

Article

A Fast Maritime Target Identification Algorithm for Offshore Ship Detection

Jinshan Wu ¹, Jiawen Li ^{1,2}, Ronghui Li ^{1,2,*} , Xing Xi ³, Dongxu Gui ⁴ and Jianchuan Yin ^{1,2}

¹ Maritime College, Guangdong Ocean University, Zhanjiang 524000, China; jinshan_wu@126.com (J.W.); jiawen-li@gdou.edu.cn (J.L.); yinjianchuan@gdou.edu.cn (J.Y.)

² Technical Research Center for Ship Intelligence and Safety Engineering of Guangdong Province, Zhanjiang 524000, China

³ College of Computer, Guangdong University of Technology, Guangzhou 510006, China; xxyzll@yeah.net

⁴ College of Information Technology, Jilin Agricultural University, Changchun 130000, China; gdx_jlau@163.com

* Correspondence: lironghui@163.com

Abstract: The early warning monitoring capability of a ship detection algorithm is significant for jurisdictional territorial waters and plays a key role in safeguarding the national maritime strategic rights and interests. In this paper, a Fast Maritime Target Identification algorithm, FMTI, is proposed to identify maritime targets rapidly. The FMTI adopts a Single Feature Map Fusion architecture as its encoder, thereby improving its detection performance for varying scales of ship targets, from tiny-scale targets to large-scale targets. The FMTI algorithm has a decent detection accuracy and computing power, according to the mean average precision (mAP) and floating-point operations (FLOPs). The FMTI algorithm is 7% more accurate than YOLOF for the mAP measure, and FMTI's FLOPs is equal to 98.016 G. The FMTI can serve the demands of marine vessel identification while also guiding the creation of supplemental judgments for maritime surveillance, offshore military defense, and active warning.

Keywords: YOLOF; target identification; maritime identification; monitoring and early warning



Citation: Wu, J.; Li, J.; Li, R.; Xi, X.; Gui, D.; Yin, J. A Fast Maritime Target Identification Algorithm for Offshore Ship Detection. *Appl. Sci.* **2022**, *12*, 4938. <https://doi.org/10.3390/app12104938>

Academic Editors: Jinfen Zhang, Yong Ma, Zhouhua Peng and Le Qi

Received: 13 April 2022

Accepted: 11 May 2022

Published: 13 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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

The maritime environment and international environment are becoming increasingly complex and changing, resulting in a rise in the type of vessels. The diversification of maritime targets, including warships, fishing vessels, cargo ships, etc., presents a great challenge for effective territorial water management [1]. Automatic ship detection algorithms are essential for effective territorial water surveillance. In a variety of sea conditions, this algorithm is capable of reliably recognizing arriving and leaving vessels. Currently, China's marine surveillance monitoring methods are divided into two groups based on how the information is obtained—active and passive approaches. The active methods get information from radar, video surveillance, remote sensing, underwater sonar, etc. On the other hand, the passive approach is usually used for the ship's automated identifying system.

Recently, deep learning technology has made great progress in various fields. Additionally, Convolutional Neural Networks (CNN) have made outstanding contributions in plenty of areas [2], including image classification and recognition, video recognition, etc. The traditional approach to feature extraction required manual tasks. Because different types of targets have different dependent features, the manual extraction approach is ineffective. CNN [3] can extract probable features automatically, saving time on human extracted features. Neuro networks along with improvements in big data technology have led to a shift in maritime target identification from wasteful manual monitoring methods to deep learning automatic identification methods.

Artificial intelligence advances are pushing the field of computer vision forward and providing practical answers to the problem of recognizing objects at sea. Many CNN-based recognition models have emerged, e.g., SSD [4], Faster RCNN [5], and YOLO [6]. There are two types of target detection algorithms—one-stage and two-stage ways. The end-to-end concept is used in the one-step method. After feature extraction, the image directly outputs the target class probability together with the location coordinate prediction frame, resulting in a faster detection. SSD and YOLO are typical representatives. The two-stage recognition procedure requires the identification of probable detection zones first and then classification of the objects. Two-stage profound network models are represented, e.g., RCNN series [7,8] and SPPNet [9] have higher recognition accuracy and are independent of factors such as perspective, light, and cover.

The main contributions of the article are as follows: (1) A novel detection algorithm, FMTI, is proposed for maritime target detection. (2) A multiscale feature fusion method is proposed to enrich the information of a single map. (3) FMTI can offer essential references for marine-related government functions to make their decisions.

The rest of this paper is organized as follows. Section 2 shows the related work. Section 3 demonstrates the methodology and the implementation details of our FMTI algorithm. The performance metrics and the experimental results are presented in Section 4. Finally, Section 5 concludes the work and proposes future works.

2. Related Work

Maritime rights and interests have once again become a focus of the world since the beginning of the 21st century, and the strategic position of coastal nations has been promoted like never before. Coastal countries constantly improve their ability to protect their unique marine interests and rights. They also aim to protect the coastal and marine ecosystems and further develop the marine economy.

Although technology has progressively become mature and the speed of recognition has gradually increased, the two-step algorithm fails to achieve real-time because the model is divided into multiple stages and calculates work redundantly. To resolve this problem, scholars have introduced the YOLO series [10,11], SSD, and other single-stage target detection methods; however, these approaches listed above improved the speed while reducing the accuracy.

Marine target detection work differs from land-based recognition in that maritime vehicles are constrained by waves, which impact vessel behavior [12], including six sorts of activities [13], such as surging, swaying, heaving, rolling, pitching, and yawing. The semantic data was markedly deficient. Thus, CNN-based target detection methods are available for maritime target recognition in natural scenes. CNN was constructed by several various layers, with the network being trained to understand the relationship between the data, and the model describes the mapping relationship between the input and the output data. Both traditional and deep learning target detection methods together constitute the current dominant target detection methods.

Many scholars have conducted research to obtain the best possible precision and speed in balance. Tello et al. [14] proposed a discrete wavelet transform-based method for ship detection that relied on statistical behavior differences between ships and surrounding sea regions to help in the interpretation of visual data, resulting in more reliable detection. Chang et al. [15] introduced the You Only Look Once version 2 (YOLO V2) approach to recognizing ships in SAR images with high accuracy to overcome the computationally costly accuracy problem. Chen et al. [16] used a combination of a modified Generative Adversarial Network (GAN) and a CNN-based detection method to achieve the accurate detection of small vessels. Contemporary technologies and models have limitations, such as the inability to recognize closed-range objects. Other scholars have performed outstanding work and contributed to the implementation of neural network algorithmic methods by migration to solve practical problems in different fields. Arcos-García et al. [17] analyze target detection algorithms (Faster R-CNN, R-FCN, SSD, and YOLO V2), combined

with some extractors (Resnet V1 50, Inception V2, darknet-19, etc.) to improve and adapt the traffic sign detection problem through migration learning. It is worth noting that ResNet's network structure has been introduced, resulting in exceptional performance in fields including non-stationary GW signal detection [18], Magnetic Resonance Imaging identification [19], the CT image recognition [20], and agricultural image recognition [21]. Feature Enrichment Object Detection (FEOD) framework with weak segmentation Loss based on CNN is proposed by Zhang et al. [22], the Focal Loss function is introduced to improve the algorithm performance of the algorithm. Li et al. [23] proposed a new decentralized adaptive neural network control method, using RBF neural networks, to deal with unknown nonlinear functions to construct a tracking controller. A new adaptive neural network control method was proposed by Li et al. [24] for uncertain multiple-input multiple-output (MIMO) nonlinear time-lag systems. To address the problem of Slow Feature Discriminant Analysis (SFDA) which cannot fully use discriminatory power for classification, Gu et al. [25] proposed a feature extraction method called Adaptive Slow Feature Discriminant Analysis (ASFDA). A fast face detection method based on convolutional neural networks to extract Discriminative Complete Features (DCFes) was proposed by Guo et al. [26] it detaches from image pyramids for multiscale feature extraction and improves detection efficiency. Liu et al. [27] establish a multitask model based on the YOLO v3 model with Spatial Temporal Graph Convolutional Networks Long Short-Term Memory to design a framework for robot-human interaction for judgment of human intent. To resolve the real-time problem of recognition, Zheng et al. [28] introduce an attention mechanism and propose a new attention mechanism-based real-time detection method for traffic police, which is robust. Yu et al. [29] proposed a multiscale feature fusion method based on bidirectional feature fusion, named Adaptive Multiscale Feature (AMF), which improves the ability to express multiscale features in backbone networks.

Additionally, scholars have been working on meaningful improvements based on them and making assistance for further enhancement of maritime target recognition. The integrated classifier MLP-CNN was proposed by Zhang et al. [30] to exploit the complementary results of CNN based on deep spatial feature representation and MLP based on spectral recognition to compensate for the limitations of object boundary delineation and loss of details of fine spatial resolution of CNN due to the use of convolutional filters. Sharifzadeh et al. [31] proposed a neural network with a hybrid CNN and multilayer perceptron algorithm for image classification, which detected target pixels based on the statistical information of adjacent pixels, trained with real SAR images from Sentinel-1 and RADARSAT-2 satellites, and obtained good performance. For the pre-processed data, Wu et al. [32] employed a support vector machine (SVM) classifier to classify the ships by assessing the feature vectors by calculating the average of kernel density estimates, three structural features, and the average backward scattering coefficients. Tao et al. [33] proposed a segmentation-based constant false alarm rate (CFAR) detection algorithm for multi-looked intensity SAR images, which solves the problems related to the target detection accuracy of the non-uniform marine cluster environment, and the detection scheme obtains good robustness on real Radarsat-2 MLI SAR images. Meanwhile, a robust CFAR detector based on truncation statistics was proposed by Tao et al. [34] for single- and multi-intensity synthetic aperture radar data to improve the target detection performance in high-density cases. SRINIVAS et al. [35] applied a probabilistic graphical model to develop a two-stage target recognition framework that combines the advantages of different SAR image feature representations and differentially learned graphical models to improve recognition rates by experimenting on a reference moving and stationary target capture and recognition dataset.

In order to tackle the collision avoidance problem for USVs in complex scenarios, Ma et al. [36] suggested a negotiation process to accomplish successful collision avoidance for USVs in complicated conditions. Li et al. [1] suggested employing the EfficientDet model for maritime ship detection and defined simple or complex settings with a positive recognition rate in the above circumstances, which provides an important reference for

maritime security defense. For USV systems with communication delays, external interference, and other issues, Ma et al. [37] suggested an event-triggered communication strategy. Additionally, an event-based switched USV control system is proposed, and the simulation results show that the proposed co-design process is effective.

Traditional target detection methods include different color specificities of own color space models and manual design to extract features. This method is susceptible to visual angle, light, etc., and has a large volume of computation, low recognition efficiency, and a slow speed, which cannot meet the requirements of detection efficiency, performance, and speed. Target detection based on deep learning brings a new trend for maritime target recognition.

The acquisition and transmission of maritime data are growing sophisticated and becoming crucial in maritime supervision increasingly. However, at this stage of maritime regulation, the active early warning technology is eager to improve. The early warning of proactive detection requires quick and efficient detection of surrounding targets, but an unavoidable problem is that it will be impacted by a reduction in detection speed, as the algorithm accuracy rate rises. Therefore, to balance the speed and accuracy of the algorithm detection, this paper adopts a deep learning technique to design an FMTI model for maritime vessel detection.

3. Methodology

The successful one-stage detector adopts a Feature Pyramid Network (FPN) owing to the divide-and-conquer scheme of the FPN for the optimizations in object detection, which has not employed multi-scale feature fusion. In terms of optimization, Chen et al. [38] introduced the You Only Look One-level Feature (YOLOF), instead of complex feature pyramids, only single-level features are applied for detection. Extensive experiments on the COCO benchmark are to verify the effectiveness of the proposed model. Additionally, the YOLOF model is partially updated to fit the demands of offshore operations, based on the research presented in this paper.

3.1. Process of FMTI

The FMTI algorithm is proposed in this paper for the detection of maritime targets, and its specific process is described subsequently. When there are one or more targets (including multiple targets) in the image to be recognized, the network is required to make a judgment for each prediction frame. Thus, the model divides the process into the following three steps.

1. The classified image is gridded and there are the Bounding Boxes (Bbox) in the grid cell. Each Bbox contains five features, $(x, y, w, h, Score_{confidence})$. Where (x, y) is the offset of the Bbox center relative to the cell boundary, (w, h) denotes the ratio of width and height in the whole image, and $Score_{confidence}$ is the Confidence Score.

$$Score_{Confidence} = Pr(object) \times GIOU_{pred}^{true} \quad (1)$$

$Pr(object)$ means whether the target exists or not. The existing value is 1, and the opposite value is 0.

$$Pr(object) = \begin{cases} 1, & \text{Exist.} \\ 0, & \text{None.} \end{cases} \quad (2)$$

The $GIOU$ [39] was optimized from the IOU, (Figure 1A). The intersection of Prediction and Ground Truth is shown by IOU. Where $Area(pred)$ denotes the area of the detection boxes and $Area(true)$ denotes the area of the true value.

$$IOU_{pred}^{true} = \frac{Area(true) \cap Area(pred)}{Area(true) \cup Area(pred)} \quad (3)$$

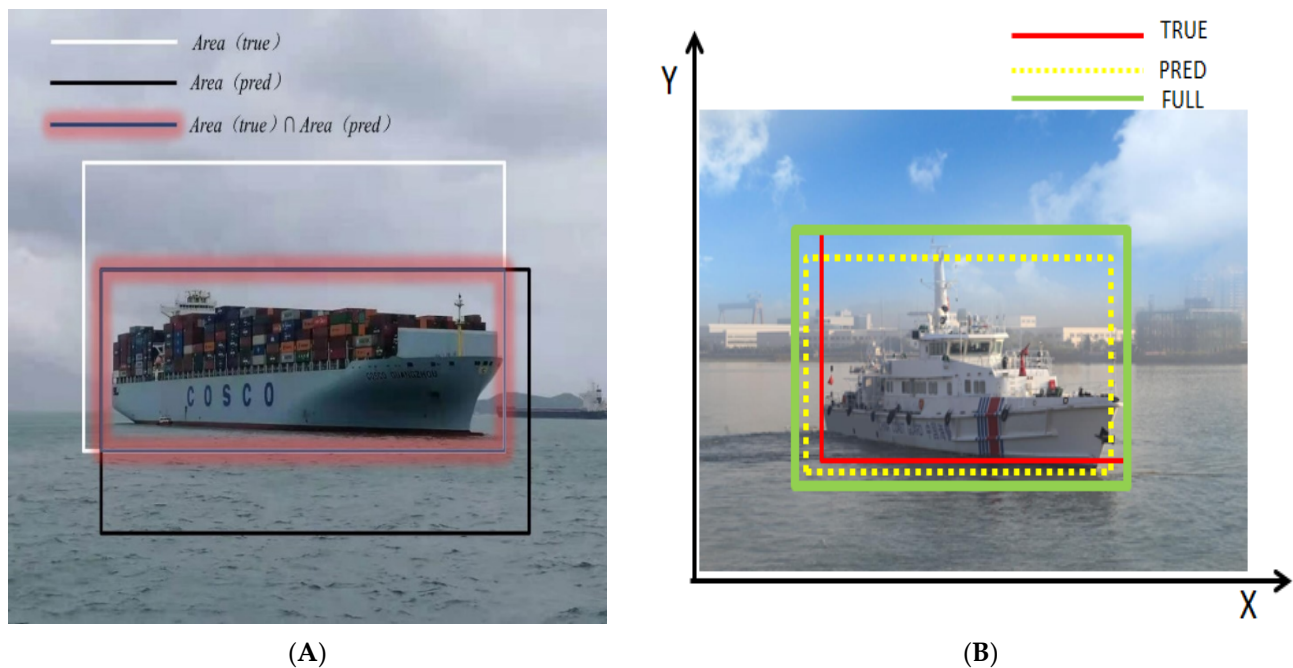


Figure 1. Schematic diagram of (A). *IOU* and (B). *GIoU*.

To calculate *GIoU*, it is necessary to find the smallest box that can fully cover the Prediction box ($Area(pred)$) and the Ground Truth box ($Area(true)$), named $Area(full)$. The schematic diagram is indicated in Figure 1.

$$GIoU_{pred}^{true} = IOU_{pred}^{true} - \frac{Area(full) \setminus Area(true) \cup Area(pred)}{Area(full)} \quad (4)$$

- The second step is feature extraction and prediction. Target prediction is performed in the final layer of the fully connected. If the target exists, the Cell gives the $Pr(class|object)$, and the probability of each class in the whole network is calculated, then the detection $Score_{confidence}$ is calculated. The comprehensive calculation is as

$$\begin{aligned} Conf_{object_i} &= Pr(Class|object) \times Score_{Confidence} \\ &= Pr(Class|object) \times Pr(object) \times GIoU_{pred}^{true} \\ &= Pr(Class_i) \times GIoU_{pred}^{true} \end{aligned} \quad (5)$$

- Setting the detection limitation of $Score_{confidence}$, adjusting and filtering the borders with scores lower than the default value. The remaining borders are the correct detection boxes and the final judgment results are outputted sequentially.

3.2. Multi-Scale Feature Fusion

Scholars strove to find better feature fusion methods for the greater robustness of information. The initial development of the target detector was used to obtain the whole logical information of the object by a single layer for making prediction judgments. For example, the last layer's output was adopted for subsequent processing in a series of R-CNN.

A typical representative application of multiscale feature fusion is FPN [40]. The multi-scale information obtained from feature fusion and improved the network performance for different scale targets (including tiny targets).

YOLOF involves two key modules of a projector and residual blocks. In the projector, 1×1 convolution is applied to reduce the number of parameters, and then 3×3 convolution

is done to extract contextual semantic information (similar to FPN). Residual blocks are four residual modules with different rates of dilation stacked to generate output features with multiple fields of perception, in Figure 2. For residual blocks, all convolution layers are followed by a BatchNorm layer [41] and a ReLU layer [42], but just convolution layers and BatchNorm layers are used in Projector. To accept varying target sizes, 4 consecutive residual units are employed to allow the integration of numerous features with different perceptual fields in a one-level feature.

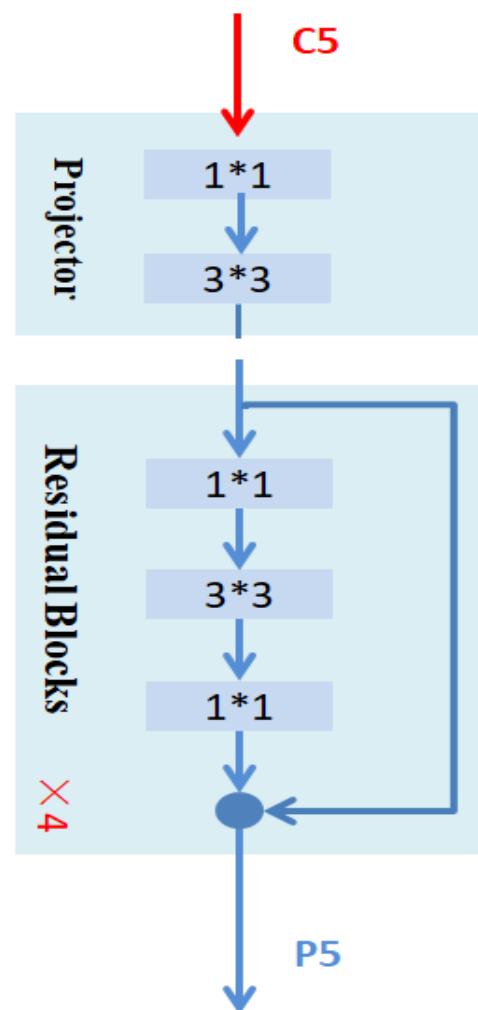


Figure 2. Encoder of YOLOF.

An encoder called Single Feature Map Fusion (SFMF) is presented here because it has been developed as the key component of the detector, distinguished from a feature pyramid based on multiple maps. It was obtained from the optimizations of YOLOF, to design the featured fusion components upon a single feature layer [38]. By the residual module, the YOLOF encoder obtains semantic information on multiple scales.

In Figure 3, L1–L5 is generated on the backbone paths with feature maps containing different scale information, path-1 integrates the results of L1–L4 and L5. The results of path-1 produce the final outcome P5. Remarkably, it ignores the preprocessing. In practice, the use of ReLU in a backbone network may result in the loss of information about the destination. This study tries to employ Meta-ACON [43] (refer to Section 3.3), which is employed in the backbone network to learn to activate or inactivate automatically.

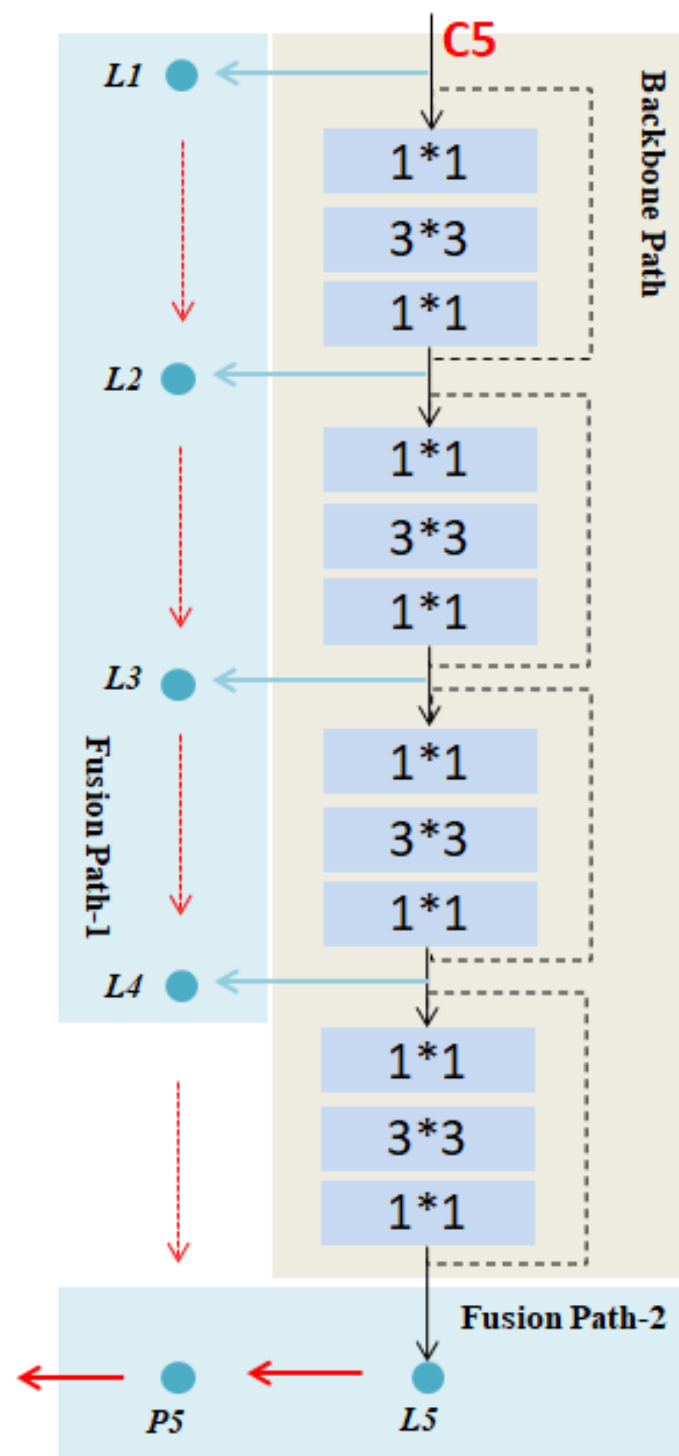


Figure 3. SFMF Architecture.

Preliminary validation of the fusion path method is tested in the 2007-COCO dataset, and the results are shown in Table 1.

Table 1. Comparison of Fusion Paths.

Fusion Path	AP
SMSF	38.5
PANet [44]	37.8
BiFPN [1,45]	37.5

Consideration of additional channels or whether to use shortcuts results are shown in Table 2.

Table 2. Additional case results.

	Additional Channels		Shortcuts	
	YES	NO	YES	NO
AP	38.5	37.4	38.5	37.1

The shortcut retains the original information and overwrites all scale targets in YOLOF. The SFMF retains the lower-scale information for subsequent fusion. The results indicate that the SFMF can create better results with shortcuts.

3.3. Activation and Loss Function

The most ordinary nonlinear functions, including Sigmoid and ReLU, are employed to activate the outputs in deep learning. Ma [43] proposed a novel Meta-ACON to learn automatically to activate the output. Likewise, the same activation function uses a smoothed maximum for approximating the extremum. Its smooth and differentiable approximation is 6, which x represents input. It considers the standard maximum function $\max(x_1, \dots, x_n)$ of n values.

$$S_{\beta}(x_1, x_2, \dots, x_n) = \frac{\sum_{i=1}^n x_i e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}} \quad (6)$$

Additionally, the switch factor β is

$$\begin{cases} \beta \rightarrow \infty, S_{\beta} \rightarrow \max \\ \beta \rightarrow 0, S_{\beta} \rightarrow \text{arithmetic mean} \end{cases} \quad (7)$$

The loss functions are categorized into classification and regression loss. The classification [46] is optimized via a focal loss (FL) algorithm at the one-stage detector. The function of the focal loss is to calculate the cross-entropy loss of the predicted outcomes for all non-ignored categories. The loss function serves to evaluate the comparison between the predicted and actual values of the model, where the smaller the loss function, the better the model performance. This work follows the original settings in YOLOF, e.g., FL and *GIOW*.

4. Experiments

4.1. Dataset Composition

Currently, most datasets are designed for land targets. However, maritime target images lack open datasets because maritime targets differ greatly from land targets. In this paper, the typical target objects as maritime ships are divided into five typical types, including passenger ships, container ships, bulk carriers, sailboats, and other ships. It is worth noting that the island can be accurately judged by the model, so the boxes are hidden. The purpose of this operation is to keep the display tidy.

The images in the dataset have been augmented with data to minimize overfitting the model in order to improve detection accuracy. How can I deal with the overfitting issue? The most efficient method is to enhance the data set. The purpose of supplementing the data will be to allow the model to meet more ‘exceptions’, allowing it to constantly correct itself and provide better results. This is usually accomplished by either gathering or enhancing more of the initial data from the source, or by copying the original data and adding random disturbance or faulty data, which accounts for 3% of the total in this study. To improve the model’s generalization and practical application by enriching the dataset, a selection of real-world ship images was obtained from the open-source network to supplement the dataset. Horizontal and vertical flipping, random rotation, random scaling,

random cropping, and random expansion are all common augmentation procedures. It is worth noting that a detailed annotation of the dataset is necessary, although this is a time-consuming and complicated operation. There are 4267 images in total in the dataset, with 20% designated for the test set, and the rest for the training set. In COCO, the batch size is set to 48, the learning rate is set to 0.06, and the maximum number of iterations is set to 8 k. Additionally, use the parameters in YOLOF for supplemental choices, such as FL and *GIOU*. In the own self-built dataset, these parameters are recommended. The batch size is set at 24 and the learning rate is set to 0.03. For debugging purposes, there is personal experience data, batch size set to 8/GPU.

4.2. Establishment of Computer Platform

The experimental platform includes the following components. An Intel(R) Xeon(R) Gold 6130 CPU @ 2.10 GHz, three NVIDIA TAITAN RTX 24 G GPUs, ResNet50 as the basic algorithm framework, Python 3.7.0 as the programming language, Opencv4.5 as the graphics processing tool, and Detectron2 from FACEBOOK as the training framework, as shown in Table 3.

Table 3. Platform configuration information.

Name	Version
CPU	Intel(R) Xeon(R) Gold 6130 CPU @ 2.10 GHz
GPU	NVIDIA TAITAN RTX 24 G
OS	Ubuntu 20.04
python	3.7.0

4.3. Evaluation Indexes

In this paper, the indexes including Frames per second (FPS), mAP, and FLOPs are used to evaluate the overall performance of detection results. The C_{TP} indicates the number of ships classified as true positives. *Precision* is denoted by $precision = \frac{C_{TP}}{all\ detections}$, the *Recall* rate $recall = \frac{C_{TP}}{all\ ground\ truths}$. Typically, the higher recall, the lower accuracy, and vice versa. *AP* combining the different accuracy and recall rates reflects the overall performance of the model as

$$AP = \int_0^1 P(R) dR \quad (8)$$

Mean Average Precision (*mAP*) denotes the average of each *AP* category as

$$mAP = \frac{\sum_{i=1}^n AP_i}{n} \quad (9)$$

Additionally, the floating-point operations per second (FLOPs), the number of floating-point operations performed per second, are used to measure the computing power of a computer.

4.4. Results Analysis

The results of target recognition by the model are shown in Figure 4. A target detector on a ship with good performance can provide maritime authorities with an objective reference for data visualization and reduce the ship's collision risk due to human negligence.

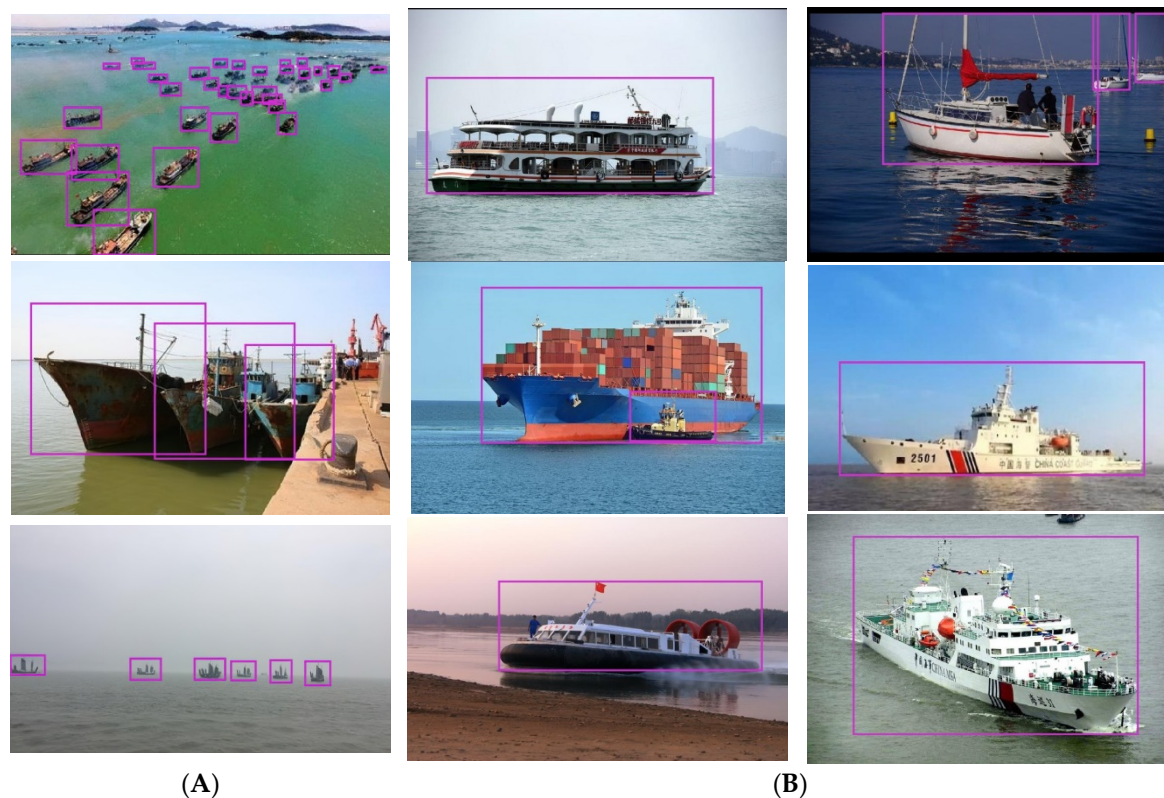


Figure 4. Recognition results of FMTI: (A) multi-objective, (B) simple or single objective.

For the first image of Figure 4A, the far ship targets were not labeled in detail at the beginning of the experiment, which was subject to an ‘accident’ of erroneous recognition. The recognition effect by the FMTI algorithm was so accurate that it surpassed the labeling, i.e., the number of targets identified successfully was more than the number of the ones labeled manually. Similarly, the hull pieces in the second image are partially overlapping but can still be distinguished. In the third photo, the ships are separated, and this allowed for the best recognition.

The FMTC algorithm has a good performance of recognition for not only the multitarget tasks but also the simple or single target(s) tasks, like in Figure 4B.

In particular, ResNet-101 [47] was introduced as a backbone network for cross-sectional comparison of models, denoted by Res101. Table 4’s data is rounded; however, this does not affect the overall assessment. Unavailable or useless data is indicated by /.

Table 4. Preliminary comparison results of different combinations.

Name	FMTI	YOLOF	YOLOF + SFMF	YOLOF (Res101)	YOLOF (Res101) + SFMF
FPS	39	40	38	25	29
FLOPs	91 G	86 G	91 G	/	/
Model_parameter	48 M	44 M	49 M	64 M	68 M
mAP	>37.7%	37.7%	37%	36%	37%

Table 4 was obtained in the 2017 COCO validation set, and Table 5 was acquired from self-built datasets. The data is generated on an identically equipped device. Following a comprehensive analysis of Table 4, the FMTI and YOLOF models were chosen to be applied to the self-built dataset. The results are given in Table 5, $Score_{confidence} = 0.5$.

Table 5. FMTI and YOLOF comparison results.

Name	FMTI	YOLOF
FPS	36.66	37.53
FLOPs	98.016 G	93.52 G
mAP	0.47529	0.40382
Model_parameter	47.137 M	42.488 M

As shown in Table 4, we acquired 37 percent mAP (YOLOF + SFMF) and 36 percent mAP (YOLOF (Res101)), respectively. FMTI achieves more than 0.7 percent mAP improvement (Baseline: YOLOF + SFMF or YOLOF (Res101) + SFMF) and better than YOLOF (Res101) over 1.7 percent mAP, respectively. Furthermore, YOLOF received 37 percent mAP, an increase of one percent above the YOLOF (Res101) mAP. In terms of mAP, FMTI exceeds YOLOF and the other models, although it has a slightly lower FPS than YOLOF, which does not affect the processing performance of the FMTI model.

The results are clearly shown in Table 5. Additionally, it is worth highlighting that the improvement of mAP is over 7%, which is significant. The computational power has been advanced in parallel with mAP. The model is frequently improved along with memory changes. The accompanying increase in model parameters is so normal that it is within acceptable ranges.

More particularly, when the FMTI algorithm is applied for maritime monitoring to provide early warning for potential danger signals in offshore areas proactively, the occurrence probability of maritime accidents must be reduced. The FMTI model proposed in this paper applies to maritime target detection, meanwhile, it also can broad application prospects in the fields of maritime rescue, maritime traffic monitoring, and maritime battlefield situational awareness and assessment.

5. Conclusions

This paper addressed an encoder, known as SFMF, which enables multi-scale feature fusion on a map. A cross-sectional assessment of the different component compositions was conducted prior to the experimental application of model choices, then the YOLOF model was selected for comparison with the FMTI model. Although the FMTI model had a slightly lower speed evaluation metric of FPS than the YOLOF model, it had more computational power in the COCO dataset, so the two models mentioned above were chosen for the next experimental comparison. Combining speed and processing power, the FMTI algorithm was able to outperform the previous YOLOF in the marine ship detection data, so it has the potential for future applications.

The FMTI algorithm could offer technical support in the areas of smart coastal transit, naval defense, and smart maritime construction. It could be employed on the video surveillance equipment to detect offshore ships, detecting the ships entering and departing ports, the field of illegal fighting or military defense by recognizing and pre-warning dangerous boats along the shoreline.

However, the image data for the majority of the training data set are captured during good weather conditions in this paper, so further studies may still be done to ensure better performance of the model. Future work will focus on the diversity of test samples.

Author Contributions: Conceptualization, J.W. and X.X.; methodology, J.W., D.G. and X.X. soft-ware, X.X.; validation, X.X. and D.G.; investigation, J.W.; resources, J.W.; writing—original draft preparation, J.W.; writing—review and editing, J.W., J.L. and R.L.; visualization, X.X.; supervision, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (52171346), the Natural Science Foundation of Guangdong Province (2021A1515012618), the special projects of key fields (Artificial Intelligence) of Universities in Guangdong Province (2019KZDZX1035), and the program for scientific research start-up funds of Guangdong Ocean University (R19055).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Li, R.; Wu, J.; Cao, L. Ship target detection of unmanned surface vehicle base on efficientdet. *Syst. Sci. Control. Eng.* **2022**, *10*, 264–271. [\[CrossRef\]](#)
- Chen, X.; Liu, Y.; Achuthan, K.; Zhang, X. A ship movement classification based on Automatic Identification System (AIS) data using Convolutional Neural Network. *Ocean. Eng.* **2020**, *218*, 108182. [\[CrossRef\]](#)
- Shen, S.; Yang, H.; Yao, X.; Li, J.; Xu, G.; Sheng, M. Ship Type Classification by Convolutional Neural Networks with Auditory-like Mechanisms. *Sensors* **2020**, *20*, 253. [\[CrossRef\]](#) [\[PubMed\]](#)
- Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In *Computer Vision—ECCV 2016*; Springer: Cham, Switzerland, 2016; Volume 9905, pp. 21–37. [\[CrossRef\]](#)
- Girshick, R. Fast R-CNN. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 1440–1448. [\[CrossRef\]](#)
- Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788. [\[CrossRef\]](#)
- Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [\[CrossRef\]](#)
- Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [\[CrossRef\]](#)
- Purkait, P.; Zhao, C.; Zach, C. SPP-Net: Deep Absolute Pose Regression with Synthetic Views. *arXiv* **2017**, arXiv:1712.03452.
- Redmon, J.; Farhadi, A. YOLO9000: Better, Faster, Stronger. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6517–6525. [\[CrossRef\]](#)
- Redmon, J.; Farhadi, A. YOLOv3: An Incremental Improvement. *arXiv* **2018**, arXiv:1804.02767.
- Taimuri, G.; Matusiak, J.; Mikkola, T.; Kujala, P.; Hirdaris, S. A 6-DoF maneuvering model for the rapid estimation of hydrodynamic actions in deep and shallow waters. *Ocean. Eng.* **2020**, *218*, 108103. [\[CrossRef\]](#)
- Rezaadegan, F.; Shojaei, K.; Sheikholeslam, F.; Chatraei, A. A novel approach to 6-DOF adaptive trajectory tracking control of an AUV in the presence of parameter uncertainties. *Ocean. Eng.* **2015**, *107*, 246–258. [\[CrossRef\]](#)
- Tello, M.; Lopez-Martinez, C.; Mallorqui, J.J. A Novel Algorithm for Ship Detection in SAR Imagery Based on the Wavelet Transform. *IEEE Geosci. Remote Sens. Lett.* **2005**, *2*, 201–205. [\[CrossRef\]](#)
- Chang, Y.-L.; Anagaw, A.; Chang, L.; Wang, Y.; Hsiao, C.-Y.; Lee, W.-H. Ship Detection Based on YOLOv2 for SAR Imagery. *Remote Sens.* **2019**, *11*, 786. [\[CrossRef\]](#)
- Chen, Z.; Chen, D.; Zhang, Y.; Cheng, X.; Zhang, M.; Wu, C. Deep learning for autonomous ship-oriented small ship detection. *Saf. Sci.* **2020**, *130*, 104812. [\[CrossRef\]](#)
- Arcos-García, Á.; Álvarez-García, J.A.; Soria-Morillo, L.M. Evaluation of deep neural networks for traffic sign detection systems. *Neurocomputing* **2018**, *316*, 332–344. [\[CrossRef\]](#)
- Lopac, N.; Hrzić, F.; Vuksanović, I.P.; Lerga, J. Detection of Non-Stationary GW Signals in High Noise From Cohen’s Class of Time-Frequency Representations Using Deep Learning. *IEEE Access* **2022**, *10*, 2408–2428. [\[CrossRef\]](#)
- Javed Awan, M.; Mohd Rahim, M.S.; Salim, N.; Mohammed, M.A.; Garcia-Zapirain, B.; Abdulkareem, K.H. Efficient Detection of Knee Anterior Cruciate Ligament from Magnetic Resonance Imaging Using Deep Learning Approach. *Diagnostics* **2021**, *11*, 105. [\[CrossRef\]](#) [\[PubMed\]](#)
- Qiblawey, Y.; Tahir, A.; Chowdhury, M.E.H.; Khandakar, A.; Kiranyaz, S.; Rahman, T.; Ibtehaz, N.; Mahmud, S.; Maadeed, S.A.; Musharavati, F.; et al. Detection and Severity Classification of COVID-19 in CT Images Using Deep Learning. *Diagnostics* **2021**, *11*, 893. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, X.; Yang, Y.; Li, Z.; Ning, X.; Qin, Y.; Cai, W. An Improved Encoder-Decoder Network Based on Strip Pool Method Applied to Segmentation of Farmland Vacancy Field. *Entropy* **2021**, *23*, 435. [\[CrossRef\]](#) [\[PubMed\]](#)
- Zhang, T.; Hao, L.-Y.; Guo, G. A feature enriching object detection framework with weak segmentation loss. *Neurocomputing* **2019**, *335*, 72–80. [\[CrossRef\]](#)
- Li, T.; Li, R.; Li, J. Decentralized adaptive neural control of nonlinear interconnected large-scale systems with unknown time delays and input saturation. *Neurocomputing* **2011**, *74*, 2277–2283. [\[CrossRef\]](#)
- Li, T.; Li, R.; Wang, D. Adaptive neural control of nonlinear MIMO systems with unknown time delays. *Neurocomputing* **2012**, *78*, 83–88. [\[CrossRef\]](#)

25. Gu, X.; Liu, C.; Wang, S.; Zhao, C. Feature extraction using adaptive slow feature discriminant analysis. *Neurocomputing* **2015**, *154*, 139–148. [[CrossRef](#)]
26. Guo, G.; Wang, H.; Yan, Y.; Zheng, J.; Li, B. A fast face detection method via convolutional neural network. *Neurocomputing* **2020**, *395*, 128–137. [[CrossRef](#)]
27. Liu, C.; Li, X.; Li, Q.; Xue, Y.; Liu, H.; Gao, Y. Robot recognizing humans intention and interacting with humans based on a multi-task model combining ST-GCN-LSTM model and YOLO model. *Neurocomputing* **2021**, *430*, 174–184. [[CrossRef](#)]
28. Zheng, Y.; Bao, H.; Meng, C.; Ma, N. A method of traffic police detection based on attention mechanism in natural scene. *Neurocomputing* **2021**, *458*, 592–601. [[CrossRef](#)]
29. Yu, X.; Wu, S.; Lu, X.; Gao, G. Adaptive multiscale feature for object detection. *Neurocomputing* **2021**, *449*, 146–158. [[CrossRef](#)]
30. Zhang, C.; Pan, X.; Li, H.; Gardiner, A.; Sargent, I.; Hare, J.; Atkinson, P.M. A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS J. Photogramm. Remote Sens.* **2018**, *140*, 133–144. [[CrossRef](#)]
31. Sharifzadeh, F.; Akbarizadeh, G.; Seifi Kaviani, Y. Ship Classification in SAR Images Using a New Hybrid CNN–MLP Classifier. *J. Indian Soc. Remote Sens.* **2019**, *47*, 551–562. [[CrossRef](#)]
32. Wu, F.; Wang, C.; Jiang, S.; Zhang, H.; Zhang, B. Classification of Vessels in Single-Pol COSMO-SkyMed Images Based on Statistical and Structural Features. *Remote Sens.* **2015**, *7*, 5511–5533. [[CrossRef](#)]
33. Tao, D.; Anfinsen, S.N.; Brekke, C. Robust CFAR Detector Based on Truncated Statistics in Multiple-Target Situations. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 117–134. [[CrossRef](#)]
34. Tao, D.; Doulgeris, A.P.; Brekke, C. A Segmentation-Based CFAR Detection Algorithm Using Truncated Statistics. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 2887–2898. [[CrossRef](#)]
35. Srinivas, U.; Monga, V.; Raj, R.G. SAR Automatic Target Recognition Using Discriminative Graphical Models. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 591–606. [[CrossRef](#)]
36. Ma, Y.; Zhao, Y.; Incecik, A.; Yan, X.; Wang, Y.; Li, Z. A collision avoidance approach via negotiation protocol for a swarm of USVs. *Ocean. Eng.* **2021**, *224*, 108713. [[CrossRef](#)]
37. Ma, Y.; Nie, Z.; Hu, S.; Li, Z.; Malekian, R.; Sotelo, M. Fault Detection Filter and Controller Co-Design for Unmanned Surface Vehicles under DoS Attacks. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 1422–1434. [[CrossRef](#)]
38. Chen, Q.; Wang, Y.; Yang, T.; Zhang, X.; Cheng, J.; Sun, J. You Only Look One-level Feature. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13034–13043. [[CrossRef](#)]
39. Rezaatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 658–666. [[CrossRef](#)]
40. Lin, T.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature Pyramid Networks for Object Detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 936–944. [[CrossRef](#)]
41. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv* **2015**, arXiv:1502.03167.
42. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning, Haifa, Israel, 21–24 June 2010; pp. 807–814.
43. Ma, N.; Zhang, X.; Liu, M.; Sun, J. Activate or Not: Learning Customized Activation. In Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 8028–8038. [[CrossRef](#)]
44. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path Aggregation Network for Instance Segmentation. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768. [[CrossRef](#)]
45. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and Efficient Object Detection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 10778–10787. [[CrossRef](#)]
46. Lin, T.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)] [[PubMed](#)]
47. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *arXiv* **2015**, arXiv:1512.0338.