantial set of 20,000 frames with weaker annotations. This exhaustive dataset encompasses an extensive range of urban object classes, including cars, pedestrians, bicycles, and buildings, each presenting unique challenges in terms of segmentation and recognition.

Cityscapes' detailed and diverse dataset serves as a crucial benchmark for the development and evaluation of advanced semantic segmentation algorithms. The dataset's complexity and real-world variability make it an ideal proving ground for models intended for intricate tasks in computer vision, such as panoptic segmentation and object tracking. The depth and breadth of its data contribute significantly to the training of models, enabling them to learn and accurately identify a wide range of objects and scenarios typical of urban landscapes.

Trained on the Cityscapes dataset, Panoptic DeepLab has demonstrated exceptional proficiency, achieving state-of-the-art performance across several metrics. This high level of accuracy is particularly evident in the model's panoptic segmentation of city views, where it successfully differentiates and segments a multitude of elements within dense urban scenes. Table 1 clearly illustrates that the Cityscapes dataset encompasses a wide array of living-thing categories; however, it notably lacks labeled categories for aquatic environments. This limitation hinders its ability to efficiently process data related to aquatic areas, which, significantly, constitute the majority of the Earth's surface.

**Table 1.** The legend of Cityscapes.

ID	Name	RGB Value	Color Name
0	unlabeled	0, 0, 0	Black
1	ego vehicle	0, 0, 0	Black
2	rectification border	0, 0, 0	Black
3	out of roi	0, 0, 0	Black
4	static	0, 0, 0	Black
5	dynamic	111, 74, 0	Dark brown
6	ground	81, 0, 81	Purple
7	road	128, 64, 128	Medium purple
8	sidewalk	244, 35, 232	Bright pink
9	parking	250, 170, 160	Light salmon
10	rail track	230, 150, 140	Light taupe
11	building	70, 70, 70	Dim gray
12	wall	102, 102, 156	Slate gray
13	fence	190, 153, 153	Dusty rose
14	guard rail	180, 165, 180	Grayish-pink
15	bridge	150, 100, 100	Sienna
16	tunnel	150, 120, 90	Olive
17	pole	153, 153, 153	Gray
18	polegroup	153, 153, 153	Gray
19	traffic light	250, 170, 30	Yellow
20	traffic sign	220, 220, 0	Electric yellow
21	vegetation	107, 142, 35	Olive green
22	terrain	152, 251, 152	Pale green
23	sky	70, 130, 180	Steel blue
24	person	220, 20, 60	Crimson
25	rider	255, 0, 0	Red
26	car	0, 0, 142	Deep Blue
27	truck	0, 0, 70	Navy Blue

Table 1. Cont.

ID	Name	RGB Value	Color Name
28	bus	0, 60, 100	Dark cerulean
29	caravan	0, 0, 90	Dark blue
30	trailer	0, 0, 110	Dark midnight blue
31	train	0, 80, 100	Dark teal
32	motorcycle	0, 0, 230	Blue
33	bicycle	119, 11, 32	Maroon
34	license plate	0, 0, 142	Deep blue

### 3.2. WaSR

Cityscapes has established itself as a popular and valuable dataset for urban scene segmentation [47]. It possesses inherent limitations, particularly in its applicability to environments beyond urban landscapes. One notable area where Cityscapes falls short is in the segmentation of aquatic environments [48]—a domain vastly different from urban settings in terms of visual features and segmentation challenges [49]. Recognizing this gap, researchers [20] started to develop a specialized model tailored for aquatic environments, leading to the creation of the WaSR model.

The WaSR network structure is a specialized deep learning architecture designed for maritime obstacle detection, with several distinctive features and functionalities. Below is a detailed breakdown of its network structure, functionalities, and advantages:

#### 1. Encoder–decoder architecture:

- (a) **Encoder:** Based on ResNet101 with atrous convolutions for extracting rich visual features.
- (b) **Decoder:** Integrates features from the encoder, upsampling them to construct the segmentation map. Includes multiple fusion modules for handling various water appearances.

#### 2. Fusion modules:

- Essential for combining visual features with inertial measurements, addressing maritime challenges such as reflections and wakes.

#### 3. Inertial measurement unit (IMU) integration:

- Incorporates IMU data to determine the horizon line and camera orientation, enhancing accuracy in ambiguous conditions.

#### 4. IMU feature channel encoding:

- (a) Encoding methods: Drawing a horizon line, encoding a signed distance to the horizon, and creating a binary mask below the horizon.
- (b) These encoded channels are fused into the decoder for improved segmentation accuracy.

The WaSR network is meticulously designed for precise semantic segmentation in maritime environments, a pivotal functionality for autonomous navigation and surveillance in marine settings. Its primary capability lies in its adeptness at distinguishing various elements within a maritime scene, such as water, sky, ships, and other pertinent objects. This specificity is crucial, as it directly informs navigation decisions for unmanned marine vehicles, ensuring safety and efficiency in navigation.

An essential aspect of WaSR's functionality is its robustness in handling the complex appearances of water. Maritime environments are inherently dynamic, with varying conditions such as different lighting scenarios, reflections, and diverse water textures. WaSR's sophisticated architecture can navigate these challenges, ensuring accurate segmentation even under these fluctuating conditions.

A standout feature of the WaSR network is its integration of data from an inertial measurement unit (IMU). This integration is not just a supplementary enhancement but a core aspect of its functionality. The IMU data play a critical role in accurately determining

the horizon line and the camera's orientation relative to the horizon. This is particularly vital in the ambiguous visual conditions often encountered at sea, such as foggy or glare-heavy scenarios. By fusing this inertial data with visual cues, WaSR achieves a higher level of precision in horizon detection and orientation, which is instrumental in correctly interpreting maritime scenes.

Moreover, the network's ability to incorporate IMU data for horizon estimation and to adaptively respond to various sea states showcases its advanced approach to maritime scene understanding [50,51]. This results in a significant reduction in false positives, a common challenge in water segmentation due to the reflective and dynamic nature of marine environments. The WaSR network presents several substantial advantages that set it apart in the field of maritime obstacle detection and navigation. These advantages underscore its potential as a transformative tool for a wide array of maritime applications.

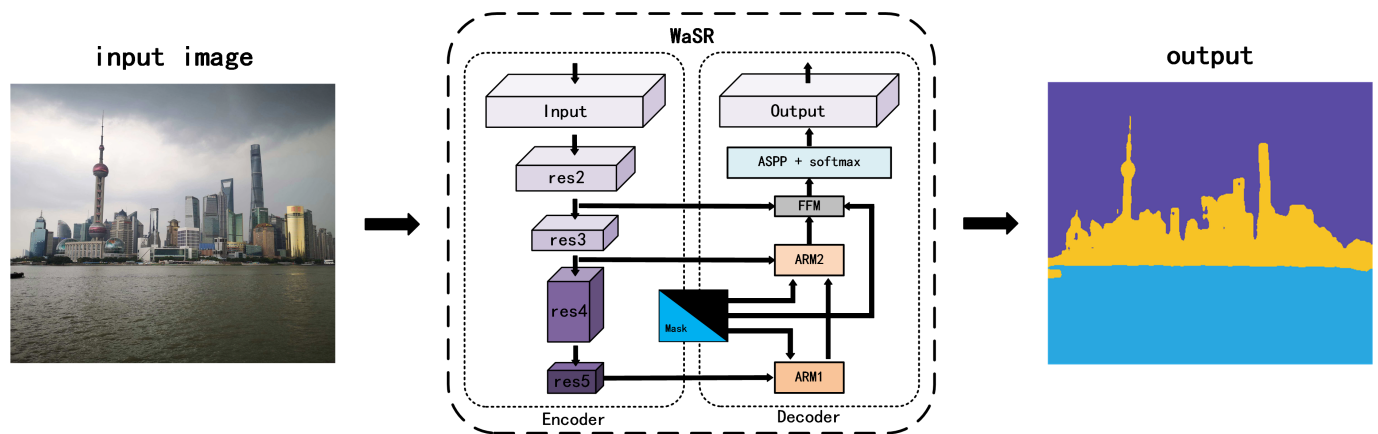
One of the primary advantages of the WaSR network is its remarkable capability to reduce false positives. In the context of maritime environments, where the reflective and dynamic nature of water often leads to visual ambiguities [52,53], reducing false positives is crucial. Traditional segmentation methods can struggle with differentiating between actual obstacles and reflections or other water-related phenomena [54–56]. WaSR, with its sophisticated fusion of visual and inertial data, excels in accurately distinguishing between these elements. This precision is particularly beneficial in ensuring the safety and efficiency of autonomous marine navigation, where accurate detection of obstacles is vital.

Another significant advantage of WaSR is its impressive generalization capabilities. The network has been tested and has shown commendable performance across various datasets and hardware setups. This ability to generalize ensures that WaSR is not limited to the specific conditions or environments it was trained, making it a versatile and robust tool for maritime obstacle detection. Whether deployed on different types of unmanned surface vehicles or in varied geographical locations, WaSR maintains a consistent level of accuracy and reliability.

Additionally, WaSR is exceptionally robust in challenging conditions. Maritime environments can be highly unpredictable, with factors like fog, glare, and varying light conditions often impeding visibility [57–59]. WaSR's design and integration of IMU data make it adept at navigating these challenges, ensuring accurate segmentation even in less-than-ideal visual conditions. This robustness enhances the network's applicability in real-world scenarios, where such conditions are commonplace.

Figure 4 illustrates the main procedure of the WaSR model, showcasing the intricate process of how the model handles aquatic imagery. The model's procedure involves a series of steps that include initial segmentation, feature extraction, and refinement stages, each meticulously designed to address the unique challenges posed by water bodies. The WaSR network presents several substantial advantages that set it apart in the field of maritime obstacle detection and navigation. These advantages underscore its potential as a transformative tool for a wide array of maritime applications.

A primary limitation of the WaSR model is its specialization for maritime environments, which inherently restricts its effectiveness in nonmaritime contexts as shown in Table 2, which only includes three different items, obstacle, water, and sky. Its algorithms and data processing techniques are optimized for water, sky, and marine obstacles shown in Figure 5, which means it might not yield the same level of accuracy or reliability in terrestrial or aerial environments. This specialization, while a strength in marine settings, limits its versatility across diverse environmental applications, especially for large and complex terrestrial environments.



**Figure 4.** WaSR. WaSR is a neural network with an input and output size of  $512 \times 384 \times 3$ . It consists of an encoder module that includes four neural networks with a ResNet-101 backbone, combined with max pooling layers. In the decoder, attention refinement modules (ARMs) are used to learn an optimal fusion strategy. Additionally, the feature fusion module (FFM) is used to extract low-level and high-level features and combine them in the CNNs. The atrous spatial pyramid pooling (ASPP) module is used to improve the segmentation performance for small structures while generating minimal computing load. These design choices make WaSR a powerful and efficient segmentation model for aquatic environments.



**Figure 5.** Results of WaSR.

**Table 2.** The legend of WaSR.

ID	Name	RGB Value	Color Name
0	Obstacle	247, 195, 37	Goldenrod
1	Water	41, 167, 224	Dodger blue
2	Sky	90, 75, 164	Dark purple

Another significant challenge is the requirement for substantial training data [60–63]. To achieve its high level of accuracy, WaSR needs to be trained on extensive datasets that comprehensively cover various maritime scenarios and conditions. Gathering such large and diverse datasets can be resource-intensive and may not be feasible for all applications, especially those with limited access to maritime environments or those operating under constrained research budgets. The model's demand for significant computational resources is also a drawback. To process complex datasets and perform real-time segmentation and detection, WaSR requires powerful processing capabilities. This requirement can pose a barrier to its deployment on systems with limited computational power or in scenarios where minimizing power consumption is critical, such as on unmanned, battery-operated marine vehicles.

In terms of operational limitations, the WaSR model shows sensitivity to lighting conditions, particularly in low-light environments. Its performance can decrease under such conditions, as the visual sensors may not capture enough detail to accurately differentiate

between various elements in the scene. This sensitivity could be a hindrance in operations conducted during nighttime or in areas with poor visibility [64–66]. Lastly, the model's ability to detect small obstacles is an area of concern, especially in safety-critical applications. While WaSR excels in identifying larger objects, it may sometimes miss smaller obstacles, which, in a maritime setting, can be just as hazardous as larger ones. This limitation necessitates additional caution and possibly supplementary detection systems to ensure comprehensive safety in navigation.

### 3.3. Model Fusion

In our endeavor to refine segmentation for both aquatic and terrestrial environments, we innovated a fusion model that capitalizes on the distinct strengths of the WaSR and Panoptic DeepLab models. This model is designed to overcome each system's specific limitations when dealing with complex environmental scenes that include both land and water.

The aquatic model excels in segmenting aquatic areas but falls short in providing intricate details for land segments. Conversely, Panoptic DeepLab, trained with the extensive Cityscapes dataset, offers comprehensive labeling for land features but cannot segment aquatic regions effectively. Our fusion model aims to harness these disparate capabilities for a more robust segmentation solution.

The process begins with a standard preprocessing step where an input image is resized to a uniform dimension, making it compatible with analysis by the two models. The aquatic model processes the image to segment it into three categories: aquatic, sky, and other. Following this, we employ a custom mask that filters out the 'sky' and 'other' segments, isolating the aquatic region in the image. Concurrently, Panoptic DeepLab performs its segmentation, producing detailed split instance graphs based on its terrestrial-focused training labels. However, it is important to note that Panoptic DeepLab's output for aquatic areas is not reliable due to the absence of such labels in its training set.

The challenge then lies in effectively merging these outputs to produce a cohesive segmented map. To address this, our fusion model initially discards the aquatic segment from Panoptic DeepLab's output, acknowledging its inherent inaccuracy. We replace this segment with the aquatic mask generated by the aquatic result, ensuring precise delineation of water bodies. However, a critical issue arises in extracting nonaquatic segments from Panoptic DeepLab's output, given its inaccuracy in classifying aquatic regions. To resolve this, we introduce an innovative color-based tag finder. This tool analyzes the RGB output from the WaSR model, focusing on identifying the blue hues that correspond to aquatic areas. By accurately pinpointing these regions, we can seamlessly integrate the detailed land segmentation from Panoptic DeepLab with the aquatic segmentation from the other model.

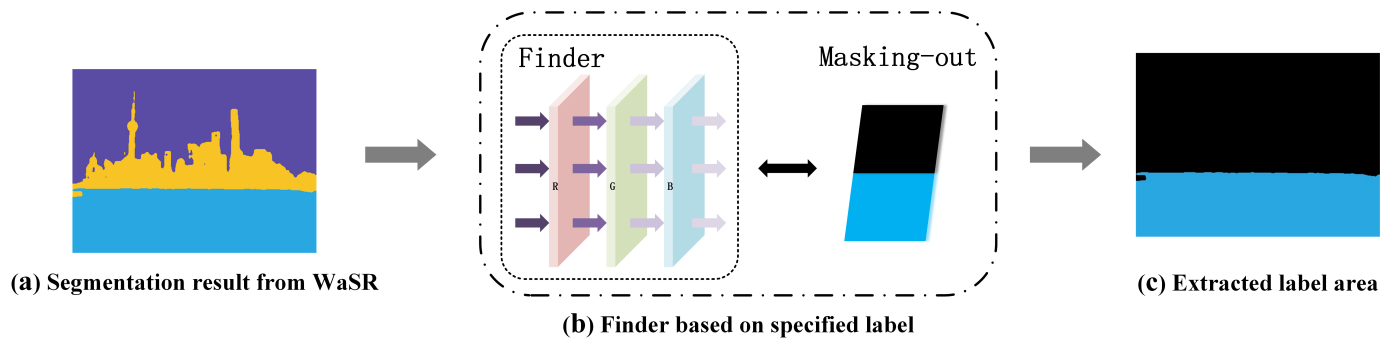
Through this sophisticated fusion approach, our model effectively combines the detailed terrestrial segmentation of Panoptic DeepLab with the precise aquatic delineation of WaSR. This results in a comprehensive and accurate representation of both land and water environments, significantly enhancing the capabilities of environmental segmentation and analysis.

Our novel fusion model adeptly merges the segmentation capabilities of Panoptic DeepLab and WaSR, resulting in a system that can accurately segment both land and aquatic areas by utilizing the strengths of both models. The comprehensive structure and workflow of this fusion model are depicted in Figure 1, illustrating how the outputs of the two models are integrated to achieve superior segmentation performance.

In our fusion model, a critical component is the color-based tag finder, whose structure and function are detailed in Figure 6. This finder is tasked with processing the RGB images from the label classification results of the aquatic elaboration. These images are then converted into LAB value images, a format that significantly enhances color differentiation, making it easier to isolate specific color ranges. The finder operates by pinpointing and extracting the areas that match the predefined aquatic color values. The resulting mask, as shown in Figure 6b, uses black to denote the extracted aquatic area. This mask is a pivotal

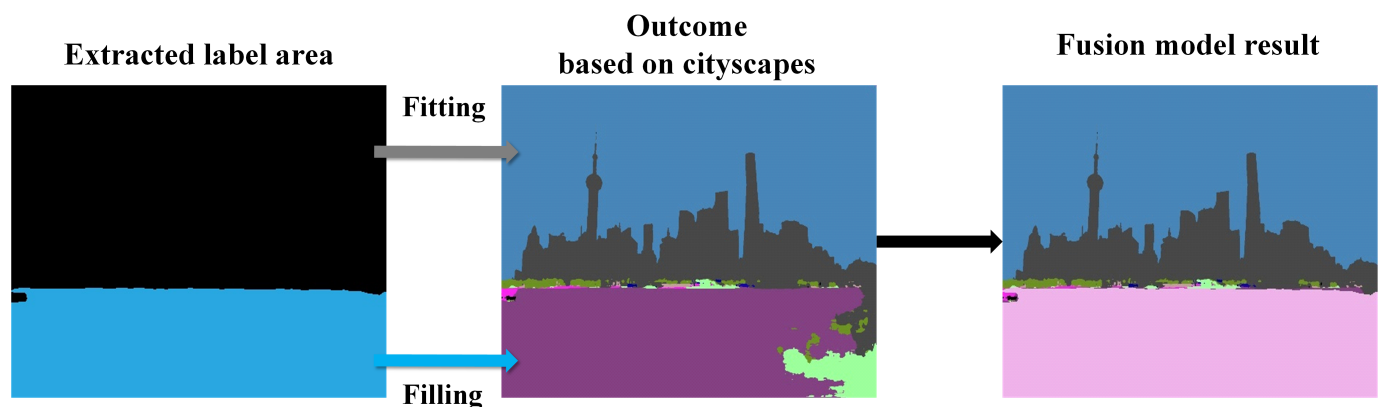


element in the fusion process, allowing us to overlay it onto the combined output of WaSR and Panoptic DeepLab, effectively replacing inaccurately segmented areas.



**Figure 6.** Extracting label.

The fusion process itself is visually elucidated in Figure 7. Here, we demonstrate how the extracted mask image from Figure 6 and the output from Panoptic DeepLab are blended. In this synthesis, the split map from Panoptic DeepLab automatically aligns with the corresponding areas on the mask. Simultaneously, the remaining areas in the mask, representing aquatic segments, are overlaid onto the final output map. This method ensures that the final output not only maintains WaSR's high accuracy in identifying aquatic areas but also benefits from the diverse and precise labeling of terrestrial features provided by Cityscapes.



**Figure 7.** Model fusion.

Moreover, we introduced a color modification step for the extracted labeled areas. This alteration is implemented to enhance the visual distinction of these areas, making it easier to observe and analyze the segmented results. By modifying the color of these areas, we provide a clearer visual demarcation, which is especially useful in applications requiring quick identification and differentiation of various segments.

In summary, our fusion model represents a significant advancement in environmental segmentation, combining the aquatic accuracy of the aquatic model result with the extensive land labeling capabilities of Panoptic DeepLab. The model not only achieves high accuracy in segmenting diverse environments but also presents the results in a visually intuitive manner, enhancing both the usability and applicability of the segmented data.

#### 4. Evaluation

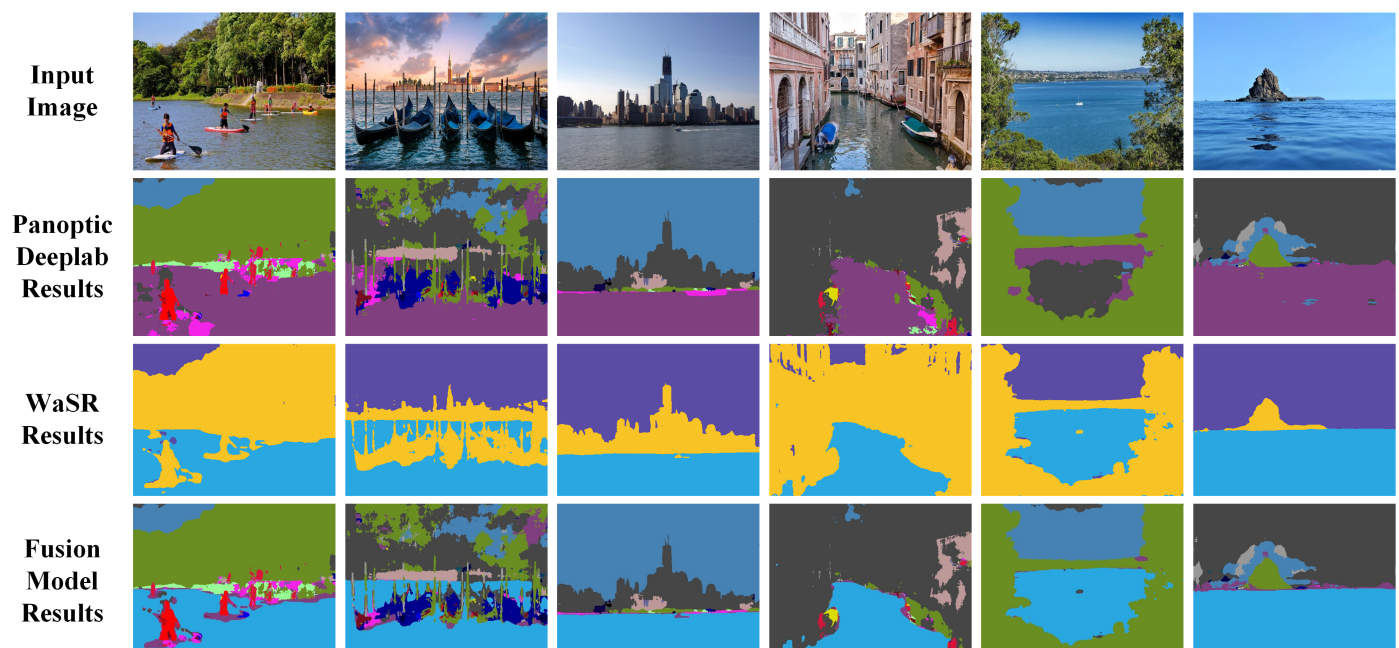
In this section, we delve into the practical application and testing of our fusion model, implemented using Google Colab's CPU environment. The focus of our testing was to assess the model's ability to effectively combine aquatic and terrestrial data in image

segmentation. We selected a diverse set of images, each containing elements of both aquatic and terrestrial environments, to evaluate the model's performance.

The initial step involved setting up the computational environment in Google Colab for both the WaSR and Panoptic DeepLab neural networks. This setup included loading their respective pretrained models, a crucial step to leverage their advanced segmentation capabilities. To ensure compatibility with each network, we resized the input images according to the specific requirements of WaSR and Panoptic DeepLab. These resized images were then processed through each network to generate their segmentation predictions.

Following the segmentation, we undertook a critical step of size normalization to align the outputs of both networks. This alignment was essential for the subsequent fusion process. We extracted the aquatic segment from the WaSR output using the previously discussed masking technique. This extracted segment was then meticulously integrated with the output from Panoptic DeepLab, resulting in a composite image that effectively combined the detailed terrestrial data from Panoptic DeepLab with the precise aquatic segmentation from WaSR.

The results of this fusion process are showcased in Figure 8, which includes a series of comparative images: the original input, the standalone outputs from DeepLab and WaSR, and the final fused output. A critical observation from these results is the seamless overlay of the aquatic region from the WaSR output onto the Panoptic DeepLab output. This overlay ensures that the aquatic regions are accurately represented in the final image.



**Figure 8.** Results of fusion model.

To quantify the effectiveness of our fusion model, we calculated the proportion of aquatic regions in both the fused output and the original WaSR output.

#### 4.1. Panoptic DeepLab Trained on Cityscapes

In the Panoptic DeepLab neural network, we integrated three essential performance metrics: panoptic quality (PQ), average precision (AP), and mean intersection over union (MIoU). The Panoptic DeepLab is initially trained using the Cityscapes dataset, which does not include aquatic images. To represent aquatic scenes more accurately, we established two separate MIoU metrics:  $MIoU_{origin}$  for standard scenarios and  $MIoU_{water}$  for aquatic contexts. The specific values are as follows.

Table 3 illustrates the performance of Panoptic DeepLab based on Cityscapes.  $MIoU_{origin}$  denotes the metric calculated from the original Cityscapes dataset, which excludes aquatic

areas. Conversely,  $MIoU_{water}$  refers to the metric derived from test data that include aquatic regions, yet it is still evaluated using the neural network trained on the original Cityscapes dataset. Consequently, it is expected that  $MIoU_{water}$  will be lower than  $MIoU_{origin}$ , reflecting the network's limited exposure to aquatic imagery during training.

**Table 3.** Panoptic DeepLab performance on Cityscapes.

Method	PQ	AP	$MIoU_{origin}$	$MIoU_{water}$
Panoptic DeepLab	63.0%	35.3%	80.5%	76.47%

#### 4.2. WaSR

The WaSR model, specifically designed for aquatic data, was trained on the MaStr1325 dataset, which includes only three classes, underscoring its targeted focus. In scenarios with notable class imbalance, a common feature of such datasets, the F1 score is often a more suitable evaluation metric [67]. This score, balancing precision and recall, offers a fairer assessment of performance, particularly for minority classes. WaSR is typically evaluated using three metrics: precision (Pr), recall (Re), and F-score (F1) shown in Table 4. To facilitate a more effective comparison with Panoptic DeepLab, we also recalculated WaSR's  $MIoU$ . The specific values for these metrics are outlined in the table below.

**Table 4.** WaSR performance on MaStr1325.

Method	Pr	Re	F1	$MIoU$
WaSR	94.60%	96.50%	95.50%	99.80%

The data in the table show that the  $MIoU$  values do not correlate with the Pr, Re, and F1 values, which are generally determined based on different evaluation objectives. In this study, we recalculated the  $MIoU$  values specifically to facilitate a comparison between the outcomes of this experiment and the results from the initial experiments.

#### 4.3. Fusion Model

We extracted the segmentation capability of WaSR, particularly for aquatic regions, and integrated it with Panoptic DeepLab. This combination produced the following  $MIoU$  results.

The  $MIoU$  for WaSR, when trained on the MaStr1325 dataset, reached 99.80%. In contrast, the original Panoptic DeepLab, trained with the Cityscapes dataset, achieved an  $MIoU$  of 80.5%. However, as the Cityscapes dataset lacks aquatic categories, the  $MIoU$  of Panoptic DeepLab decreased to 76.47% when processing images containing aquatic regions. By integrating aquatic labels derived from WaSR training into the Cityscapes dataset, we effectively extend its labeled categories. This extension results in a higher  $MIoU$  for the newly added aquatic classification compared to the original Cityscapes labels. Thus, our approach not only expands the classification range of the Cityscapes dataset but also enhances the corresponding  $MIoU$  values.

This result shown in Table 5 is particularly significant as it confirms the successful augmentation of the cityscape-based Panoptic DeepLab output with aquatic data without necessitating any retraining of the model. The accuracy of the aquatic segmentation is thus maintained, while the extensive terrestrial labeling from Panoptic DeepLab is effectively incorporated, showcasing the efficacy of our fusion approach in enhancing environmental segmentation.

**Table 5.** Performance comparison.

Method	MIoU	Dataset
WaSR	99.80%	MaSTr1325
Panoptic DeepLab	76.47%	Cityscapes
<b>Fusion Model (Ours)</b>	<b>81.46%</b>	<b>Cityscapes</b>

Table 6 illustrates the legend of our fusion model. As the model merges Panoptic DeepLab with WaSR, we simply extract the aquatic label from WaSR and integrate it into the Panoptic DeepLab results, which forms the basis of the legend.

**Table 6.** The Legend of the fusion model.

ID	Name	RGB Value	Color Name
0	unlabeled	0, 0, 0	Black
1	ego vehicle	0, 0, 0	Black
2	rectification border	0, 0, 0	Black
3	out of roi	0, 0, 0	Black
4	static	0, 0, 0	Black
5	dynamic	111, 74, 0	Dark brown
6	ground	81, 0, 81	Purple
7	road	128, 64, 128	Medium purple
8	sidewalk	244, 35, 232	Bright pink
9	parking	250, 170, 160	Light salmon
10	rail track	230, 150, 140	Light taupe
11	building	70, 70, 70	Dim gray
12	wall	102, 102, 156	Slate gray
13	fence	190, 153, 153	Dusty rose
14	guard rail	180, 165, 180	Grayish-pink
15	bridge	150, 100, 100	Sienna
16	tunnel	150, 120, 90	Olive
17	pole	153, 153, 153	Gray
18	polegroup	153, 153, 153	Gray
19	traffic light	250, 170, 30	Yellow
20	traffic sign	220, 220, 0	Electric yellow
21	vegetation	107, 142, 35	Olive green
22	terrain	152, 251, 152	Pale green
23	sky	70, 130, 180	Steel blue
24	person	220, 20, 60	Crimson
25	rider	255, 0, 0	Red
26	car	0, 0, 142	Deep blue
27	truck	0, 0, 70	Navy blue
28	bus	0, 60, 100	Dark cerulean
29	caravan	0, 0, 90	Dark blue
30	trailer	0, 0, 110	Dark midnight blue
31	train	0, 80, 100	Dark teal
32	motorcycle	0, 0, 230	Blue
33	bicycle	119, 11, 32	Maroon
34	license plate	0, 0, 142	Deep blue
35	water	41, 167, 224	Dodger blue

## 5. Conclusions

This research introduces an innovative fusion model designed to seamlessly integrate aquatic environment labels into a neural network initially trained on the Cityscapes dataset. This model eliminates the need for additional training, a significant advancement in the field of neural network application. Our strategy involved utilizing two cutting-edge neural networks, Panoptic DeepLab and WaSR, each with its unique strengths. Panoptic DeepLab excels in identifying a wide range of environmental features but falls short in

aquatic environment recognition. Conversely, WaSR demonstrates exceptional proficiency in aquatic region detection but lacks in-depth analysis of contemporary terrestrial features.

To bridge this gap, we developed a specialized finder tool for WaSR. This tool is adept at locating specific aquatic areas based on their labels and creating an accurate mask. This mask is then intricately fused with the output from Panoptic DeepLab, effectively combining the capabilities of both networks. As a result, our fusion model not only integrates aquatic labels into the Panoptic DeepLab's cityscape results but also does so without necessitating retraining of the network.

A significant advantage of our model is its compatibility with CPU processing. This compatibility streamlines the implementation process, making it more accessible and resource-efficient, particularly beneficial for applications at the edge. Running on a CPU also ensures that the model can be easily integrated into various systems without the need for specialized, high-end hardware.

We aim to refine our approach by incorporating a lighter neural network for the identification of aquatic regions. This development is expected to further simplify the implementation process and enhance deployability, particularly on edge hardware like FPGAs [68,69]. Such a lightweight network would not only retain the model's effectiveness but also increase its practicality for real-world applications, especially in edge-computing scenarios.

**Author Contributions:** Conceptualization, Z.Y., C.-Y.L. and C.-W.S.; methodology, Z.Y. and C.-Y.L.; software, Z.Y.; validation, Z.Y., R.W. and L.M.; formal analysis, Z.Y.; investigation, Z.Y.; resources, C.-W.S.; data curation, Z.Y.; writing—original draft preparation, Z.Y.; writing—review and editing, L.M. and C.-W.S.; visualization, Z.Y.; supervision, L.M. and C.-W.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author.

**Acknowledgments:** The authors would like to express sincere gratitude to William Cheung and Chung Yim Yiu from the Department of Property, The University of Auckland, for their invaluable contributions to this paper. Their insightful ideas, feedback, and dedication significantly enriched the content and quality of this work. The authors are truly thankful for their collaboration on this project.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Yuan, K.; Abe, H.; Otsuka, N.; Yasufuku, K.; Takahashi, A. Impact of the COVID-19 Pandemic on Walkability in the Main Urban Area of Xi'an. *Urban Sci.* **2022**, *6*, 44. [\[CrossRef\]](#)
2. Verma, D.; Jana, A.; Ramamritham, K. Quantifying Urban Surroundings Using Deep Learning Techniques: A New Proposal. *Urban Sci.* **2018**, *2*, 78. [\[CrossRef\]](#)
3. Chaturvedi, V.; de Vries, W.T. Machine Learning Algorithms for Urban Land Use Planning: A Review. *Urban Sci.* **2021**, *5*, 68. [\[CrossRef\]](#)
4. Leya, R.S.; Jodder, P.K.; Rahaman, K.R.; Chowdhury, M.A.; Parida, D.; Islam, M.S. Spatial Variations of Urban Heat Island Development in Khulna City, Bangladesh: Implications for Urban Planning and Development. *Earth Syst. Environ.* **2022**, *6*, 865–884. [\[CrossRef\]](#)
5. Ballouch, Z.; Hajji, R.; Poux, F.; Kharroubi, A.; Billen, R. A Prior Level Fusion Approach for the Semantic Segmentation of 3D Point Clouds Using Deep Learning. *Remote Sens.* **2022**, *14*, 3415. [\[CrossRef\]](#)
6. Feng, Y.; Diao, W.; Sun, X.; Li, J.; Chen, K.; Fu, K.; Gao, X. Npaloss: Neighboring Pixel Affinity Loss for Semantic Segmentation in High-Resolution Aerial Imagery. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *2*, 475–482. [\[CrossRef\]](#)
7. Karimi, J.D.; Corstanje, R.; Harris, J.A. Bundling ecosystem services at a high resolution in the UK: Trade-offs and synergies in urban landscapes. *Landsc. Ecol.* **2021**, *36*, 1817–1835. [\[CrossRef\]](#)
8. Cordts, M.; Omran, M.; Ramos, S.; Rehfeld, T.; Enzweiler, M.; Benenson, R.; Franke, U.; Roth, S.; Schiele, B. The Cityscapes Dataset for Semantic Urban Scene Understanding. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 26 June–1 July 2016; pp. 3213–3223.
9. Vergaño-Salazar, J.G.; Meyer, J.F.; Córdova-Lepe, F.; Marín, E.D. Distribution model of toxic agents and runoff phenomenon in flat aquatic regions. *J. Phys. Conf. Ser.* **2020**, *1514*, 012004. [\[CrossRef\]](#)
10. Pritikin, B.; Prochaska, J.X. AI based Out-Of-Distribution Analysis of Sea Surface Height Data. *arXiv* **2023**, arXiv:2306.06072.



11. Lewis, J.A.; Ernstson, H. Contesting the Coast: Socioecological Cleavages and Coastal Planning in the Mississippi River Delta. Available online: <https://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-119412> (accessed on 7 December 2023).
12. Vaishnav, M.; Cadene, R.; Alamia, A.; Linsley, D.; VanRullen, R.; Serre, T. Understanding the Computational Demands Underlying Visual Reasoning. *Neural Comput.* **2021**, *34*, 1075–1099. [\[CrossRef\]](#)
13. Noguchi, R.; Sankur, O.; Jéron, T.; Markey, N.; Mentré, D. Repairing Real-Time Requirements. In Proceedings of the Automated Technology for Verification and Analysis, Virtual Event, 25–28 October 2022.
14. Lo, C.Y.; Sham, C.W. Energy Efficient Fixed-point Inference System of Convolutional Neural Network. In Proceedings of the 2020 IEEE 63rd International Midwest Symposium on Circuits and Systems (MWSCAS), Springfield, MA, USA, 9–12 August 2020; pp. 403–406.
15. Tenorio, R.H.V.; Sham, C.W.; Vargas, D.V. Preliminary Study of Applied Binary Neural Networks for Neural Cryptography. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion. Association for Computing Machinery, Cancún, Mexico, 8–12 July 2020; pp. 291–292.
16. Valencia, R.; Sham, C.W.; Sinnen, O. Evolved Binary Neural Networks Through Harnessing FPGA Capabilities. In Proceedings of the 2019 International Conference on Field-Programmable Technology (ICFPT), Tianjin, China, 9–13 December 2019; pp. 395–398.
17. Valencia, R.; Sham, C.W.; Sinnen, O. Using Neuroevolved Binary Neural Networks to solve reinforcement learning environments. In Proceedings of the 2019 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), Bangkok, Thailand, 11–14 November 2019; pp. 301–304.
18. Lo, C.Y.; Lau, F.C.M.; Sham, C.W. Fixed-Point Implementation of Convolutional Neural Networks for Image Classification. In Proceedings of the 2018 International Conference on Advanced Technologies for Communications (ATC), Ho Chi Minh City, Vietnam, 18–20 October 2018; pp. 105–109.
19. Cheng, B.; Collins, M.D.; Zhu, Y.; Liu, T.; Huang, T.S.; Adam, H.; Chen, L.C. Panoptic-DeepLab: A Simple, Strong, and Fast Baseline for Bottom-Up Panoptic Segmentation. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 12472–12482.
20. Bovcon, B.; Kristan, M. WaSR—A Water Segmentation and Refinement Maritime Obstacle Detection Network. *IEEE Trans. Cybern.* **2021**, *52*, 12661–12674. [\[CrossRef\]](#)
21. Yue, Z.; Sham, C.W.; Lo, C.Y.; Cheung, W.; Yiu, C.Y. Sea View Extension for Semantic Segmentation in Cityscapes. In Proceedings of the 2023 9th International Conference on Applied System Innovation (ICASI), Chiba, Japan, 21–25 April 2023; pp. 33–35.
22. Wu, Q.; Yang, Q.; Zheng, X. A Multi-Task Model for Sea-Sky Scene Perception with Information Intersection. In Proceedings of the 8th International Conference on Computing and Artificial Intelligence, Tianjin, China, 18–21 March 2022.
23. Zhou, Z.; Liu, S.; Duan, J.; Aikaterini, M. A Superpixel-based Water Scene Segmentation Method by Sea-sky-line and Shoreline Detection. In Proceedings of the 2021 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), Chengdu, China, 18–20 June 2021; pp. 413–418.
24. Seyedhosseini, M.; Tasdizen, T. Semantic Image Segmentation with Contextual Hierarchical Models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 951–964. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Li, Y.; Huang, J.J.; Hu, M.; Yang, H.; Tanaka, K. Design of low impact development in the urban context considering hydrological performance and life-cycle cost. *J. Flood Risk Manag.* **2020**, *13*, e12625. [\[CrossRef\]](#)
26. Bovcon, B.; Muhovič, J.; Pers, J.; Kristan, M. The MaSTr1325 dataset for training deep USV obstacle detection models. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 3431–3438.
27. Bovcon, B.; Muhovič, J.; Vranac, D.; Mozetic, D.; Pers, J.; Kristan, M. MODS—A USV-Oriented Object Detection and Obstacle Segmentation Benchmark. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 13403–13418. [\[CrossRef\]](#)
28. Gui, Y.; Zhang, X.H.; Shang, Y.; Wang, K.P. A Real-Time Sea-Sky-Line Detection Method under Complicated Sea-Sky Background. *Appl. Mech. Mater.* **2012**, *182–183*, 1826–1831. [\[CrossRef\]](#)
29. Song, H.; Ren, H.; Song, Y.; Chang, S.; Zhao, Z.G. A Sea-Sky Line Detection Method Based on the RANSAC Algorithm in the Background of Infrared Sea-Land-Sky Images. *J. Russ. Laser Res.* **2021**, *42*, 318–327. [\[CrossRef\]](#)
30. Zhan, W.; Xiao, C.; Wen, Y.; Zhou, C.; Yuan, H.; Xiu, S.; Zou, X.; Xie, C.H.; Li, Q. Adaptive Semantic Segmentation for Unmanned Surface Vehicle Navigation. *Electronics* **2020**, *9*, 213. [\[CrossRef\]](#)
31. Yu, Y.; Bao, Y.; Wang, J.; Chu, H.; Zhao, N.; He, Y.; Liu, Y. Crop Row Segmentation and Detection in Paddy Fields Based on Treble-Classification Otsu and Double-Dimensional Clustering Method. *Remote Sens.* **2021**, *13*, 901. [\[CrossRef\]](#)
32. Ghosh, S.; Das, A. Modelling urban cooling island impact of green space and water bodies on surface urban heat island in a continuously developing urban area. *Model. Earth Syst. Environ.* **2018**, *4*, 501–515. [\[CrossRef\]](#)
33. Li, Y.; Zhao, H.; Qi, X.; Wang, L.; Li, Z.; Sun, J.; Jia, J. Fully Convolutional Networks for Panoptic Segmentation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 214–223.
34. Said, K.A.M.; Jambek, A.B. DNA Microarray Image Segmentation Using Markov Random Field Algorithm. *J. Phys. Conf. Ser.* **2021**, *2071*, 012032. [\[CrossRef\]](#)
35. Sun, B.; Kuen, J.; Lin, Z.; Mordohai, P.; Chen, S. PRN: Panoptic Refinement Network. In Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–7 January 2023; pp. 3952–3962.

36. Schachtschneider, J.; Brenner, C. Creating Multi-Temporal Maps of Urban Environments for Improved Localization of Autonomous Vehicles. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* **2020**, *XLIII-B2-2020*, 317–323. [\[CrossRef\]](#)
37. Li, Y.; Zhao, H.; Qi, X.; Chen, Y.; Qi, L.; Wang, L.; Li, Z.; Sun, J.; Jia, J. Fully Convolutional Networks for Panoptic Segmentation With Point-Based Supervision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *45*, 4552–4568. [\[CrossRef\]](#)
38. Wang, H.; Ji, X.; Peng, K.; Wang, W.; Wang, S. PVONet: Point-voxel-based semi-supervision monocular three-dimensional object detection using LiDAR camera systems. *J. Electron. Imaging* **2023**, *32*, 053015. [\[CrossRef\]](#)
39. Gasperini, S.; Mahani, M.A.N.; Marcos-Ramiro, A.; Navab, N.; Tombari, F. Panoster: End-to-End Panoptic Segmentation of LiDAR Point Clouds. *IEEE Robot. Autom. Lett.* **2020**, *6*, 3216–3223. [\[CrossRef\]](#)
40. Sodano, M.; Magistri, F.; Guadagnino, T.; Behley, J.; Stachniss, C. Robust Double-Encoder Network for RGB-D Panoptic Segmentation. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 23–27 May 2022; pp. 4953–4959.
41. Kao, P.Y.; Zhang, R.; Chen, T.; Hung, Y.P. Absolute Camera Pose Regression Using an RGB-D Dual-Stream Network and Handcrafted Base Poses. *Sensors* **2022**, *22*, 6971. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Zheng, Z.; Hu, Y.; Zhang, Y.; Yang, H.; Qiao, Y.; Qu, Z.; Huang, Y. CASPPNet: A chained atrous spatial pyramid pooling network for steel defect detection. *Meas. Sci. Technol.* **2022**, *33*, 085403. [\[CrossRef\]](#)
43. Thanasutives, P.; Fukui, K.-I.; Numao, M.; Kijirikul, B. Encoder-Decoder Based Convolutional Neural Networks with Multi-Scale-Aware Modules for Crowd Counting. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2020; pp. 2382–2389.
44. Sholomov, D.L. Application of shared backbone DNNs in ADAS perception systems. In Proceedings of the International Conference on Machine Vision, Online, 25–27 July 2021.
45. Eicken, H.; Danielsen, F.; Sam, J.M.; Fidel, M.; Johnson, N.; Poulsen, M.K.; Lee, O.; Spellman, K.V.; Iversen, L.; Pulsifer, P.L.; et al. Connecting Top-Down and Bottom-Up Approaches in Environmental Observing. *Bioscience* **2021**, *71*, 467–483. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Kumar, U.; Mishra, S.; Dash, K. An IoT and Semi-Supervised Learning-Based Sensorless Technique for Panel Level Solar Photovoltaic Array Fault Diagnosis. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [\[CrossRef\]](#)
47. Choi, S.; Jung, S.; Yun, H.; Kim, J.T.; Kim, S.; Choo, J. RobustNet: Improving Domain Generalization in Urban-Scene Segmentation via Instance Selective Whitening. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 29–25 June 2021; pp. 11575–11585.
48. Jeong, M.; Li, A.Q. Efficient LiDAR-based In-water Obstacle Detection and Segmentation by Autonomous Surface Vehicles in Aquatic Environments. In Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Prague, Czech Republic, 27 September–1 October 2021; pp. 5387–5394.
49. Barman, R.; Ehrmann, M.; Clematide, S.; Oliveira, S.A.; Kaplan, F. Combining Visual and Textual Features for Semantic Segmentation of Historical Newspapers. *arXiv* **2020**, arXiv:abs/2002.06144.
50. Clemente, F.M.; Akyildiz, Z.; Pino-Ortega, J.; Rico-González, M. Validity and Reliability of the Inertial Measurement Unit for Barbell Velocity Assessments: A Systematic Review. *Sensors* **2021**, *21*, 2511. [\[CrossRef\]](#)
51. Coviello, G.; Avitabile, G. Multiple Synchronized Inertial Measurement Unit Sensor Boards Platform for Activity Monitoring. *IEEE Sens. J.* **2020**, *20*, 8771–8777. [\[CrossRef\]](#)
52. Neagoe, I.C.; Coca, M.; Vaduva, C.; Datcu, M. Cross-Bands Information Transfer to Offset Ambiguities and Atmospheric Phenomena for Multispectral Data Visualization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 11297–11310. [\[CrossRef\]](#)
53. Hong, S.; Chung, D.; Kim, J.; Kim, Y.; Kim, A.; Yoon, H.K. In-water visual ship hull inspection using a hover-capable underwater vehicle with stereo vision. *J. Field Robot.* **2018**, *36*, 531–546. [\[CrossRef\]](#)
54. Prunier, J.G.; Poésy, C.; Dubut, V.; Veyssi re, C.; Loot, G.; Poulet, N.; Blanchet, S. Quantifying the individual impact of artificial barriers in freshwaters: A standardized and absolute genetic index of fragmentation. *Evol. Appl.* **2020**, *13*, 2566–2581. [\[CrossRef\]](#)
55. Lim, J.H.X.; Phang, S.K. Classification and Detection of Obstacles for Rover Navigation. *J. Phys. Conf. Ser.* **2023**, *2523*, 012030. [\[CrossRef\]](#)
56. Cheng, C.; Liu, D.; Du, J.H.; Zheng, L., Y. Research on Visual Perception for Coordinated Air–Sea through a Cooperative USV-UAV System. *J. Mar. Sci. Eng.* **2023**, *11*, 1978. [\[CrossRef\]](#)
57. Radzki, G.; Nielsen, I.E.; Golińska-Dawson, P.; Bocewicz, G.; Banaszak, Z.A. Reactive UAV Fleet’s Mission Planning in Highly Dynamic and Unpredictable Environments. *Sustainability* **2021**, *13*, 5228. [\[CrossRef\]](#)
58. Meng, J.; Humne, A.; Bucknall, R.W.G.; Englot, B.; Liu, Y. A Fully-Autonomous Framework of Unmanned Surface Vehicles in Maritime Environments Using Gaussian Process Motion Planning. *IEEE J. Ocean. Eng.* **2022**, *48*, 59–79. [\[CrossRef\]](#)
59. Schauer, S.; Kalogeraki, E.M.; Papastergiou, S.; Douligeris, C. Detecting Sophisticated Attacks in Maritime Environments using Hybrid Situational Awareness. In Proceedings of the 2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM), Paris, France, 18–20 December 2019; pp. 1–7.
60. Caro, M.C.; Huang, H.Y.; Cerezo, M.; Sharma, K.; Sornborger, A.T.; Cincio, L.; Coles, P.J. Generalization in quantum machine learning from few training data. *Nat. Commun.* **2021**, *13*, 4919. [\[CrossRef\]](#)
61. Elmes, A.; Alemohammad, S.H.; Avery, R.; Caylor, K.K.; Eastman, J.R.; Fishgold, L.; Friedl, M.A.; Jain, M.; Kohli, D.; Bayas, J.C.L.; et al. Accounting for Training Data Error in Machine Learning Applied to Earth Observations. *Remote. Sens.* **2019**, *12*, 1034. [\[CrossRef\]](#)

62. Zheng, X.; Fu, C.; Xie, H.; Chen, J.; Wang, X.; Sham, C.W. Uncertainty-aware deep co-training for semi-supervised medical image segmentation. *Comput. Biol. Med.* **2022**, *149*, 106051. [[CrossRef](#)] [[PubMed](#)]
63. Chen, J.; Fu, C.; Xie, H.; Zheng, X.; Geng, R.; Sham, C.W. Uncertainty teacher with dense focal loss for semi-supervised medical image segmentation. *Comput. Biol. Med.* **2022**, *149*, 106034. [[CrossRef](#)] [[PubMed](#)]
64. Sharma, M. *Identification of Night Time Poor Visibility Areas in Urban Streets*; CSIR-NIScPR: New Delhi, India, 2020.
65. Gudwani, H.; Singh, V.J.; Mahajan, S.; Mittal, D.; Das, A. Identification of poor visibility conditions in urban settings. In Proceedings of the 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, 3–5 July 2017; pp. 1–5.
66. Malygin, I.G.; Tarantsev, A.A. On ensuring the safe movement of emergency service vehicles under hazardous driving conditions. *Pozharovzryvobezopasnost/Fire Explos. Saf.* **2022**, *30*, 97–107. [[CrossRef](#)]
67. Chicco, D.; Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genom.* **2020**, *21*, 6. [[CrossRef](#)]
68. Lo, C.Y.; Ma, L.; Sham, C.W. CNN Accelerator with Non-Blocking Network Design. In Proceedings of the 2021 IEEE 10th Global Conference on Consumer Electronics (GCCE), Kyoto, Japan, 12–15 October 2021; pp. 813–815.
69. Lo, C.Y.; Sham, C.W.; Fu, C. Novel CNN Accelerator Design With Dual Benes Network Architecture. *IEEE Access* **2023**, *11*, 59524–59529. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.