Semantic segmentation and CNN – A review of challenges , solutions, and future perspectives

Mohamed Moin Irfan, *Dept of CSE RV College of Engineering, Bangalore, India Mohamedmoini.cs19@rvce.edu.in*

Prerana Shekar M S, *Dept of CSE RV College of Engineering, Bangalore, India Preranashekarms.cs18@rvce.edu.in*

Purnodeep Ranjankar, *Dept of CSE RV College of Engineering, Bangalore, India* purnodeepr.cs19@rvce.edu.in

Sathvik Gowda M, Dept of CSE RV College of Engineering, Bangalore, India sathvikgowdam.cs18@rvce.edu.in

Prof Manonmani S, Assistant Professor Dept of CSE RV College of Engineering,India manonmanis@rvce.edu.in

Abstract- Worldwide, sea mines pose a serious threat to ships at sea, and MCM(mine countermeasure) equipment is used to safeguard them. Minesweeping is one tactic employed by anti-missile forces, which comprises a suspicious search for any mines in the area. The process has four steps: search, categorization, analysis, and destruction. Sonar is generally used for detection and classification, while military personnel review images of the ocean floor to find targets. To reduce work and after-hours time, automatic target recognition (ATR), computer-aided design (CAD), and computer-aided design (CAC) approaches have been developed. This article examines the various image processing, machine learning, and deep learning methods that are used in these systems for landscape categorization and recognition. Side-scan sonar images were used in early research, however they can have low resolution in difficult lighting circumstances. The objective of this project is to create a system that uses technology to provide the military with trustworthy information as quickly as is practical.

Keywords: mine countermeasure, computer-aided design, automated target recognition, minesweeping

1. INTRODUCTION

Naval mines Naval mines are a serious danger to maritime operations and the security of ships, making it essential for naval combat to identify and classify them. Traditional mine detection techniques frequently rely on sonar devices, but obtaining high-quality data for neural network training is difficult due to the scarcity of publicly accessible datasets and the secrecy concerns involved with military operations.

Deep neural networks-generated synthetic underwater photos present a viable approach to get over these restrictions and enhance mine detecting skills. Computer-generated depictions of underwater sceneries known as synthetic underwater pictures imitate the intricate visual qualities of underwater habitats. For training and testing deep learning models for a variety of underwater vision tasks, such as mine detection and segmentation, these synthetic pictures are an invaluable resource

Synthetic underwater images are computer-generated representations of underwater scenes that simulate the complex visual characteristics of underwater environments. For training and testing deep learning models for a variety of underwater vision tasks, such as mine detection and segmentation, these synthetic pictures are an invaluable resource.

Image segmentation is a computer vision technique that involves dividing an image into multiple regions or segments, each representing a distinct object or region of interest. A sort of image segmentation called semantic segmentation seeks to give each pixel in a picture a semantic label in order to enable a thorough comprehension of the scene. By reaching astounding accuracy and resilience, deep learning algorithms like the U-Net and PSPNet topologies have revolutionised picture segmentation.

In order to conduct pixel-level segmentation, the U-Net architecture, a common deep learning model for image segmentation, mixes contracting and expanding channels. It has been customised for underwater scene analysis and has demonstrated tremendous effectiveness in a variety of medical imaging applications. The U-Net model may be taught to segment underwater scenes and efficiently identify naval mines by training on synthetic underwater photos..

Similarly, the PSPNet (Pyramid Scene Parsing Network) architecture is another deep learning model designed for semantic segmentation. It utilizes pyramid pooling modules to capture multi-scale contextual information, enabling accurate and detailed segmentation of complex scenes. By applying PSPNet to synthetic underwater images, it becomes possible to identify and classify objects, including naval mines, in underwater environments.

For training, deep learning algorithms like U-Net and PSPNet need a lot of excellent labelled data. The collecting of such data is hampered by the dearth of publicly accessible underwater datasets and the confidentiality surrounding military mine detecting missions. This gap is filled by synthetic underwater picture synthesis utilising deep neural networks, which enables the production of realistic and varied underwater datasets for the training and evaluation of deep learning models for precise mine identification and segmentation.

Deep neural network-based synthetic underwater picture synthesis is essential for improving naval mine detection capabilities. Effective underwater scene segmentation, mine detection, and situational awareness may all be achieved by training deep learning models like U-Net and PSPNet on synthetic underwater photos. The progress of underwater computer vision and the creation of more reliable and effective mine detecting systems are both benefits of this combination of synthetic data generation, picture segmentation algorithms, and deep learning approaches.

2. REVIEW PROCESS

The current research took into account a thorough categorization and analysis of the literature.The methods listed below was used :

1) The database was updated by selecting the latest literature. The collected literature was reviewed until 2021.
2) Both hard copies in reputed local libraries and soft copies from the Internet were accessed for the literature review.
3) Popular search engines, such as www.google.com, www.altavista.com, etc., were employed to gather the subject-related literature from a multitude of sources. Although it has been attempted to include as many relevant works as possible, this list is by no means either complete or exhaustive.
4) The classification scheme was developed by looking at the nature of the studies. Later, the studies were examined to detect commonalities, content, advantages, and disadvantages.
5) Finally, the studies were examined with a view to suggest future avenues for research.

3. PRIOR KNOWLEDGE TO THE TOPIC

A branch of computer science and artificial intelligence (AI) called machine learning focuses on utilising data and algorithms to simulate human learning processes and gradually improve accuracy. Artificial neural networks are the core of the "deep learning" subfield of machine learning. It has the capacity to spot complex relationships and patterns in data. Deep learning does not require any explicit programming. It has lately become more well-liked as a result of advancements in processing power and the accessibility of enormous datasets. as it is based on artificial neural networks (ANNs), also known as deep neural networks (DNNs). These neural networks were developed to learn from enormous amounts of data and are motivated by the structure and function of real neurons. The main characteristic of deep learning is the use of deep neural networks, which include several layers of connected nodes. These networks may create intricate representations of the data by locating hierarchical patterns and attributes in the data. Deep learning systems may automatically learn from data and improve without active feature building. In a variety of fields, including speech recognition, image identification, natural language processing, and recommendation systems, deep learning has made significant strides. Some of the well-known Deep Learning architectures include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and deep belief networks (DBNs).

Machine learning's strong deep learning strategy incorporates a number of data processing methods. It may be used for problems involving reinforcement learning, unsupervised learning, and both. In supervised learning, neural networks are taught to categorise data or make predictions based on labelled datasets while reducing prediction errors using methods like backpropagation. Unsupervised learning is the process of finding patterns or grouping datasets without the use of labels, enabling the computer to identify unobservable links in the data. Deep learning is used to learn rules that maximise cumulative rewards, whereas reinforcement learning focuses on decision-making in a setting to maximise rewards. neural networks with convolutions, Among the deep learning methods utilised for these tasks are generative models, autoencoders, and recurrent neural networks. In general, deep learning provides a flexible and efficient framework for addressing a range of machine learning issues, such as image

recognition, clustering, language translation, robotics, and game playing.The most often used architectures in deep learning are feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs).A subset of deep learning algorithms called convolutional neural networks (CNNs) are particularly good at processing and identifying pictures. This structure is composed of several layers, including convolutional layers, pooling layers, and totally connected layers.

The most important component of a CNN is its convolutional layers, where filters are used to extract details from the input image such edges, textures, and shapes. Pooling layers are then used to down-sample the feature maps, save the most important information while reducing the spatial dimensions, and send the output of the convolutional layers. The output of the pooling layers is then applied to one or more fully connected layers to predict or categorise the image.

The structure of the visual cortex of the human brain, which comprises specialised cells that react to particular regions of the visual field, served as an inspiration for CNNs. Similar to this, CNNs include layers of linked neurons that have been trained to recognise and extract features using convolutional operations from input data.

The training process of a CNN involves feeding it with labeled training data and optimizing the network's parameters through a process called backpropagation. During training, the CNN learns to recognize and classify patterns in the input data, adjusting its internal parameters to minimize the prediction error.
CNNs have achieved remarkable success in various domains, including computer vision, medical imaging, speech recognition, and natural language processing. Their ability to automatically learn features from data, coupled with their hierarchical structure, makes CNNs a powerful tool for analyzing and understanding complex visual information.

The convolutional layer, which applies filters or kernels to the input data to achieve local feature extraction, is one of the crucial elements of CNNs. The incoming data is scanned by these filters, which identify patterns and details like edges, textures, and forms. Convolutional neural networks (CNNs) are able to progressively acquire more complicated and abstract representations of the input data through a process of convolution, non-linear activation, and pooling.Additional layer types found in CNNs include pooling layers and. By downsampling the feature maps using pooling layers, the spatial dimensions are reduced but the key characteristics are preserved. High-level thinking and decision-making are made possible by fully linked layers, which link all neurons from the preceding layer to the following layer.The strength of CNNs resides in their capacity to learn hierarchical data representations, starting with basic characteristics and progressing to more intricate and significant representations. In tasks like image classification, object recognition, picture segmentation, and even natural language processing, CNNs are incredibly successful as a result.

A CNN is trained by feeding it labelled training data and then optimising the network's parameters using a technique known as backpropagation. The CNN adjusts its internal settings to minimise the prediction error as it trains to identify and categorise patterns in the input data.
CNNs have excelled in a number of fields, including speech recognition, computer vision, medical imaging, and natural language processing. CNNs are an effective tool for analysing and comprehending complicated visual data because of its hierarchical structure and capacity to automatically learn characteristics from input.

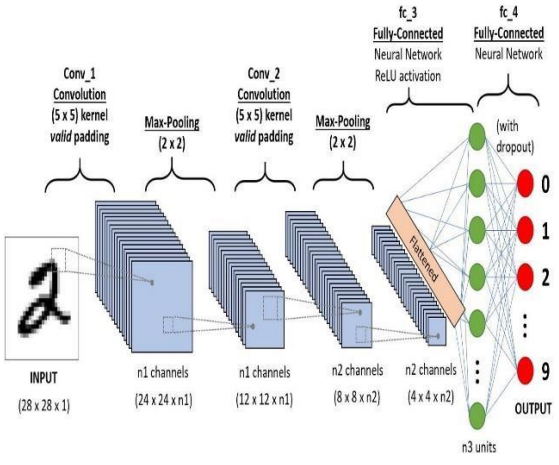


Fig 3.1: CNN architecture
(*Source:Towardsdatascience.com)*

3.1 Convolutional Layer

It tries to learn the feature representation of the images, whether the inputs are pictures of cats vs dogs or numbers. The different feature maps are computed using a variety of kernels and matrices. As a result, the (n*n) matrix's filter or kernel, which is then applied to the input data (or image) to create the convolutional feature, is determined by the sort of problem we are trying to solve. This convolution feature is applied to the next layer after biassing and applying any required activation functions..

3.2 Pooling Layer

The pooling layer comes in between the convolutional layers. The resolution of the feature maps is decreased in order to generate shift invariance. The two most popular pooling procedures are average pooling and maximal pooling. In essence, fewer connections between convolutional layers mean that the processing units are under less computational burden.
The following is a list of a few distinct pooling methods:
1.Lp Pooling
2.Max Pooling
3.Average Pooling
4.Mixed Pooling and so on

3.3 Fully-Connected Layer

There may be a number of fully connected layers after several convolutional and pooling layers. All of the neurons in the layer above the current layer are connected to one another.The output layer of the last CNN layer makes the final predictions. For classification tasks, the Softmax function is typically used when several classes are sought for prediction (example: MNIST Dataset) and the Sigmoid function is used for binary classification (example: Cats vs. Dogs).

4. Introduction to Semantic Segmentation and its Connection to Deep Learning

Assigning a semantic label to each pixel in an image as part of the computer vision job known as semantic segmentation separates the picture into meaningful parts. By assigning each pixel a relevant class or category, it seeks to give a thorough comprehension of the scene. This method is essential for a number of applications, such as augmented reality, autonomous driving, and scene interpretation.

Convolutional neural networks (CNNs), in particular, have completely changed the area of semantic segmentation. CNNs have excelled in deciphering complex patterns and identifying spatial connections in pictures. They possess the capacity to automatically learn hierarchical representations of visual input, making segmentation more precise and effective.

The construction and training of CNNs are where the relationship between deep learning and semantic segmentation is found. CNNs are made to use convolutional processes to extract significant characteristics from the input picture. The network can learn more complicated representations thanks to several layers of convolution and pooling. Backpropagation is used to optimise the network's parameters while modifying the weights to reduce prediction error.
There are different types of semantic segmentation approaches, each with its own characteristics and applications:
1. Semantic Segmentation: This method treats all instances of the same class as a single object and applies a single class label to each pixel. It aims to capture the scene's overall meaning.
2. Instance Segmentation: The objective of this approach is to identify and categorise each distinct instance or item that is present in the image. By differentiating across instances of the same class, it offers a more thorough knowledge.
3. Panoptic Segmentation: Panoptic segmentation combines semantic and instance segmentation with the goal of providing a coherent representation of both the scene's objects (such as the road and sky) and its objects (such as the vehicles and pedestrians). Each pixel is given a semantic label, and it makes distinctions between various occurrences.
4. Real-time Segmentation: This technique is best suited for real-time applications like robotics and autonomous systems since it concentrates on attaining quick and effective inference.

Deep learning methods in conjunction with semantic segmentation have made substantial advancements in computer vision. Numerous applications, including autonomous driving, object identification, image editing, and augmented reality, have become possible because to the capacity to precisely and quickly identify every pixel in an image. Semantic segmentation models are becoming more precise, quick, and robust as a result of continuing research and development in this field, making them essential tools for visual comprehension and analysis

Three popular model architectures for semantic segmentation are U-Net, FCN (Fully Convolutional Network), and DeepLab.

1.U-Net: Although it was created primarily for biomedical image segmentation, U-Net is a widely used semantic segmentation architecture that has applications in many other fields. It consists of a symmetric expanding path that provides accurate localization and a contracting path that captures context. Convolutional and pooling layers are used in the contracting route to remove features and shrink the spatial dimensions. The segmentation masks are produced by the expanding route by gradually recovering the spatial resolution through upsampling and concatenation operations. Precision localization is made possible through U-Net's skip connections, which allow data from earlier levels to be transferred straight to subsequent layers. U-Net is appropriate for applications requiring precise segmentation because of its ability to manage both fine-grained information and high-level context thanks to these skip connections.

2.FCN: Another notable design for semantic segmentation is the FCN (Fully Convolutional Network). Convolutional layers are used in place of completely linked layers to maintain spatial information. To upsample the feature maps and restore the spatial resolution, FCN uses transposed convolutions, sometimes referred to as deconvolutions. Additionally, skip links are added to better combine data from several levels and improve segmentation efficiency. FCN comes in a variety of forms, including FCN-32s, FCN-16s, and FCN-8s, which gradually improve segmentation by adding data from lower levels. Application areas where FCN has been extensively employed include scene parsing, object identification, and picture segmentation.

3. DeepLab: DeepLab is a state-of-the-art model architecture for semantic segmentation, known for its excellent performance and accuracy. It utilizes atrous (dilated) convolutions to capture multi-scale contextual information without significantly increasing the computational cost. DeepLab employs an encoder-decoder structure, where the encoder captures global context using atrous convolutions, and the decoder recovers the spatial details using upsampling. One notable variant of DeepLab is DeepLabv3+, which includes a spatial pyramid pooling module to gather multi-scale contextual information and a refinement network to improve the segmentation results. DeepLab has achieved outstanding performance in various benchmark datasets and has been widely adopted in many applications, including autonomous driving, medical image analysis, and remote sensing.

Each of these three model designs, each with its own innovations and strengths, has made a substantial contribution to the semantic segmentation area. Precision localization is made possible by U-Net's skip connections, spatial information is preserved by FCN's fully convolutional architecture, and multiscale contextual data is captured by DeepLab's arous convolutions. These architectures and their variations have undergone continual research and investigation, which has significantly improved the accuracy and efficiency of computer vision applications' semantic segmentation tasks.

It is crucial to note that when comparing the three widely used model architectures for semantic segmentation, U-Net, FCN (Fully Convolutional Network), and DeepLab, the optimum design relies on the particular job, dataset, and constraints. However, we may talk about their benefits, drawbacks, and applicability for creating synthetic underwater picture datasets using deep neural networks

4.1 : U-Net:

Advantages:

U-Net has been widely utilised in biomedical image segmentation and has demonstrated high performance in a variety of applications. It is renowned for its capacity to handle fine-grained information and exact localisation due to its skip connections.

• Context can be captured and precise segmentation masks can be produced thanks to U-Net's expanding and contracting routes.

Disadvantages:

Due to its symmetric design, U-Net could be limited in its ability to handle large-scale context, and a lot of labelled data may be necessary for both good training and performance.

• Possibility of Generating Synthetic Underwater Image Data Sets:

• Because U-Net enables accurate localization and handling of tiny features, which are crucial in

underwater image processing tasks, it might be a good option for creating synthetic underwater picture datasets.

4.2 : FCN (Fully Convolutional Network):

Advantages:

- With the use of fully convolutional layers, FCN is renowned for its capacity to maintain spatial information. It can handle a range of picture sizes and generate dense pixel-wise predictions.
- • It has been demonstrated that FCN with skip connections enhances segmentation performance by combining data from several levels..

Disadvantages:

In comparison to U-Net, FCN could have trouble performing fine localisation and processing small-scale information.

• The upsampling processes may need the use of extra computational resources.

• Possibility of Generating Synthetic Underwater Image Data Sets:

• Due to its ability to efficiently maintain spatial information, which is crucial for underwater sceneries that may contain fine details, FCN may be used to create synthetic underwater picture datasets.

4.3. DeepLab:

Advantages:

• DeepLab makes use of atrous convolutions to collect multi-scale contextual data without dramatically raising the cost of computing.

• In tasks requiring semantic segmentation, it has attained cutting-edge performance.

• The DeepLabv3+ variant's extra modules enhance segmentation outcomes.

Disadvantages:

When compared to more straightforward designs, DeepLab may have larger computing requirements. To perform at its best, DeepLab may also need more training data and a longer training period.

• Possibility of Generating Synthetic Underwater Image Data Sets:

• DeepLab may be used to create synthetic underwater picture datasets since it is good at gathering contextual information and performs exceptionally well on a variety of segmentation tasks.

Given their different advantages in handling minute details and collecting contextual information, U-Net and DeepLab can both be good options for dataset development. It is advised to test out several designs and gauge how well they work on the particular underwater picture collection in order to select the best one.

5. IMPLEMENTATION OF SYNTHETIC UNDERWATER IMAGE DATA SET GENERATION USING U-NET :

A convolutional neural network called U-Net was created at the University of Freiburg's Department of Computer Science for the purpose of segmenting biological images.

UNet, which developed from the conventional convolutional neural network, was created and used for the first time in 2015 to process pictures used in biomedicine. A standard convolutional neural network focuses on classifying images, with an input of an image and an output of a single label. However, in biomedical applications, it is necessary to identify both the presence of a disease and the location of the abnormality. UNet is devoted to finding a solution to this issue. It can localise and identify boundaries since every pixel is classified, ensuring that the input and output are of same size.

For example, for an input image of size 2x2:

[[255, 230], [128, 12]] # each number is a pixel the output will have the same size of 2x2:

[[1, 0], [1, 1]] # could be any number between [0, 1]

Now let’s get to the detail implementation of UNet.

1. Show the overview of UNet
2. Breakdown the implementation line by line and further explain it

Overview

The network has basic foundation looks like:

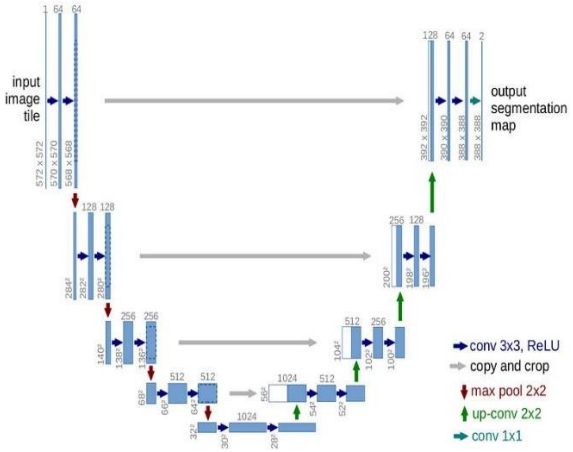


Fig 4.1: UNet architecture

First sight, it has a “U” shape. The architecture is symmetric and consists of two major parts — the left part is called contracting path, which is constituted by the general convolutional process; the right part is expansive path, which is constituted by transposed 2d convolutional layers(you can think it as an upsampling technic for now).

For semantic segmentation problems, the U-Net architecture is a well-known deep learning model. It is built on an encoder-decoder structure, where the encoder absorbs contextual information and the decoder produces a high-resolution segmentation map.

The convolutional and pooling layers that make up the encoder portion of the U-Net gradually decrease the spatial dimensions while increasing the number of feature channels.

Up-sampling and concatenation techniques are used in the U-Net's decoder section to restore the spatial resolution that was lost during encoding.

Since the encoder and decoder layers are connected via skip connections in the U-Net design, the model may access both low-level and high-level information.

Due to the combination of coarse contextual information and fine-grained features, these skip connections enable accurate localization.

The vanishing gradient problem is further lessened and faster convergence during training is made possible by the skip connections in U-Net.

U-Net combines pooling layers with tiny strides with convolutional layers with wide receptive fields to improve the model's capacity to capture spatial information.

The overlap between the predicted segmentation map and the ground truth labels is measured by the dice coefficient, a modified version of the loss function used by U-Net.

The resilience and generalisation abilities of the model are frequently improved by the application of data augmentation techniques including flipping, rotation, and elastic deformations.

Due to its capacity to manage sparse training data and generate precise and thorough segmentation results, U-Net has exhibited exceptional performance in a variety of medical picture segmentation tasks, including organ and tumour segmentation

Advantages of U-Net:

1. Effective Feature Learning: The deep convolutional neural network architecture used by U-Net enables strong feature learning. The network contains a contracting path that uses convolutional layers to collect high-level abstract characteristics and an expanding path that enables precise localisation using skip connections and upsampling. In applications like biological picture segmentation, where precise delineation of structures is essential, this architecture aids U-Net in doing very well.

2. Effective Training with Limited Labelled Data: U-Net has the advantage of being able to train efficiently with little labelled data. The training set is augmented artificially by the architecture using methods like flipping, rotating, and scaling. Additionally, the model can learn from both low-level and high-level characteristics because to the usage of skip connections, which improves its performance with less annotated data.

3. Versatility and Adaptability: U-Net is a versatile architecture that can be applied to various image segmentation tasks. It has been successfully employed in diverse domains, including biomedical imaging, satellite imaging, and natural scene understanding. Moreover, U-Net can be adapted and extended to address specific challenges or incorporate additional modules, such as attention mechanisms or dilated convolutions, to improve performance on specific tasks.

Disadvantages of U-Net:

1. High Memory Consumption: One of U-Net's drawbacks, especially in deeper systems, is its comparatively high memory consumption. Upsampling procedures on the expanding route need the storage of intermediate feature maps, which raises memory requirements for training and inference. When processing large-scale photos or working with limited computer resources, this might provide difficulties.

2. Sensitivity to Class Imbalance: Class imbalance problems in the training data may be detected by U-Net. The network may prioritise the majority classes and find it difficult to effectively segment the minority classes in situations where some classes or locations of interest are underrepresented in comparison to others. To overcome this difficulty, class imbalance must be addressed using methods like data augmentation, weighted loss functions, or specialised sampling procedures.

3. Limited Contextual Knowledge: U-Net may have limited contextual knowledge outside of the network's receptive field, despite the fact that it can record specific local information through its contracting and expanded routes. The region of the input picture that affects a pixel's prediction is referred to as the receptive field. The restricted receptive field of U-Net may cause subpar performance in activities demanding extensive contextual information or comprehensive scene knowledge. Larger receptive fields, multi-scale methods, or the incorporation of outside contextual data may all be necessary to overcome this constraint.

It's crucial to remember that these benefits and drawbacks are unique to the U-Net design and may change based on the implementation, dataset, and job requirements.

Particularly in the area of medical image analysis, the U-Net architecture has evolved into a very productive and widely accepted approach for semantic segmentation. Its distinct encoder-decoder structure and skip connections make it possible to localise objects or areas of interest within pictures with accuracy and precision. Skip connections are used to overcome the difficulties in collecting both low-level and high-level characteristics and to solve the vanishing gradient issue. The segmentation skills of U-Net are further improved by its capacity to regain spatial resolution through up-sampling and concatenation operations. The performance of the model during training is improved by using a modified loss function, such as the dice coefficient. The U-Net model is more resilient and generalised thanks to data augmentation approaches. The U-Net design has generally shown to be a powerful tool in various medical imaging tasks, providing accurate and detailed segmentation results even with limited training data.

An inventive method for overcoming the difficulties in obtaining labelled underwater images for various computer vision applications is the production of synthetic underwater picture datasets using U-Net. Because of the restricted visibility, colour distortion, and light absorption seen in underwater situations, computer vision algorithms perform far worse in these settings. Large-scale labelled underwater picture dataset collection, however, is frequently costly, time-consuming, and logistically complex. A possible alternative is provided by synthetic data creation methods, which produce fake underwater photos that closely resemble genuine underwater sceneries.Due to its capacity to recognise complex characteristics and precisely define object borders, the convolutional neural network architecture known as U-Net has become more prominent in the field of picture segmentation. It comprises of a decoder and an encoder network that collects high-level characteristics. Researchers and practitioners may get beyond the constraints of real data collecting by creating synthetic underwater picture datasets using the capability of U-Net.

Researchers may create synthetic underwater photographs from non-underwater settings by utilising U-Net to train a model on already-existing underwater images. The model can understand the fundamental patterns and properties of underwater photography thanks to this method, which makes use of the transfer learning capabilities of deep neural networks. A more complete and varied dataset is possible because to the ability of U-Net's synthetic pictures to accurately reproduce the variety of underwater environments, including various types of water, light intensities, and visibility levels.

There are several uses for the created synthetic underwater picture datasets in underwater computer vision research. They may be used to develop and test models for a variety of jobs, such as object identification, semantic segmentation, picture augmentation, and categorization of underwater scenes. Additionally, these datasets give researchers the ability to run extended tests and compare various methods in a controlled setting without the restrictions and expenses related to gathering actual undersea data.

Synthetic data combined with U-Net gives a number of benefits. First off, it offers a cheap and effective way to produce big labelled datasets, especially in situations where getting real data is difficult or impossible. Second, using synthetic data enables controlled experimentation that may be used to explore the effects of different underwater circumstances on the effectiveness of computer vision systems. Additionally, by offering a wider variety of training examples, the created synthetic pictures can aid in strengthening the resilience and generalisation of models.

In conclusion, the development of synthetic underwater picture datasets using U-Net offers a potent solution to the dearth and constraints of genuine underwater data. Researchers and professionals may create a variety of realistic synthetic underwater photos by utilising the capabilities of U-Net. Leveraging the capabilities of U-Net, researchers and practitioners can generate diverse and realistic synthetic underwater images, enabling advancements in underwater computer vision tasks. By facilitating the training and evaluation of models, synthetic datasets pave the way for improved algorithms and solutions to tackle the unique challenges of underwater environments.

6. IMPLEMENTATION OF SYNTHETIC UNDERWATER IMAGE DATA SET GENERATION USING PSP-NET

We discuss the increasing use of digital imaging in marine research driven by advances in camera technology. It offers a variety of modern sensors and algorithms for underwater photography, including general applications to marine science and specific projects such as coastal marine biodiversity, monitoring of the same Human impact on the marine environment, automatic identification of fish species, fish linkage analysis and exploration. . The purpose of this article is to explore the use of deep learning for fish segmentation and contouring in real underwater scenes. Semantic segmentation is done not only to recognize objects and their positions, but also to place text in a single pixel, extract object contours, and provide accurate area estimation. Correct mineral segmentation is important to identify morphological features such as overall length and weight, and to identify specific fish by finding the area of the profile.

The proposed algorithm is an important component of an underwater sensor platform designed for non-invasive mine assessment. This article addresses the growing need for using deep learning techniques on limited hardware, especially for remote control and control of underwater vehicles. To this end, this study explores different models of popular segmentation models, particularly the pruned variants of DeepLabv3 and PSPNet, and evaluates their performance and inference times for fish segmentation. The underwater video used in the study was recorded with a low-profile camera designed for underwater use, eliminating the need for illumination during data capture. The main contributions of this work are the alternative configurations of PSPNet and DeepLabv3 which are better for hardware and comparison of segmentation models

for extraction time and performance of mine

segmentation. ×

This document is divided into several sections that provide an overview of related publications, explaining the dataset, algorithms used and performance indicators. It also presents the experimental measurements performed and summarizes the findings. This article highlights the importance of digital images in marine science, particularly for fish segmentation

, and its implications for fish population analysis and species identification. It highlights the need for deep segmentation models to suit limited applications and provides alternative configurations to popular models when comparing their performance.

6.1 :PRINCIPLE AND WORKING OF PSPNET

PSPNet (Pyramid Scene Segmentation Network) is a deep learning model designed for semantic segmentation tasks, including image segmentation in underwater scenes. It aims to accurately assign a name to each pixel in an image, providing a detailed understanding of the structure of the scene. The principle of PSPNet can be given as follows:

Pyramid Pooling Module: The principle of PSPNet is in the Pyramid Pooling Module. This module captures various data points using layers of different sizes. It uses the idea that different objects in an image should have different dimensions to preserve information content.

Pooling layers collect and encode information at different scales, providing richer image representation. Multi-joint processing helps the network capture global context and local context, thus improving semantic partitioning results

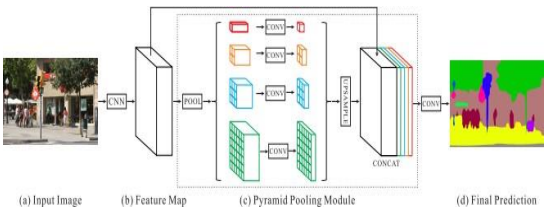


Fig 6.1: Overview of our proposed PSPNet

The pyramid pooling module provides the properties of four different pyramids. The roughest significant level in red is the earth pool that provides urine output. The lower levels of the pyramid divide certain maps into different regions and create common representations of different areas. The results of different levels in the pyramid pooling module have a special report of various variables. To preserve the weight of the global features, we use a 1 × 1 convolutional layer after each pyramid level to reduce the dimensionality of the elements represented by 1/N of the original representative if the pyramid level dimension is N.

Next, we directly upsample the low-dimensional feature maps to get features the same size as the original feature maps with bilinear interpolation. Finally, features at different levels are combined into spherical features with a final pyramid pool. Note that

Pyramid number and size of each level can be adjusted. They are related to the size of the feature map that enters the pyramid pooling layer. This

model abstracts different regions using nuclei of different sizes in a few steps.
Therefore, multilevel kernels should have a reasonable difference in notation. Our pyramid pooling module is a four-level module with dimensions of 1 x 1, 2 x 2, 3 x 3 and 6 x 6. We run general tests to show the difference for Max and Medium sharing mode.

Feature Fusion: Feature aggregation: In addition to the Pyramid Pooling Module, PSPNet also includes a join process to efficiently combine multiple features. After pooling, the features are upsampled and combined. This fusion step allows the network to use local and global information. By combining features of different scales, PSPNet improves the network's ability to accurately classify objects of various sizes and shapes. The consolidated feature captures high-quality content and high-level contextual information, improving overall segmentation performance.

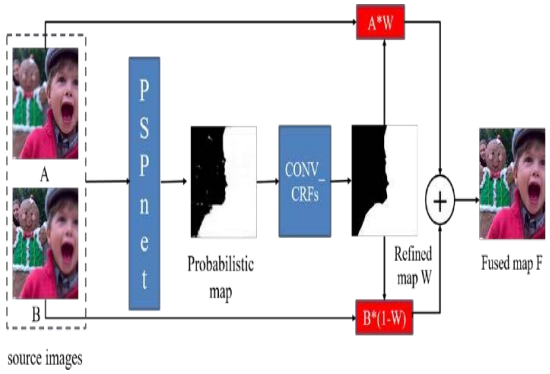


Fig 6.2: Framework of the proposed method for multi-focus image fusion

As shown in Figure 6.2, the plan can be summarized as follows: a multifocal image segmentation method using PSPnet for image compositing is proposed. Unlike the original PSPnet, the auxiliary loss and BCE loss are removed from the network and used. Change in softmax loss to show the final classifier of the pre-trained CNN model Resnet101. It can also be configured as a 6-channel input layer for the first layer of the network, allowing several images to enter the network at the same time; then ConvCRFs are adapted for further processing. Therefore, we consider two multifocal

images of the same location and the fusion process is written as

i. A pair of images A and B is entered into the PSPnet to generate the resulting map of image A, the details of the process are described in Chapter 2.

1 A pair of source images A and B are input to the PSPnet, which is utilized to extract the probabilistic map of image A,and the process details can be seen in section

2. Set ConvCRFs as optimization method to obtain optimization map W of probability map A, see fig. Section 2.2;

3. The final fusion plot F is obtained by the following process:

F A * W B * (1 - W)

Encoder-Decoder Architecture: PSPNet follows an encoder-decoder architecture, which is a common design pattern in semantic segmentation models. The encoder part, often based on a pre-trained convolutional neural network (CNN), extracts hierarchical features from the input image. In PSPNet, the encoder is responsible for capturing low-level visual features. The decoder part, including the Pyramid Pooling Module and feature fusion, takes the extracted features and performs upsampling and fusion operations to generate the final segmentation map. The decoder gradually recovers the spatial resolution and refines the segmented regions. By combining the multi-scale features and leveraging the global and local contextual information, PSPNet achieves accurate semantic segmentation in real-world underwater scenes.

Most semantic segmentation models contains two parts, i.e an Encoder and a Decoder. The Encoder is responsible for the extracting out features from the image, the decoder is the one which predicts the class of the pixel at the end. A typical Encoder-Decoder for segmentation task looks like the architecture shown below:

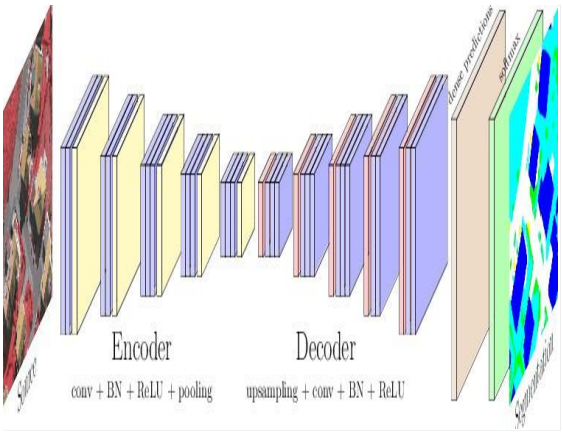


Fig 6.3: Encoder-Decoder Networks for Semantic Segmentation

PSPNet Encoder

The PSPNet encoder contains the CNN backbone with dilated convolutions along with the pyramid pooling module.

Dilated Convolutions

In the last layers of the backbone, we replace the traditional convolutional layers with Dilated convolution layers, which helps in increasing the receptive field. This Dilated convolution layers are placed in the last two blocks of the backbone. Hence the feature recieved at the end of the backbone contains richer features. The illustration[2] shows what dilated convolutions do and how is it different from convolutions.

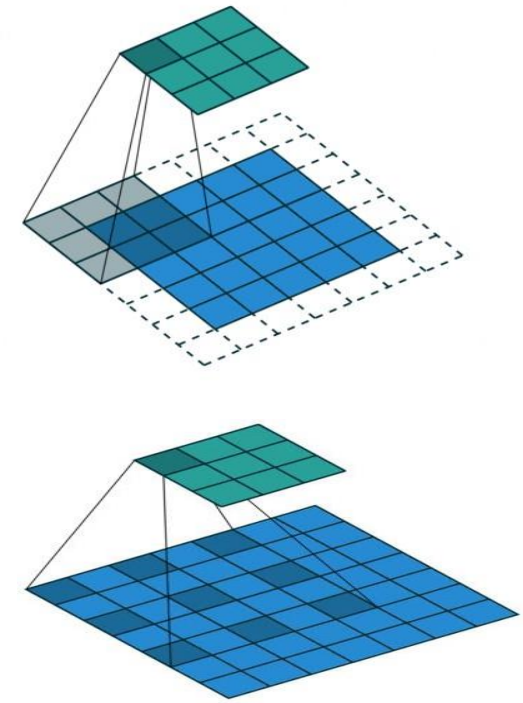


Fig 6.4: Animation of convolution with dilation=2 (left) and dilation=1(right).

In the above fig 4.4, When dilation=1, it is just the standard convolution operation.The value of dilation specifies the sparsity while doing the convolution. We can see that the receptive field for dilated convolution is larger as compared to the standard convolution. The size of the receptive field indicates how much context information we use. In PSPNet, the last two blocks of the backbone have dilation values 2 and 4 respectively.

PSPNet Decoder

After the encoder has extracted out features of the image, it is the turn of the decoder to take those features and convert them into predictions by passing them into its layers. The decoder is just another network which takes in features and results into predictions.

8x upsampling decoder

The PSPNet model is not a complete segmentation model in itself, it is just an encoder, which means it is just half of what is required for image segmentation. The most common decoders that are found in various implementations of PSPNet is a convolution layer followed by a 8x bilinear-upsampling.

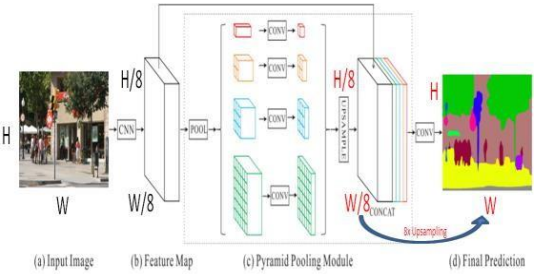


Fig 6.5: PSPNet with 8x upsampling decoder

There is a downside of having a 8x upsmapling decoder in the end is that there are no learnable parameters in them hence the results that we get are blobby and it fails to capture high resolution information from the image.

6.2 : ADVANTAGES AND DISADVANTAGES OF PSPNET

Advantages:

• Multiple data points: PSPNet leverages a wide variety of data points from the pyramid pooling module, allowing the model to understand objects in the environment and improve segmentation accuracy.

• Decision Making: Using pyramid pooling and feature aggregation, PSPNet can generate detailed segmentations with good boundaries by performing accurate product analysis and evaluation.

• Optimization for scaled objects: PSPNet can handle objects of different scales and sizes, adapting to situations where objects may appear from different camera angles or present a big change.

• Strong Semantic Segmentation Performance: PSPNet demonstrates the performance of semantic partitioning tasks and achieves good results on a variety of models and data, including underwater image partitioning

Contextual Reasoning and Global Information Integration: Pyramid pooling and feature aggregation mechanisms allow PSPNet to reason about objects in the entire scene context and

combine global information to improve partitioning results.

Multi-field adaptability: PSPNet's design and structure make it applicable to many fields beyond underwater photography. It has been successfully applied to many computer vision tasks such as scene analysis, medical image analysis, and remote sensing.

Disadvantages:

High Computing Requirements: PSPNet's complex architecture is computationally demanding, especially due to the work of pyramid pooling and feature merging. Training and decision making can require significant computing resources and limit its effectiveness on limited hardware.
Large memory space: PSPNet's memory space can be large, especially when processing high resolution images or processing large files. This can cause problems for deployment on devices with limited memory.
Do not rely on large datasets: Like many deep learning models, PSPNet relies on large datasets for training. Obtaining and interpreting diverse and comprehensive data can be time consuming and expensive, especially in specialized fields such as underwater photography.
Susceptibility to noise or poor quality data: PSPNet's performance may suffer for poor quality or underwater images. Poor visibility, noise, and artifacts can adversely affect segmentation accuracy by requiring additional steps or data augmentation techniques.
Interpretation and Interpretation: Deep learning models, including PSPNet, are often criticized for their lack of interpretation and explanation. The complexity of the model makes the decision difficult to understand and explain, which may limit its usefulness in some applications where interpretation is important.
Limitation of the ability to blind cases: When applied to blind cases or areas different from the training data, the performance of the PSPNet may degrade.
Adapting to a new environment or managing change may require additional strategies such as adaptive or adaptive learning.

6.3 : CONCLUSION AND FUTURE WORK :

A layered layer fusion semantic segmentation method PSPNet based on the pyramid pool, which can reduce the error of the model, is proposed. After the images are extracted by the backbone feature extraction network, overall fine point information is obtained with the pyramid pool, and the shallow features of the similarity are further fused during the decoding process to support the information of the feature map. In the feature extraction process, together with the channel display mechanism and the spatial attention mechanism, different parts of the feature map are given weight, the expression of the features is reliable, and the perception of the world features is improved. achieve the goal. improving the segmentation effect. Experiments show that the proposed PSPNet model can classify high-quality and useful images of publicly available data. The rise of deep learning has led to the rapid growth of computer vision.

Although deep learning segmentation algorithms have solved many segmentation problems, there are still some flaws. In the future, the accuracy and speed of the current model can be further improved through refinement and improvement. In addition, less research has been done on the incorporation of increasingly sophisticated elements in this area. In this case the correct segmentation will be affected and there may be differences in brightness, overlapping of multiple targets etc. will lead to negative consequences. Therefore, the generalization ability of the model should be studied further.

7. Effective Underwater mine detection Techniques using Resnet-50 algorithm

ResNet is a conventional neural network for numerous computer vision
applications/tasks that is short for Residual Networks. The major advance with ResNet was that we were able to effectively train incredibly deep neural networks with over 150 layers.
Extensive training of neural networks was extremely hard because of the disappearance
problem before ResNet. The notion of skip
connection was originally developed. The
following graphic shows the connection skipping. The picture on the right, we always build
convolution layers like before, but now we also add the original input to the output.The picture
on the left stacks the layers of convolution one by one. This is called skip connection.

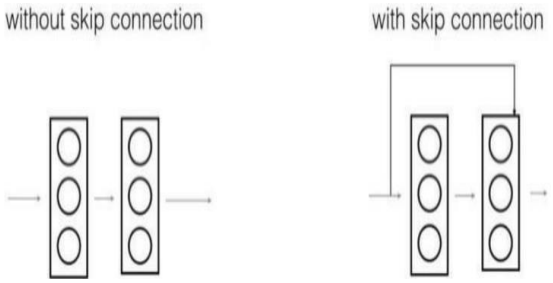


Fig 7.1: Skip Connections in resnet 50

Two lines of code can be written. X shape = X # X Store in a variable the initial value of X. ## Convert + batch standards on X. ## X = Add() # SKIP Connection ([X, X shortcut]). The coding is straightforward but one crucial aspect is that, given that X, X shortcut above aretwo matrices, we can only add them if they have the same form. Then, if the operations of the Convolution + batch standard are performed to the same output form, we may add them simply.

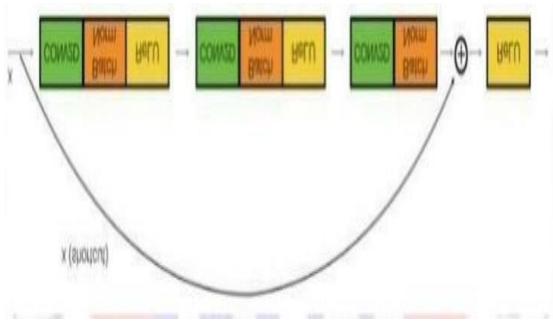


Fig 7.2: Resnet Architecture explaining how skip connections work.

7.1 : LITERATURE SURVEY

The detection of mines has made extensive use of underwater image processing. Utilising autonomous underwater vehicles (AUVs), this is accomplished. These AUVs are placed in the mine detection area to collect mine location data. The base receives the obtained data and takes the appropriate action there.Underwater camera sensors are used by these autonomous vehicles. These sensors have significant disadvantages. Due to noise and scattering, the AUV's underwater images are not accurate. The light transit properties of water and the biological activity on the sea floor are the main causes of this phenomenon.

It is annoying to use underwater optical imaging. When compared to regular photography, there are a lot of challenges. The scattering effect blurs underwater images. Due to wavelength absorption, the colour is diminished. The image that was obtained has noise and water traces. Since artificial lighting is not used in underwater photography, the quality of the image is diminished. In the daytime, the flicker cannot be avoided.

Side-Scan Sonar was the subject of research. The network is supplied the sonar signatures right here. For surveying, this method maps a substantial portion of the ocean floor. The data are first processed by the signal before the training method is applied to the network. After training, the sonar image is separated into subframes during segmentation. Then, using feature extraction, the object property is determined, and the data set is divided.

Mine detection is a barrier for side-scan sonar imaging. The environment in which the system is deployed may change. The image's accuracy suffers as a result. The target mines have a variety of shapes, which the side scan sonar finds challenging to handle. It is possible to misidentify underwater habitations like coral and reefs. These can even conceal things that resemble mines. The sonar signal's slow detection time makes it also unreliable. The side scan sonar maps the target object after sending a sonar wave to it in order to collect data. The process takes a long time, and side scan sonar is very imprecise in rough ocean conditions.

The concept of transformable template matching can also be used to find underwater objects. With this technique, feature extraction is done by creating a templet out of sonar video clips. This is accomplished by analysing acoustic shadows and identifying areas. The target region is found using a method called fast saliency detection. The following step is to extract the normalised gradient feature, which is followed by calculating how similar the target and template are.
Studying the Threshold-based approach revealed that it has historically been widely employed. For mine detecting purposes, this model is not the most effective. The active contour model in the model requires a lot of calculation, and it has an impact on the initial contour.

The sonar image of the model is subpar. Accurate recognition is not enabled by noise and deceptive targets. Towards the Future The integral-image representation is used in sonar imagery. In a short period of time, this offers competitive features. There is a significant reduction in computational load. Small areas of the image are the focus of the piece. The real-time object detecting capabilities of this technique are good. The increased demand for real-time signal processing makes it difficult to detect. There is no density filtering in the method. Here, the algorithm must disregard the fact that multiple mines will be located close to one another.

Mine-cast shadows that resemble other objects are also picked up.

Another technique for underwater image detection is multi beam sonar image processing. This technique makes advantage of the BluView (BV) Sonar. Real-time data is gathered, turned into a picture, and then preprocessed. The foreground and background are distinguished via the contour detection technique. The algorithm used in this process tracks objects. The particle filter tracking approach is used to achieve this. The tracking strategy uses an adaptive fusion tracking strategy. The limited contrast and considerable noise in sonar images make the method problematic.

Another intriguing technique for researching the underwater objects is the adaptive fuzzy neural network. Here, pattern recognition and feed forward are employed. A key tool utilised in its creation is MATLAB. It computes texture features. Features like autocorrelation, total variance, sum average, and sum entropy are used to categorise objects. The model is tested for classification after the texture feature has been trained.

Sensors with monocular vision are used to find objects beneath water. In this technology, light transmission is utilised.The global contrast feature is used to identify the Region of Interest (ROI). To construct the dataset, a monocular camera was employed. The testing model is advanced as various photos are produced. It can reduce noise and improve the system's precision. The approach has shortcomings. The employed camera experiences intensity degradation. The colour distortion and hazy effect compound this drawback.

7.2 Effective Underwater mine detection Techniques using VGG16 algorithm

VGG16 is the architecture of the CNN. VGG16 Until now, it has been considered one of the best vision model designs. Most unusual about VGG16 is that they
concentrated on having 3x3 filter coalescing layers with step 1, and always have the same
padding and maxpool layer of 2x2 step filter 2 instead of having a lot of hyper parameters. It
is included throughout the design of this combination of convolution and max pool layers. It finally has two FC layers (completely linked) and a softmax for output. The 16 in VGG16
refersto a weight of 16 layers. The network is a big network that has around 138 million (about) parameters.

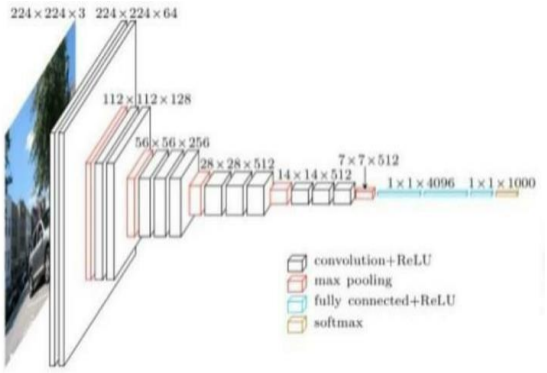


Fig 7.3: Architecture of VGG16

The layer input to a conv is 224 × 224 RGB picture. This picture passes through a stack of
convolutionary (conv.) layers, in which a tiny
reception field is employed with filters:3×3
(which is the smallest size to capture the notion of left/right, up/down, center). It also uses 1 di 1 convolution filters in one configuration that may be considered as a linear change of the input channels (followed by non-linearity). The tapered strand is set at 1 pixel; the spatial padding is done using the input of the core layer such that after tapering, the spatial resolutionis kept. Five levels of max pooling which follow some of the levels are spatial pooling (notall the conv.layers are followed by max-pooling). A 2-2 pixel window with step 2 allows max pooling.Three fully connected (FC) layers follow a stack of convolutive (different) layers . The lastlayer is the layer of soft-max. The completely linked layers in all networks are the same configuration. The non-linearity of all buried layers (ReLU) is provided. It is also highlighted that none (save one) of the networks have Local Answer Normalization (LRN), nor does such standardization improve ILSVRC data set performance, but leads to increased memory consumption and calculation time.

Table 7.1: Summary of MLO detection and classification methods reviewed in this article.

Application	Authors	Year	Technique	Remarks
Detection Classification	Tucker et al.	2007	Classical image processing	Employments canonical relationship investigation Tall exactness (88%).
Detection	Barngrover et al.	2016	ML	Employments brain–computer interface, Haar-like highlight classifier and SVM.
Detection	Thanh et al.	2020	DL	Uses Gabor-based detector Achieves competitive performance compared to the existing approaches.
Detection	Attaf et al.	2016	Classical image processing	Livelihoods the adequacy overpowering component analysis Exploits the saliency of sonar pictures.
Detection Classification	Saisan et al.	2008	ML	Treats mine discovery as a two-dimensional question acknowledgment and localisation issue Appears great comes about on benchmarking information made from the mine dataset.
Detection	Abu et al.	2019	ML	Employments a back vector machine over the factual features.Does not require information almost the target's shape or estimate.
Detection	Rao et al.	2009	ML	Utilized for real-time application Great comes about on the database given by the Maritime Surface Warfare Center.
Detection Classification	Reed et al.	2003	Classical image processing	Employments unsupervised Markov irregular field model. Utilises a priori data almost the spatial relationship between highlights and shadows.
Classification	Fei et al.	2015	DL	Uses an ensemble learning scheme in the Dempster–Shafer theory framework
Detection	Acosta et al.	2015	Classical image processing	Employments Cell Average–Constant False Alarm Rate Suitable for independent submerged vehicles

Table 7.1: *Cont.*

Application	Authors	Year	Technique	Remarks
Classification	Neumann et al.	2008	Classical image processing	Employments the Hough change Altogether diminishes the number of wrong discoveries
Classification	McKay et al.	2017	DL	Employments convolutional neural systems and exchange learning
Classification	Yao et al.	2002	DL	Employments include determination and neural arrange classifier Introduces a sub-band combination component for wideband information
Classification	Dobeck et al.	1997	DL	Employments the k-nearest neighbour attractor-based neural network classifier
Detection Classification	Ciany et al.	2003	Classical image processing	Employments signal-to-noise proportion and shape of the objects Suitable for independent submerged vehicles
Detection Classification	Williams et al.	2016	DL	Employments profound systems learned for a few parallel classification errands
Classification	Galusha et al	2019	DL	Employments convolutional neural systems for area based on cross-validation
Detection Classification	Dura et al.	2008	Classical image processing	Employments a superellipse fitting approach. The classification rate is higher than 80%.
Detection	Wu et al.	2019	DL	Employments the productive convolutional arrange engineering for semantic division.

Table 7.1: *Cont.*

Application	Authors	Year	Technique	Remarks
Semantic segmentation	Yanming Guo Yu Liu ·	2017	CNN	Deep learning advancements revolutionize image segmentation, impacting various visual tasks.
Semantic segmentation	Rishipal Singha , Rajneesh Rani	2020	DL AND DCNN	This paper explores DCNNs for semantic segmentation and emphasizes the need for improved real-time performance.
Semantic segmentation	Xiaolong Liu1 · Zhidong Deng	2018	CNN AND FCN	Comprehensive review of recent progress in DCNN-based semantic image segmentation methods.
instance segmentation	Wenchao Gu a , Shuang Bai	20212	DCNN	This paper discusses instance segmentation, including evaluation, methods, backbones, and future directions
instance segmentation	Kuo-Kun Tseng a , Jiangrui Lin	20212	CNN	A fast instance segmentation algorithm combining YOLOv3 and Mask-RCNN is proposed, with improved speed and accuracy.
instance segmentation	Farhana Sultana a , Abu Sufia	20202	DCNN AND CNN	This article provides an overview of image segmentation models, their optimization, and application areas
instance segmentation	Hao Chena,∗ , Xiaojuan Q	2016	Object detection	Deep contour-aware network improves histological object segmentation with superior performance.

8. CONCLUSION AND FUTURE WORK

In summary, U-Net, Fully Convolutional Network (FCN), Semantic Segmentation, PSPNet (Pyramid Scene Parsing Network) and Image Segmentation are all important and useful concepts in the field of computer vision, especially later in image segmentation.

U-Net is a powerful tool for image segmentation, especially in the field of biomedical imaging, as it can capture local and global features by shrinking and expanding. It excels in tasks that require a clear definition of the structure and can be adapted to many different situations with the necessary modifications.
FCN, including

U-Net, revolutionized image segmentation by providing end-to-end pixel-level prediction. They replace all layers with convolutional layers, so spatial information is preserved.
FCNs have proven effective in many applications and are widely used as the base model for semantic segmentation tasks.

Semantic segmentation aims to assign a semantic tag to each pixel in an image and provide pixel-level detailed information about the image. FCNs with U-Net are often used in semantic segmentation tasks because of their ability to capture spatial information and learn discrimination.

PSPNet (Pyramid Scene Parsing Network) is another well-known semantic segmentation architecture. It includes a pyramid pooling module to capture multiple data points, providing more powerful and accurate segmentation.
PSPNet performs competitively on a variety of metrics and is particularly useful for capturing global content.

Determining which of these methods is better depends on specific tasks, data characteristics, and performance measures. U-Net is advantageous where accurate localization and detailed boundary information are important. FCNs, including U-Net, are versatile and can be widely used in many fields. PSPNet is good at capturing the global context, which is important for projects that require a comprehensive understanding of the situation.
Ultimately, the choice of the most appropriate method depends on the specific requirements of the application and the balance between accuracy, efficiency and budget. It is recommended to evaluate and compare these project datasets and project models to determine the best approach for a given situation.

Regarding future work of U-Net, FCN, Semantic Segmentation, PSPNet, and Image Segmentation, there are several areas that researchers can focus on to further develop these models:

Architectural Enhancements: One of the aspects of future work is to explore Architecture enhancements and enhancements. This may include combining monitoring systems, spatial and channel monitoring modules, or cross-linking to improve data flow and performance.

Efficiency and Speed: Although this system performs well, there are still areas for improvement in terms of efficiency and speed.
Future work may focus on developing more robust modeling techniques or exploring techniques such as network convolution and quantization to reduce model size and speed up inference time.

Domain Adaptation and Generalization: Developing the general capabilities of these processes is important for practical applications. Future studies may explore transfer and transfer learning strategies to improve the performance of the model on different data and different image formats.

Weakly Supervised and Semi-supervised Learning: Training deep learning models for image segmentation often requires a lot of recording data. Future work may explore unsupervised and semi-supervised learning strategies where models can use limited or incomplete explanations to better perform partitioning tasks.
Understanding context: While semantic segmentation focuses on pixel-level classification,

combined with higher context understanding can improve performance. Future research may explore ways to capture long-term dependencies and integrate data points from different scales and abstraction levels.

Deciding which method is better between U-Net, FCNs, semantic segmentation and PSPNet depends on the specific task, dataset and evaluation criteria. Each method has advantages and limitations. It is recommended to evaluate and compare these project data and project models to determine which method is more accurate, efficient and suitable for the specific applications designed.
Overall, future work in these areas aims to advance image segmentation by solving current limitations and challenges, well improving model performance, and extending the applicability of this process to many places and situations around the world.

9.REFERENCES

1. Del Rio Vera, J.; Coiras, E.; Groen, J.; Evans, B. Automatic Target Recognition in Synthetic Aperture Sonar Images Based on
Geometrical Feature Extraction. *EURASIP J. Adv. Signal Process.* 2009, *2009*, 109438. https://doi.org/10.1155/2009/109438.
2. Dura, E.; Bell, J.; Lane, D. Superellipse Fitting for the Recovery and Classification of Mine-Like Shapes in Sidescan Sonar Images.
IEEE J. Ocean. Eng. 2008, *33*, 434–444. https://doi.org/10.1109/JOE.2008.2002962.
3. Neupane, D.; Seok, J. A review on deep learning-based approaches for automatic sonar target recognition. *Electronics* 2020, *9*,
1972.
4. Barngrover, C.; Kastner, R.; Belongie, S. Semisynthetic versus real-world sonar training data for the classification of mine-like
objects. *IEEE J. Ocean. Eng.* 2015, *40*, 48–56. https://doi.org/10.1109/JOE.2013.2291634.
5. Cerqueira, R.; Trocoli, T.; Neves, G.; Joyeux, S.; Albiez, J.; Oliveira, L. A novel GPU-based sonar simulator for real-time
applications. *Comput. Graph.* 2017, *68*, 66–76. https://doi.org/10.1016/j.cag.2017.08.008.
6. Borawski, M.; Forczmański, P. Sonar Image Simulation by Means of Ray Tracing and Image Processing. In *Enhanced Methods in*
Computer Security, Biometric and Artificial Intelligence Systems; Kluwer Academic Publishers: Boston, MA, USA, 2005; pp. 209–
214.
7. Saeidi, C.; Hodjatkashani, F.; Fard, A. New tube-based shooting and bouncing ray tracing method. In Proceedings of the 2009
International Conference on Advanced Technologies for Communications, Hai Phong, Vietnam, 12–14 October 2009; pp. 269–
273.
8. Danesh, S.A. *Real Time Active Sonar Simulation in a Deep Ocean Environment*; Massachusetts Institute of Technology: Cambridge,
MA, USA, 2013.

9. Saito, H.; Naoi, J.; Kikuchi, T. Finite Difference Time Domain Analysis for a Sound Field Including a Plate in Water. *Jpn. J. Appl. Phys.* 2004, *43*, 3176–3179. https://doi.org/10.1143/JJAP.43.3176.
10. Maussang, F.; Rombaut, M.; Chanussot, J.; Hétet, A.; Amate, M. Fusion of local statistical parameters for buried underwater mine detection in sonar imaging. *EURASIP J. Adv. Signal Process.* 2008, *2008*, 1–19. https://doi.org/10.1155/2008/876092.
11. Maussang, F.; Chanussot, J.; Rombaut, M.; Amate, M. From Statistical Detection to Decision Fusion: Detection of Underwater Mines in High Resolution SAS Images. In *Advances in Sonar Technology*; IntechOpen: London, UK, 2009; ISBN 9783902613486.
12. Lurton, X. *An Introduction to Underwater Acoustics*; Springer: Berlin/Heidelberg, Germany, 2010; ISBN 978-3-540-78480-7.
13. Barngrover, C.M. *Automated Detection of Mine-Like Objects in Side Scan Sonar Imagery*; University of California: San Diego, CA, USA, 2014.
14. Tellez, O.L.L.; Borghgraef, A.; Mersch, E. The Special Case of Sea Mines. In *Mine Action—The Research Experience of the Royal Military Academy of Belgium*; InTechOpen: London, UK, 2017; pp. 267–322.
15. Doerry, A. Introduction to Synthetic Aperture Sonar. In Proceedings of the 2019 IEEE Radar Conference (RadarConf), Boston, MA, USA, 22–26 April 2019; pp. 1–90.
16. Atherton, M. *Echoes and Images, The Encyclopedia of Side-Scan and Scanning Sonar Operations*; OysterInk Publications: Vancouver, BC, Canada, 2011; ISBN 098690340X.
17. Rao, C.; Mukherjee, K.; Gupta, S.; Ray, A.; Phoha, S. Underwater mine detection using symbolic pattern analysis of sidescan sonar images. In Proceedings of the 2009 American Control Conference, St. Louis, MO, USA, 10–12 June 2009; pp. 5416–5421. https://doi.org/10.1109/ACC.2009.5160102.
18. Tucker, J.D.; Azimi-Sadjadi, M.R.; Dobeck, G.J. Canonical Coordinates for Detection and Classification of Underwater Objects From Sonar Imagery. In Proceedings of the OCEANS 2007—Europe, Aberdeen, UK, 18–21 June 2007; pp. 1–6.
19. Reed, S.; Petillot, Y.; Bell, J. An automatic approach to the detection and extraction of mine features in sidescan sonar. *IEEE J. Ocean. Eng.* 2003, *28*, 90–105. https://doi.org/10.1109/JOE.2002.808199.
20. Klausner, N.; Azimi-Sadjadi, M.R.; Tucker, J.D. Underwater target detection from multi-platform sonar imagery using multichannel coherence analysis. In Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 11–14 October 2009; pp. 2728–2733.
21. Langner, F.; Knauer, C.; Jans, W.; Ebert, A. Side scan sonar image resolution and automatic object detection, classification and identification. In Proceedings of the OCEANS ’09 IEEE Bremen: Balancing Technology with Future Needs, Bremen, Germany, 11–14 May 2009; pp. 1–8.
22. Hożyń, S.; Zalewski, J. Shoreline Detection and Land Segmentation for Autonomous Surface Vehicle Navigation with the Use of an Optical System. *Sensors* 2020, *20*, 2799. https://doi.org/10.3390/s20102799.
23. Hożyń, S.; Żak, B. Local image features matching for real-time seabed tracking applications. *J. Mar. Eng. Technol.* 2017, *16*, 273–282. https://doi.org/10.1080/20464177.2017.1386266.
24. Wang, X.; Wang, H.; Ye, X.; Zhao, L.; Wang, K. A novel segmentation algorithm for side-scan sonar imagery with multi-object. In Proceedings of the 2007 IEEE International Conference on Robotics and Biomimetics (ROBIO), Sanya, China, 15–18 December 2007; pp. 2110–2114.
25. Om, H.; Biswas, M. An Improved Image Denoising Method Based on Wavelet Thresholding. *J. Signal Inf. Process.* 2012,
26.Hoz ̇yn ́,S.;Z ̇ak,B.SegmentationAlgorithmUsing MethodofEdgeDetection.SolidStatePhenom.2013,1 96,206–211.[CrossRef]
27. Celik, T.; Tjahjadi, T. A Novel Method for Sidescan Sonar Image Segmentation. IEEE J. Ocean. Eng. 2011, 36, 186–194. [CrossRef]
28. Wei, S.; Leung, H.; Myers, V. An automated change detection approach for mine recognition using sidescan sonar data. In Proceedings of the 2009 IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 11–14 October 2009; pp. 553–558.
29. Neumann, M.; Knauer, C.; Nolte, B.; Brecht, D.; Jans, W.; Ebert, A. Target detection of man made objects in side scan sonar images segmentation based false alarm reduction. J. Acoust. Soc. Am. 2008, 123, 3949. [CrossRef]
30. Huo, G.; Yang, S.X.; Li, Q.; Zhou, Y. A Robust and Fast Method for Sidescan Sonar Image Segmentation Using Nonlocal Despeckling and Active Contour Model. IEEE Trans. Cybern. 2017, 47, 855–872. [CrossRef]

31. Acosta, G.G.; Villar, S.A. Accumulated CA–CFAR Process in 2-D for Online Object Detection From Sidescan Sonar Data. IEEE J.
Ocean. Eng. 2015, 40, 558–569. [CrossRef]
32. Ye, X.-F.; Zhang, Z.-H.; Liu, P.X.; Guan, H.-L. Sonar image segmentation based on GMRF and level-set models. Ocean Eng. 2010,
37, 891–901. [CrossRef]
33. Fei, T.; Kraus, D. An expectation-maximisation approach assisted by Dempster-Shafer theory and its application to sonar image
segmentation. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),
Kyoto, Japan, 25–30 March 2012; pp. 1161–1164. [CrossRef]
34. Szymak, P.; Piskur, P.; Naus, K. The Effectiveness of Using a Pretrained Deep Learning Neural Networks for Object Classification
in Underwater Video. Remote Sens. 2020, 12, 3020. [CrossRef]
35. Huo, G.; Wu, Z.; Li, J. Underwater Object Classification in Sidescan Sonar Images Using Deep Transfer Learning and Semisynthetic
Training Data. IEEE Access 2020, 8, 47407–47418. [CrossRef]
36. Attaf, Y.; Boudraa, A.O.; Ray, C. Amplitude-based dominant component analysis for underwater mines extraction in side scans
sonar. In Proceedings of the Oceans 2016—Shanghai, Shanghai, China, 10–13 April 2016. [CrossRef]
37. Wu, M.; Wang, Q.; Rigall, E.; Li, K.; Zhu, W.; He, B.; Yan, T. ECNet: Efficient Convolutional Networks for Side Scan Sonar Image Segmentation. Sensors 2019, 19, 2009. [CrossRef] [PubMed]
38. Abu, A.; Diamant, R. A Statistically-Based Method for the Detection of Underwater Objects in Sonar Imagery. IEEE Sens. J. 2019,
19, 6858–6871. [CrossRef]
39. McKay, J.; Monga, V.; Raj, R.G. Robust Sonar ATR Through Bayesian Pose-Corrected Sparse Classification. IEEE Trans. Geosci.
Remote Sens. 2017, 55, 5563–5576. [CrossRef]
40. Williams, D.P. Underwater target classification in synthetic aperture sonar imagery using deep convolutional neural networks.
In Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016;
pp. 2497–2502. [CrossRef]
41. Ciany, C.M.; Zurawski, W.C.; Dobeck, G.J.; Weilert, D.R. Real-time performance of fusion algorithms for computer aided detection
and classification of bottom mines in the littoral environment. In Proceedings of the Oceans 2003. Celebrating the Past . . .
Teaming Toward the Future (IEEE Cat. No.03CH37492), San Diego, CA, USA, 22–26 September 2003; Volume 2, pp. 1119–1125.
42. Saisan, P.; Kadambe, S. Shape normalised subspace analysis for underwater mine detection. In Proceedings of the 2008 15th IEEE
International Conference on Image Processing, San Diego, CA, USA, 12–15 October 2008; pp. 1892–1895.
43. Thanh Le, H.; Phung, S.L.; Chapple, P.B.; Bouzerdoum, A.; Ritz, C.H.; Tran, L.C. Deep gabor neural network for automatic
detection of mine-like objects in sonar imagery. IEEE Access 2020, 8, 94126–94139. [CrossRef]

DrillBit

The Report is Generated by DrillBit Plagiarism Detection Software

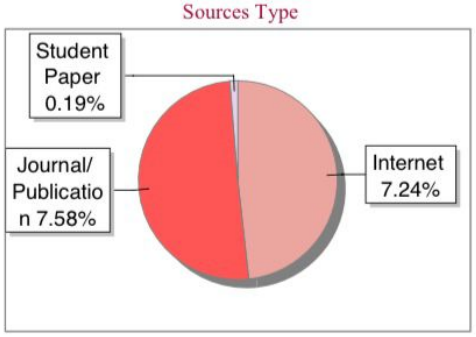
Submission Information

Author Name	moin
Title	paper4
Paper/Submission ID	757393
Submission Date	2023-05-31 08:37:44
Total Pages	18
Document type	Research Paper

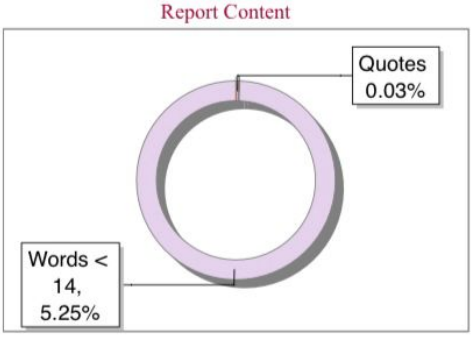
Result Information

Similarity **15 %**

Sources Type



Report Content



Exclude Information

Quotes	Not Excluded
References/Bibliography	Not Excluded
Sources: Less than 14 Words Similarity	Not Excluded
Excluded Source	**0 %**
Excluded Phrases	Not Excluded

A Unique QR Code use to View/Download/Share Pdf File

DrillBit

DrillBit Similarity Report

15	63	B	A-Satisfactory (0-10%) B-Upgrade (11-40%) C-Poor (41-60%) D-Unacceptable (61-100%)
SIMILARITY %	MATCHED SOURCES	GRADE	

LOCATION	MATCHED DOMAIN	%	SOURCE TYPE
1	www.mdpi.com	2	Internet Data
2	towardsdatascience.com	2	Internet Data
3	Emerging artificial intelligence methods in structural engineering by Salehi-2018	1	Publication
4	A deep neural networks based model for uninterrupted marine environment monitori by G-2020	1	Publication
5	Encoder-decoder CNN models for automatic tracking of tongue contours in real-tim by Hame-2020	1	Publication
6	A Comparison Different DCNN Models for Intelligent Object Detection i by Ding-2018	<1	Publication
7	A review of feature selection techniques in sentiment analysis, by Ahmad, Siti Rohaida- 2019	<1	Publication
8	Hierarchical Features Driven Residual Learning for Depth Map Super-Re, by Guo, Chunle Li, Ch- 2018	<1	Publication
9	IEEE 2011 IEEE International Conference on Computer Vision Workshops	<1	Publication
10	Explosives Detection by Terahertz SpectroscopyA Bridge Too Far by Kemp-2011	<1	Publication
11	www.acm.org	<1	Internet Data

12	www.atlantis-press.com	<1	Publication
13	www.projectpro.io	<1	Internet Data
14	hess.copernicus.org	<1	Publication
15	docplayer.net	<1	Internet Data
16	IEEE 2018 25th National and 3rd International Iranian Conference on, by Mohammadi, Elnaz O- 2018	<1	Publication
17	arxiv.org	<1	Publication
18	Pattern Recognition and Classification Introduction, by Dougherty, Geoff- 2013	<1	Publication
19	malaevents.in	<1	Internet Data
20	surge.iitk.ac.in	<1	Publication
21	dochero.tips	<1	Internet Data
22	moam.info	<1	Internet Data
23	bmcgenomics.biomedcentral.com	<1	Internet Data
24	deepai.org	<1	Internet Data
25	arxiv.org	<1	Publication
26	www.ncbi.nlm.nih.gov	<1	Internet Data
27	IEEE 2017 9th International Conference on Information Technology an, by Fazekas, Szilard Zs- 2017	<1	Publication
28	IEEE 2017 IEEE 25th Annual Symposium on High-Performance Interconne, by Lu, Xiaoyi Shi, Ha- 2017	<1	Publication
29	1library.co	<1	Internet Data

30	Advances in Machine Learning and Data Science by Damoda-2018	<1	Publication
31	DisepNet for breast abnormality recognition by Yu-2021	<1	Publication
32	ijrte.org	<1	Publication
33	koreascience.or.kr	<1	Publication
34	towardsdatascience.com	<1	Internet Data
35	www.doaj.org	<1	Publication
36	www.researchgate.net	<1	Internet Data
37	www.sciencepubco.com	<1	Publication
38	Characterization of lenticulostriate arteries with high resolution black-blood T by Ma-2019	<1	Publication
39	mdpi.com	<1	Internet Data
40	mdpi.com	<1	Internet Data
41	qdoc.tips	<1	Internet Data
42	Solving the over segmentation problem in applications of Watershed Transform by Gonzalez-2013	<1	Publication
43	www.dx.doi.org	<1	Publication
44	IEEE 2018 10th International Conference on Knowledge and Systems En, by Nguyen, Mau Uyen D- 2018	<1	Publication
45	IEEE 2018 Eighth International Conference on Image Processing Theor, by Jiang, Xiaoyue Du,- 2018	<1	Publication
46	Deep learning based automatic detection of uninformative images in pulmonary opt by Brochet-2020	<1	Publication

47	mdpi.com	<1	Internet Data
48	citeseerx.ist.psu.edu	<1	Internet Data
49	www.indiamart.com	<1	Internet Data
50	afflictor.com	<1	Internet Data
51	arxiv.org	<1	Publication
52	Demand for Fresh Vegetables in the United States 19702010	<1	Student Paper
53	docplayer.net	<1	Internet Data
54	Effective Deep Learning Approaches for Summarization of Legal Texts by Anand-2019	<1	Publication
55	ijcea.com	<1	Publication
56	Lifetime stress exposure and health A review of contemporary assessment methods by Shields-2017	<1	Publication
57	moam.info	<1	Internet Data
58	Semi-supervised learning with generative model for sentiment classification of s by Duan-2020	<1	Publication
59	The Future of Quality Measurement in the United States by Yi-2014	<1	Publication
60	towardsdatascience.com	<1	Internet Data
61	Wavelets for texture analysis an overview	<1	Student Paper
62	www.atlantis-press.com	<1	Internet Data
63	www.mecs-press.org	<1	Publication