applied sciences

MDPI

Article

ECGYOLO: Mask Detection Algorithm

Wenyi Hu [1], Jinling Zou [1], Yuan Huang [1], Hongkun Wang [1], Kun Zhao [1], Mingzhe Liu [1,*] and Shan Liu [2,*]

[1] Department of Computer and Network Security, Chengdu University of Technology, Chengdu 610059, China
[2] School of Automation, University of Electronic Science and Technology of China, Chengdu 610054, China
* Correspondence: liumz@cdut.edu.cn (M.L.); shanliu@uestc.edu.cn (S.L.)

Abstract: Of past years, wearing masks has turned into a necessity in daily life due to the rampant new coronavirus and the increasing importance people place on health and life safety. However, current mask detection algorithms are difficult to run on low-computing-power hardware platforms and have low accuracy. To resolve this discrepancy, a lightweight mask inspection algorithm ECGYOLO based on improved YOLOv7tiny is proposed. This algorithm uses GhostNet to replace the original convolutional layer with ECG module instead of ELAN module, which greatly improves the inspection efficiency and decreases the parameters of the model. In the meantime, the ECA (efficient channel attention) mechanism is led into the neck section to boost the feature fetch capability of the channel, and Mosaic and Mixup data enhancement techniques are adopted in training to obtain mask images under different viewpoints to improve the comprehensiveness and effectiveness of the model. Experiments show that the mAP (mean average precision) of the algorithm is raised by 4.4% to 92.75%, and the number of arguments is decreased by 1.14 M to 5.06M compared with the original YOLOv7tiny. ECGYOLO is more efficient than other algorithms at present and can meet the real-time and lightweight needs of mask detection.

Keywords: ECG; ECA; mask detection; YOLOv7

Citation: Hu, W.; Zou, J.; Huang, Y.; Wang, H.; Zhao, K.; Liu, M.; Liu, S. ECGYOLO: Mask Detection Algorithm. *Appl. Sci.* **2023**, *13*, 7501. https://doi.org/10.3390/app13137501

Academic Editor: Habib Hamam

Received: 31 May 2023
Revised: 19 June 2023
Accepted: 21 June 2023
Published: 25 June 2023

1. Introduction

Globally, coronaviruses are rampant [1], spreading and endangering all countries at a rate rare in the history of human medicine. Since the first outbreak in Wuhan, Hubei Province, China, in late 2019, the virus has rapidly spread around the world, becoming a global public health emergency. The pandemic has not only caused massive casualties and medical stress but has also had a profound impact on global economic, political and social life. During this difficult time, governments, social organizations, medical institutions and ordinary people around the world acted together to call for world solidarity against the epidemic and to maintain generosity and self-discipline. However, many people did not keep their distance or wear masks in crowded public places (such as stations, bars, parks), causing the global epidemic to become more serious. To address the inefficiency of preventive measures taken by government agencies, this study aims to create a deep-learning-based mask detection system using an improved ECGYOLO based on YOLOv7tiny to ensure that all people wear masks in these places, thereby reducing the risk of COVID-19 virus transmission.

At present, with the rapid increase in computing power, deep-learning-based target detection algorithms [2] are also gaining more and more attention and are widely used. Common application scenarios include small target detection [3], hidden object detection [4] and optical remote-sensing images [5], among other fields. For example, a small target detection method is proposed for feature extraction [6], a method for FPRNet is proposed for remote-sensing target detection [7], and an adaptive balanced network is proposed for remote-sensing image detection in the literature [8]. These methods include some of the

latest research results in the domain of object detection. At the moment, target detection algorithms are parted into two main kinds: two-step algorithms and one-step algorithms. Two-step algorithms are represented by FasterRCNN [9], which introduces the concept of candidate region proposals to generate a set of candidate ranges at the first step and afterwards adopts a classifier to further filter them. The one-step algorithm is represented by YOLO [10] and SSD, which directly performs dense sampling on the image and only needs to be fed into the network once to predict all the bounding boxes, so the speed is faster [11].

Although present target detection algorithms have relatively good detection speed and accuracy, there are still some shortcomings, the main problem being that for average hardware platforms, most current detection algorithms are not yet able to meet the real-time and accuracy requirements needed for mask detection. Therefore, in order to further improve the real-time and accuracy of mask detection, this paper proposes the ECGYOLO model based on YOLOv7tiny. Experiments show that ECGYOLO outperforms YOLOv7tiny in terms of accuracy and inference speed.

The contribution of this article is in three major areas:

1. This paper proposes a lightweight mask detection model ECGYOLO, further improving the ELAN module based on YOLOv7tiny, using the ECG module and adding the ECA model, replacing the normal convolution with GhostConv and reintroducing RepConv. All these upgrades can effectively improve the model in the mask-wearing detection task performance.
2. Throughout this paper, the authors evaluate and compare the performance of commonly used target detection models including YOLOv7, YOLOv7tiny, FasterRCNN and SSD with the proposed ECGYOLO model in the mask-wearing detection task. The evaluation results show that ECGYOLO achieves 92.7% in the mAP metric, which is 4.4% better than that of YOLOv7tiny and even higher than that of other models such as YOLOv7, FasterRCNN and SSD. As a result, ECGYOLO has better performance and efficiency in mask detection tasks.
3. Another contribution of this paper is in decreasing the number of model parameters of the ECGYOLO model to 5.06 M, which is 1.14 M lower than YOLOv7tiny and much smaller than that of other evaluation models. This will make the ECGYOLO model more suitable for deploying and promoting its use on devices with limited computational resources.

All this being said, the ECGYOLO model proposed throughout this paper achieves high performance in the mask detection scenario and has the advantages of small number of parameters and fast computational speed, which will make the model more practical and feasible in practical applications.

The sections of this paper are organised as follows: the Section 2 presents some related work, the Section 3 describes the improved model, the Section 4 describes the dataset and the operating environment, the Section 5 analyses the results of the experiments and the comparison of the models, and the Sections 6 and 7 provide some discussion and conclusions.

2. Related Works

YOLO is currently the most powerful open source target detection model and can be found in various fields, for example, one study reported in the literature applied YOLO to citrus orchards, which can save a lot of manpower and resources [12], another study also described in the literature used YOLO to detect whether drivers are distracted [13], and another study in the literature applied YOLO to ship detection [14]. The best of the YOLO series right now is YOLOv7, which was formerly known as YOLOv1 and has undergone several improvements in YOLOv2 [15], YOLOv3 [16], YOLOv4 [17] and YOLOv5 [18]. YOLOv7 introduces the residual module Darknet-53 [19] and the FPN [20] structure to achieve multi-scale fusion and prediction of objects at three different scales. In addition, YOLOv7 extends the original ELAN structure and proposes an Extended ELAN framework,

which can increase the self-study capability of the circuit without damaging the primary gradient path. On the whole, YOLOv7 has great advantages in terms of parameters, calculation and accuracy and is a very advanced target recognition model.

A study described in the literature uses a global attention mechanism to reduce the loss of feature information to some extent and Soft-NMS to improve the accuracy of the prediction frame [21]. The authors of another study reported in the literature use the Mish activation function to replace the LeakyReLU activation function, a dense SPP layer in feature extraction, and offer their own understanding of the detection of small targets. However, the F1 index in mask detection is poor, only 78%, and there is a large number of missed and wrong detections [22]. The authors of another study presented in the literature proposed to introduce CSPDarkNet53 into YOLOv4 with Hardswish activation function to achieve relatively high-accuracy mask recognition, but the computational cost of the model is high, and the model itself is more strenuous on low-computing-power hardware devices [23]. A YOLO mask detection framework using an improved Res2Net module is proposed in the literature, but the improved model is more complex and not as good as YOLOv7 in terms of performance [24].

However, it can be seen from the above literature that there is a lack of a mask detection system suitable for a low-computing-power platform with high accuracy and fast operation. Therefore, this paper proposes a dataset based on WIDER Face and MAFA and uses the lightweight YOLOv7tiny model, combined with ECA attention mechanism, GhostConv convolution module, ECG module and EIoU loss function for improvement. Experiments show that these mends are profitable to improve the preciseness and rate of the model and realize the lightweighting of the model. The trial run on the dataset and suggested in this article shows that the arithmetic can effectually detect the face of the mask wearer and compare other existing object detection arithmetic on the market, so the accuracy and speed are improved, and the model is lightweight.

3. YOLOv7tiny Model Improvements

3.1. ECGYOLO

The input layer of ECGYOLO model adopts various technical means, which include adaptive image adjustment and Mosaic high-order data enhancement, etc. Among them, the Mosaic technique is mainly used to process small target detection, which can effectively increase the robustness and accuracy. In addition, ECGYOLO's backbone network uses the ECG model and downsampling model, activation function uses Hardswish, and convolution layer uses GhostConv. In the Neck layer of ECGYOLO, SPPCSPC, ELAN structure and downsampling model are mainly used, which can obtain the spatial information and context information of the input features more effectively, thus improving the detection efficiency of the model for small targets. Similarly, the attention mechanism uses the ECA model, which can further improve the perceptual field and the capacity to concentrate on significant features across channels of the model. In the head layer, ECGYOLO mainly adopts the reparameterized structure RepConv to solve the problems in detecting small targets. In addition, the loss function adopts EIoU, which can enhance the effect of the localization accuracy and robustness of the model to the target edges. In summary, ECGYOLO adopts a series of advanced technical means, including Adaptive image adjustment, Mosaic high-order data enhancement, ECG model, GhostConv convolutional layer, ELAN structure, ECA attention mechanism, RepConv reparameterization structure and EIoU loss function, which make the model have high accuracy and robustness in the target detection field.

As can be seen from Figure 1, the backbone network part of ECGYOLO has nine modules: two CBS modules, four ECG modules and three MP modules. The CBS modules mainly include convolution, normalisation and activation functions; the ECG module is a modified module based on the YOLOv7tniy's ELAN structure; and the MP module is Maxpooling, which is mainly used for downsampling. In the neck FPN section, there are four CBS modules, one SPPCSPC module, four MCB modules and two UpSampling

modules. The SPPCSPC module is a special SPP (spatial pyramidal pooling) layer that introduces a CSP structure into the SPP structure; the MCB module is the ELAN structure of YOLOv7tiny. Other parts include the expansion of three RepConv modules and three YOLOhead modules.

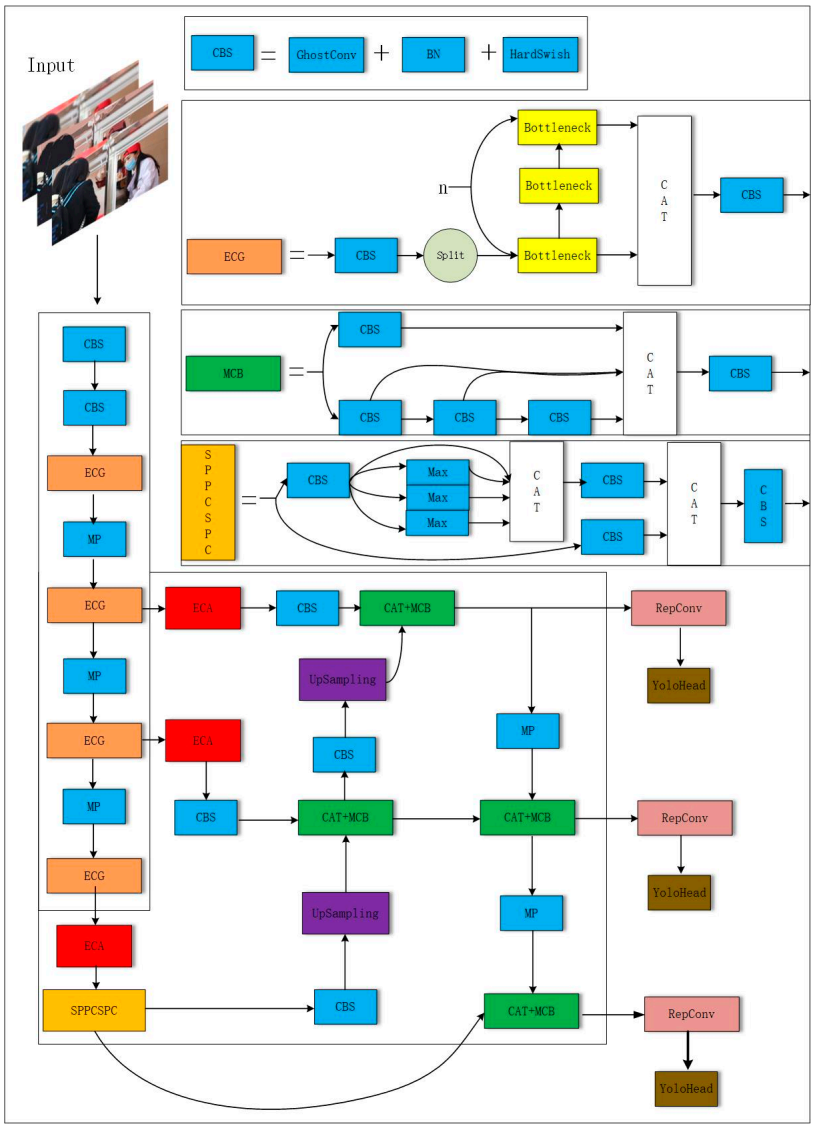

Figure 1. ECGYOLO model.

3.2. CBS Module Improvements

CBS is the convolution, normalisation (batch normalization) and activation function. In YOLOv7, CBS is the normal convolution, normalization and SiLu activation function. In YOLOv7tiny, CBS represents the normal convolution, normalization and LeakyReLU activation function. In ECGYOLO, CBS represents GhostConv, normalization and Hardswish activation function.

Compared to YOLOv7tiny, the ECGYOLO's CBS module uses the cheaper GhostConv to replace the normal convolution and uses some less computationally intensive operations to generate these redundant feature maps, which greatly reduces the number of model parameters and increases the execution speed of the model. In terms of activation functions, the Hardswish activation function can be implemented as a segmentation function to reduce the number of memory accesses compared to the LeakyReLU activation function, thus significantly reducing the waiting time cost. Therefore, the use of GhostConv instead

of normal convolution and Hardswish instead of LeakyReLU activation functions in the CBS module is a better option.

3.2.1. GhostConv

Achieving high accuracy and light weight on platforms with poor hardware configuration still has many problems. Although lightweight network models such as ShuffleNet [25] and MobileNet [26] have emerged, GhostNet has become a better alternative to traditional convolution with its unique convolution module. Compared with the traditional convolution, GhostNet uses a more efficient Ghost module, by dividing the convolution kernel of the original convolution into two parts, and the application of a few lower computations for manipulations to produce these redundant feature maps, thereby reducing the number of parameters and increasing the implementation rate of the model. Moreover, GhostNet also introduces the SE module, which can go a step further to enhance the precision of the model. These innovations enable GhostNet to maintain high precision while keeping the parameter within a reasonable range, making it suitable for deployment in resource-constrained scenarios such as mobile devices.

GhostNet is a lightweight CNN model that uses a new type of module called Ghost to decrease the parameter of the model. GhostNet decomposes each standard convolutional layer into two parts, one of which extracts features from the trunk convolution kernel, while the other smaller subconvolution kernel. The Ghost convolution kernel is used to go a step further to optimize the function extraction process [27]. In this way, GhostNet can obtain smaller model size and lower computational cost while maintaining model accuracy. GhostModule chiefly falls into the following three sections:

Foremost, the authors obtain the intrinsic feature maps $Y_{\omega' * h' * m'}$ with regular convolution; w' and h' are the width and height of the output data, and m represents m maps.

$$Y' = X \times f', \tag{1}$$

Afterwards, the feature map y_i' of apiece channel of Y' is used to generate y_{ij} of Ghost feature map by $\Phi_{i,j}$ operation. y_{ij} as shown in Equation (2).

$$y_{ij} = \Phi_{i,j}(y_i'), \tag{2}$$

Finally, the received intrinsic feature maps and Ghost feature map y_{ij} are spliced together to achieve the ultimate result OutPut. The Ghost model is as displayed in Figure 2.

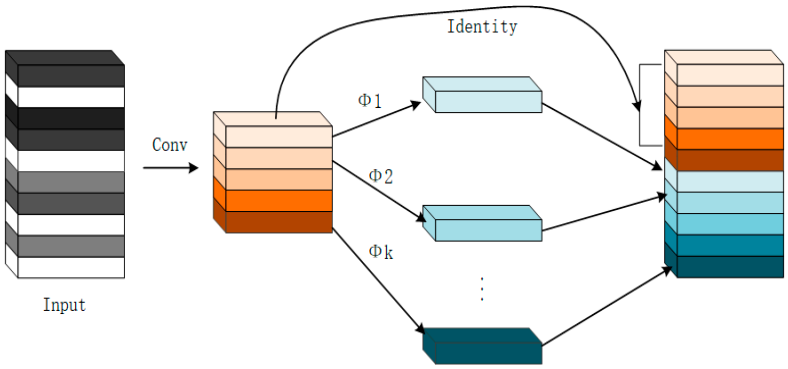

Figure 2. Ghost Model.

3.2.2. Hardswish

Hardswish [28] is an activation function that is an improvement of the Swish activation function. The Swish activation function has achieved good results in deep learning, but its computational complexity is high, so Hardswish was proposed to reduce the computational cost. Unlike Swish, Hardswish uses a segmented linear function instead of a sigmoid function. Specifically, Hardswish is equivalent to ReLU [29] for input values less than -3

or greater than 3, while smoothing is performed in a sigmoid-like form for input values between −3 and 3. Compared to Swish and other familiar activation functions, Hardswish has lower computational cost and can improve the computational efficiency of the model while maintaining similar performance; thus, using Hardswish to replace YOLOv7tiny's LeakyReLU in resource-constrained environments such as mobile devices can effectively reduce inference time and power consumption. In addition, Hardswish has a number of other advantages. For example, it is monotonically differentiable and has no negative outputs, which makes training more stable and reliable. Moreover, using Hardswish on feature maps does not lead to information bottlenecks (bottleneck) because its output range is the same as that of ReLU. In conclusion, Hardswish is a lightweight, efficient and easy-to-implement activation function. In prospect, it can reduce computational costs while improving model accuracy and has the advantages of monotone differentiability and no negative output, making it a good choice of activation function. The Hardswish function has been implemented in many deep learning frameworks.

In conclusion, Hardswish and LeakyReLU are essentially two different activation functions, although they have some similarities in some aspects. Hardswish is suitable for scenarios requiring high computational efficiency, while LeakyReLU can effectively alleviate problems such as neuron death problem and gradient disappearance. The Hardswish formula is shown in Equation (3).

$$\text{Hardswish(a)} = \begin{cases} 0 & a \leq -3, \\ a & a \geq +3, \\ \frac{a^2}{6} + \frac{a}{2} & \text{otherwise} \end{cases}, \tag{3}$$

3.3. MCB Module Improvements

This study uses ELAN idea to redesign the C3 module in YOLOv5 and uses GhostConv to replace the ordinary convolution in C3 to obtain a new target detection model ECG. Compared to the ELAN structure of YOLOv7, YOLOv7tiny is lighter with fewer branches and convolutions. Experimental results show that while ensuring light weight, ECGYOLO has richer feature expression capability and higher detection accuracy compared with YOLOv7tiny. The ECG model is shown in Figure 3. The ELAN structure and the ELAN structure of YOLOv7tiny are shown in Figures 4 and 5.

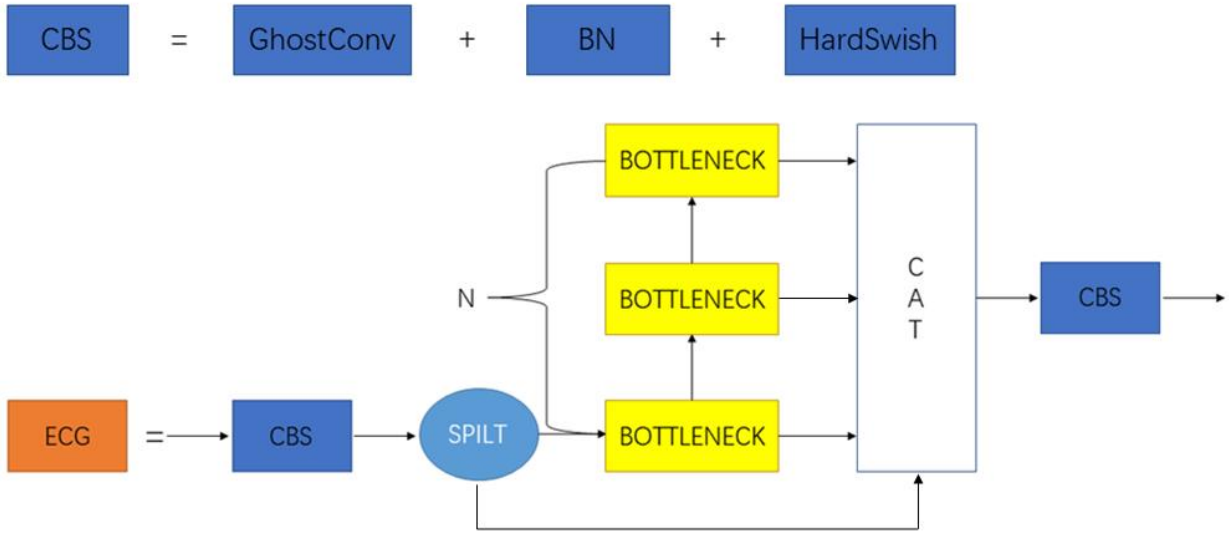

Figure 3. ECG.

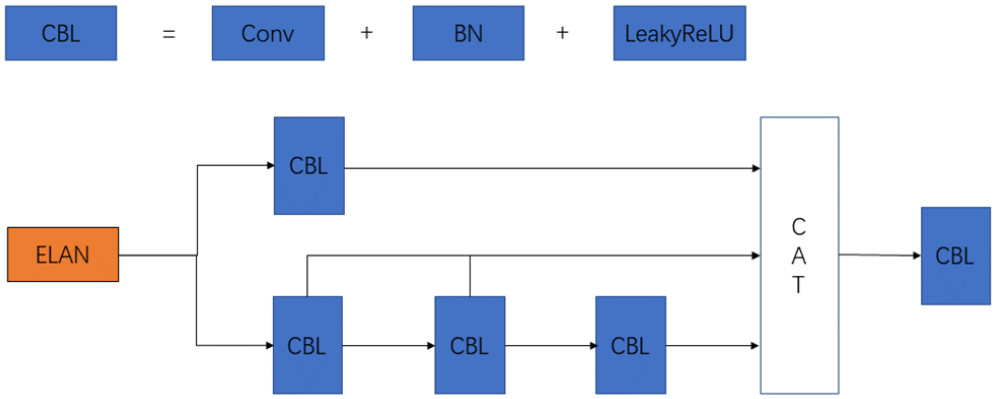

Figure 4. YOLOv7tiny ELAN.

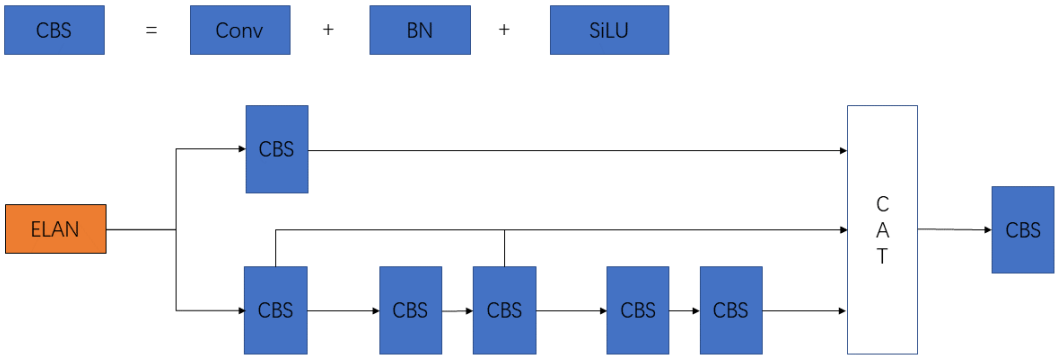

Figure 5. ELAN.

3.4. Bounding Box Loss Function Improvement

The IoU (intersection over union) is used to measure the extent of overlay between the real frame and the prediction frame in the target detection task [30]. However, IoU has a fatal flaw that the backpropagation gradient vanishes when the bounding box A and bounding box B do not overlap, so many IoU-based GIoU, DIoU, CIoU, SIoU, WIoU and EIoU appear. The loss function used in YOLOv7tiny is CIoU, while EIoU was chosen in ECG, a more important reason being that EIoU allows for better regression of the bounding box.

The GIoU (generalized IoU) [31] is a loss calculation method for bounding box prediction, which originated and extended from the IoU metric. In target detection tasks, the comparison between the predicted and actual labelled bounding boxes and the corresponding loss value calculation are crucial. Compared with IoU, GIoU considers the non-overlapping region of the bounding box and can exactly mirror the way of overlap between objects A and B. Therefore, compared to the traditional IoU indicator, GIoU is able to assess the overlap between bounding box B and bounding box A more accurately. Figure 6 shows that at IoU values all equal to 0.33, positioning from left to right becomes less and less effective, and the value of GIoU decreases in turn.

B and A express two bounding boxes, and C represents the bounding box that can contain the area of bounding box A and bounding box B. This formula considers the effect of the area difference set of B and A by the ratio of the area intersection of B and A to the area union of B and A and subtracts it from the original IoU value to obtain a more accurate assessment of the degree of overlap. The GIoU metric is an important metric used in object detection tasks, which can help improve the performance of the simulator and enhance the precision of the detection outcome [32]. The IoU is shown in Equation (4). The GIoU is shown in Equation (5).

$$\mathrm{IoU} = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A} \cup \mathrm{B}|}, \tag{4}$$

$$GIoU = IoU - \frac{|C - (A \cup B)|}{|C|}, \tag{5}$$

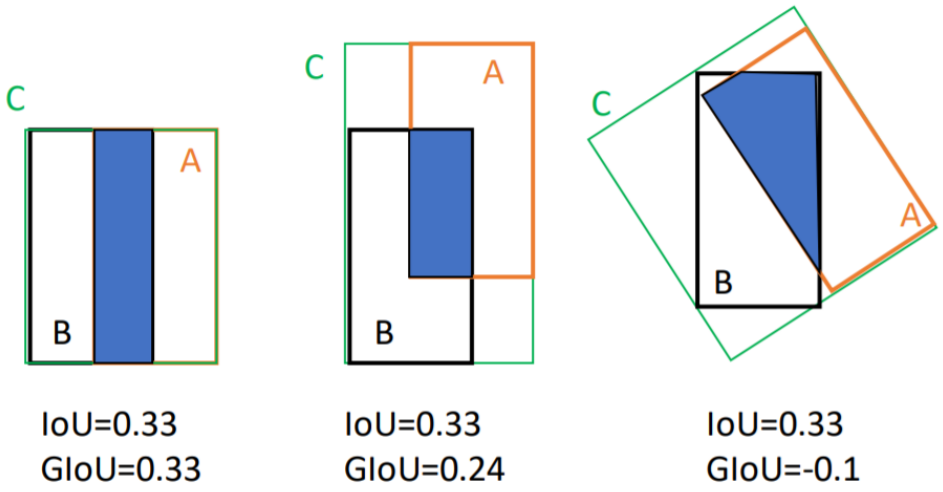

Figure 6. Different ways of overlapping detection frames.

The value range of GIoU is −1~1, which can be obtained from Equation (6), and its loss function range is 0~2.

$$L_{GIoU} = 1 - GIoU(0 \le L_{GIoU} \le 2), \tag{6}$$

The GIoU solves the problem of loss of 0 when bounding box A and bounding box B do not overlap to a certain extent, but there is also the problem that GIoU degenerates into IoU when real box and detection box are included, and the convergence of the two boxes is slow in the horizontal and vertical directions.

The CIoU is based on GIoU, which further considers the geometric factor of the aspect ratio of the bounding box, thus making the regression of the bounding box more stable and exact, where β and ν are the corresponding weights and aspect ratio coefficients, respectively. Specifically, in the calculation, β is used to balance the effect between the central point length and the length–width ratio, while ν is used as a parameter to survey the uniformity of the length–width ratio. The formula of CIoU is shown in Equation (7).

$$CIoU = IoU - \frac{\rho^2(A, B)}{c^2} - \beta v, \tag{7}$$

From GIoU to CIoU, all three loss functions have excellent performance in dealing with the bounding box of the inclusion relation, unlike IoU which degenerates. However, when the central points of B and A overlap, CIoU degenerate to IoU and do not regress well on the bounding box. Therefore, the EIoU (efficient IoU) loss function comes into being, which separates the impact factors of A and B on the basis of CIoU and calculates the width and height of A and B. The EIoU equation is as follows, where h^c and w^c are the width and height of the bounding box C. The EIoU loss equation is as Equations (8) and (9).

$$L_{EIoU} = L_{IoU} + L_{DIS} + L_{ASP}, \tag{8}$$

$$L_{EIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{(w^c)^2 + (h^c)^2} + \frac{\rho^2(w, w^{gt})}{(w^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2}, \tag{9}$$

3.5. Attention Mechanisms

Recently, introducing channel attention mechanisms into convolutional blocks has become a highly anticipated research direction. Efficient channel attention (ECA), a representative attention mechanism, has shown significant performance gains for various deep-learning network architectures after proposing an ECA module without dimensionality reduction. It effectively prevents the impact of dimensionality reduction on attention learning effect [33] and has significant performance gains for various deep-learning network architectures. Therefore, the ECA module is widely used in convolutional neural networks and performs well in feature extraction. In ECGYOLO, we use three ECA attention mechanisms, located behind the feature layers obtained by the second, third and fourth ECG modules, which allows us to focus more on the feature information of the input images.

4. Materials and Methods

4.1. Dataset Preparation and Processing

In this experiment, several rigorous measures were taken in the production of the dataset to ensure the diversity and balance of the dataset contents. A total of 10,043 multi-scene images of human faces and faces wearing masks were collected and accurately labelled according to detailed annotation files. For the web-crawled data, the authors used the LabelImg tool to annotate all images and generated xml files. In addition, the authors extracted more than 4000 face images and 5000 mask images from the public datasets WIDER Face [34] and MAFA to augment the multiplicity and number of datasets. The dataset was divided into three parts: training set, validation set and test set. The training set is used to learn data characteristics and continuously update the network arguments, and the validation set can find problems with the model or parameters in time after each round of training. The test set evaluates the trained model. These strict dataset production measures, and dataset partitioning methods can efficiently raise the accuracy and better cope with the face mask detection problem in various complex scenarios. Partial images of masks and faces in the dataset are shown in Figures 7 and 8.

Figure 7. Faces with masks.

Figure 8. Face pictures.

To reduce the risk of model overfitting and improve generalization capabilities, ECGYOLO employs a variety of data augmentation methods, including Mosaic and Mixup data enhancement and colour space conversion, among which, Mosaic data enhancement can fuse multiple images to produce a new picture, enriching the background of the picture.

While Mixup data augmentation blends two images to generate a new training sample, both strategies can efficiently enrich the number of targets and prevent the net from overfitting. These techniques can effectively enrich the diversity and complexity of training data, making the model more adaptive and robust and better handling image recognition problems in various scenarios.

4.2. Environment Configuration and Parameters

The RTX3060 graphics card is used for this training, and the whole network is fine-tuned so as to accelerate the learning efficiency of the model. The batch size is set to 24, set 1×10^{-5} is used as the value of the learning rate, the Adam algorithm is used for the optimizer and to prevent the model from overfitting. Label smoothing is used to enhance the model generalization ability [35]. The operating environment for the experiments in this study is shown in Table 1.

Table 1. Operating environment configuration.

Category	Metrics
Operating systems	Windows10
GPU	NVIDIA GeForce RTX 3060
CPU	Intel core i7 10750H
CPU main frequency	2.6 GHz
Memory	16 GB
CUDA	CUDA 11.5
Framework	PyTorch 1.11.0
Scripting languages	Python 3.9

In order to make the experiments objective, ECGYOLO mainly tests the accuracy of the model through AP and mAP. Precision calculation formula is given in Equation (10), and recall calculation formula is given in Equation (11).

$$P = \frac{TP}{TP + FP'}, \tag{10}$$

$$R = \frac{TP}{TP + FN}, \tag{11}$$

In the mask detection task, TP is the part that correctly identifies the "wearing a mask", FP is the part that mistakenly identifies "not wearing a mask" as "wearing a mask", and FN is the part that does not correctly detect "wearing a mask" or "not wearing a mask". Precision rate and recall, on the other hand, are two commonly used algorithm evaluation indicators to evaluate the capability of the algorithm to perform the mask detection task. Specifically, the precision rate is the percentage of mask wearers detected by the algorithm that are actually correct; the recall rate is the percentage of all mask wearer samples that are correctly detected by the algorithm. The recall rate is more concerned with the number of undetected mask wearers than the precision rate and provides a better measure of the comprehensiveness and effectiveness of the models. In a real-world environment, precision and recall are two indicators that are both contradictory and unified, so the authors need to consider the balance between both precision and recall, i.e., F1. The F1 calculation formula is given in Equation (12).

$$F1 = \frac{2 \times P \times R}{P + R}, \tag{12}$$

AP denotes the extent under the precision-recall chart, and the AP calculation formula is given in Equation (13). Value mAP is the mean of all APs, where N is the total value of species, and i denotes a category. Calculation formula for mAP is given in Equation (14).

$$AP = \int_0^1 P(R)dR, \tag{13}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N}, \quad (14)$$

5. Results

5.1. Network Model Comparison

The ECGYOLO model is improved by using GhostConv convolution, which simplifies the traditional convolution of complex operations. This optimization allows ECGYOLO to achieve a faster frame-per-second (FPS) time of 65.3 while maintaining high detection accuracy, which is only 6.2 less than that of the YOLOv7tiny model, 29 more than YOLOv7 and comparable to SSD. In addition, ECGYOLO is also lightweight. The model requires only 5.06M parameters, which is 94.64M smaller than the SSD model, 103.54M smaller than the highest Faster R-CNN, 31.8M smaller than YOLOv7 and even 1.04M smaller than YOLOv7tiny. This lightweight design makes ECGYOLO acceptable for most hardware devices and has good application prospects. The improved ECGYOLO model not only reduces the amount of code to achieve light weight but also improves the accuracy of the model. To specify our evaluation method, the authors set 640X640 to the resolution of all model input images, the optimizer used Adam, label smoothing, cosine annealing algorithm and non-extreme suppression. The maximum and minimum learning rates were set to 0.001 and 0.00001, and the momentum was set to 0.9. The detailed evaluation results of each model in precision, recall, F1, FLOPs and mAP parameters are listed in Table 2.

Table 2. Network Model Results.

Model	AP/%		mAP/%	Precision	Recall	F1	FLOPs(G)	FPS	Parameter /MB
	Face	Face_Mask							
YOLOv7tiny	82.5	94.2	88.35	0.89	0.817	0.851	13.8	71.5	6.2
YOLOv4	83.03	93.47	88.25	0.896	0.765	0.825	61.2	12.65	64.5
YOLOv7	83.7	93.8	88.75	0.86	0.823	0.841	104.3	36.3	36.9
SSD-vgg	76.6	92.3	84.45	0.801	0.825	0.812	272.1	64.4	99.7
FasterRCNN	80.6	91.5	86.05	0.784	0.85	0.815	371.7	8.5	136.69
(Mine)	89.1	96.4	92.75	0.962	0.876	0.917	11.3	65.3	5.06

Table 2 data show that ECGYOLO is 4.4% more accurate than YOLOv7tiny, overtops YOLOv7 by 3.0%, exceeds SSD by 8.3%, and outperforms FasterRCNN by 6.7%. The increase in precision is 6.6% above the second highest of YOLOv4, recall is 2.6% above the second highest of FasterRCNN, and F1 is also 6.6% higher than the second highest of YOLOv7tiny. These figures can reduce the number of wrong and missed checks to some extent. FPS is the second highest of that of the above models, only 6.2 below YOLOv7tiny. Parameter is 1.04M below the second lowest, and FLOPs are 2.5 lower than the second lowest, of YOLOv7tniy. The above figures are satisfactory for use and operation on most low-computing platforms. The models based on this dataset are shown in Figures 9 and 10.

Figure 9. Original images.

Figure 10. Original images.

The results of the pictures from this dataset are shown in Figures 11 and 12.

Figure 11. Renderings.

Figure 12. Renderings.

Due to the application of learning rate cosine annealing decay, the curve is fluctuating, and after 25 epochs, the curve as a whole has no decreasing trend, at which time, the Loss can be considered to have converged. The loss results and the map curve are shown in Figures 13 and 14.

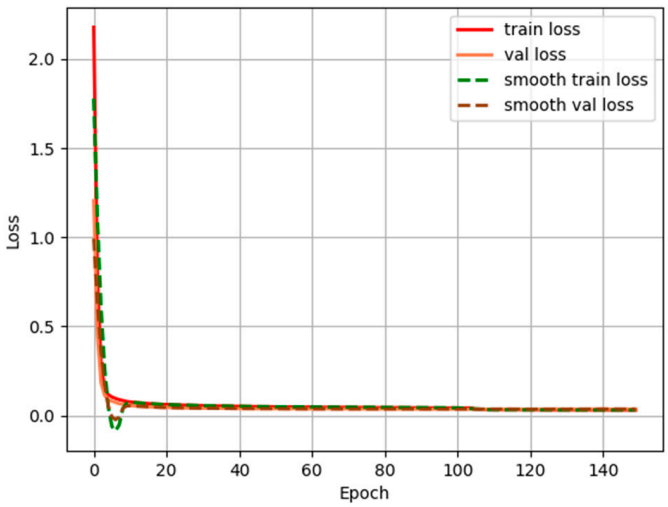

Figure 13. Convergence of ECGYOLO model.

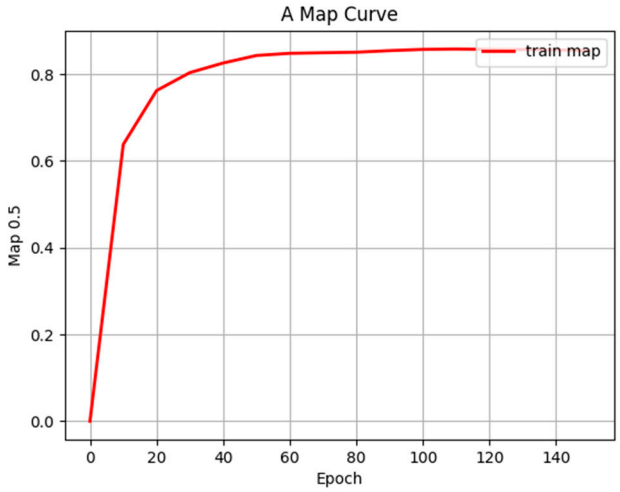

Figure 14. Map curve.

5.2. Ablation Experiments

Table 3 data show that the model accuracy, precision and recall are upgraded after replacing the normal convolution with GhostConv in YOLOv7tiny. Also, further improvements such as adding the ECA attention mechanism, replacing the EIoU loss function and replacing the activation function Hardswish can also raise the mAP of the model by 0.72, 0.3 and 0.4, respectively. In addition, when replacing the ordinary convolution with RepConv and replacing the ECG with ELAN, the accuracy is increased by 1.02 and 1.15. It is worth noting that these improved methods correspondingly increase the number of parameters by 0.94 M and 0.8 M. Although these methods increase the number of parameters partially, they also bring significant accuracy and precision and recall improvements. Therefore, these improvement methods are well worth trying and promoting. In Table 3, GC refers to GhostConv, RC denotes RepConv, and HS is Hardswish.

Table 3. Ablation experiment results.

Model	AP/%		mAP/%	Precision	Recall	Parameter /MB
	Face	Face_Mask				
YOLOv7tiny	82.5	94.2	88.35	0.89	0.817	6.2
YOLOv7tiny + GC	83.6	95	89.3	0.92	0.828	3.32
YOLOv7tiny + GC + EIoU	86.7	92.16	89.43	0.92	0.84	3.32
YOLOv7tiny + GC + EIoU + ECA	86.2	94.1	90.15	0.93	0.845	3.32
YOLOv7tiny + GC + EIoU + ECA + HS	88.2	92.96	90.58	0.93	0.856	3.32
YOLOv7tiny + GC + EIoU + ECA + HS + RC	88.3	94.9	91.6	0.95	0.86	4.26
YOLOv7tiny + GC + EIoU + ECA + HS + RC + ECG	89.1	96.4	92.75	0.962	0.876	5.06

6. Discussion

This study investigates mask-wearing testing and proposes a lightweight mask-wearing testing algorithm that has both higher detection speed and guaranteed detection accuracy and directs at the issue that present detection algorithms are slow and difficult to deploy on low-computing-power hardware platforms (e.g., embedded, mobile, etc.). Compared with YOLOv7tiny, ECGYOLO has faster speed, higher accuracy and is more lightweight. It uses some including cleaner ECG model, more efficient EIoU loss function and more efficient ECA attention mechanism. Compared with the better YOLOv7 and YOLOv7tiny on the market, ECGYOLO is better in mAP, F1 metrics and 1.14M smaller in the number of parameters than YOLOv7tniy. However, ECGYOLO also has some shortcomings. Because of the lightweight measures taken throughout this paper for deployment in low-computing-power platforms, the mAP of the face without mask is not too high. The mAP of the target result is not too high, only 89.1%. How to improve the AP of face targets

while retaining tall detection efficiency and light weight is the next problem to be solved. In summary, the ECGYOLO design scheme can meet the demand for lightweight, efficient and accurate models in practical applications compared to previous models in the target detection domain, so the emergence of the ECGYOLO model is of great significance for the large-scale application in the target detection domain and will provide great convenience for future intelligent applications.

7. Conclusions

In this document, a modified model ECGYOLO based on YOLOv7tiny is introduced, which is mainly used to resolve the detection discrepancy of mask wearing. In terms of model improvement, three main improvement methods are proposed. First of all, the ELAN module is replaced by the ECG, and the ordinary convolution is displaced by the GhostConv. Second, the ordinary convolution is replaced by RepConv to enhance precision for small target layers. Finally, in the layers of the head, neck and backbone network of the ECGYOLO model, the reorganization and optimization are completed by replacing the activation function Hardswish, replacing the loss function EIoU and appending the ECA attention.

The evaluation results suggest that the ECGYOLO model outperforms the Faster-RCNN, YOLOv7tiny, SSD and YOLOv7 models by 6.7%, 4.4%, 8.3% and 3.0%, respectively, in mAP. The model also exceeds YOLOv7tiny by 6.6%, overtops YOLOv7 by 6.7%, outperforms SSD by 10.5% and is 10.2% superior to c in terms of F1 metrics. In addition, in terms of FPS parameters, the FPS of the ECGYOLO model is 65.3, which is lower than YOLOv7tiny's 71.5 but somewhat higher than that of the rest of the models. In addition, the number of arguments of this model is 5.06 M, which is smaller than that of the smallest YOLOv7tiny by 1.14 M. Therefore, it can be seen that this model is lightweight and superior in mask detection.

Although the mask detection technique still faces some challenges and difficulties in practical scenarios, the authors believe that these problems can be gradually solved as the technology continuously upgrades, and the datasets continuously improve, thus providing people with a perfect mask detection solution. In conclusion, it is important to develop a mask inspection system based on YOLOv7tiny, and the high-performance and lightweight features of the ECGYOLO model will also make the model more advantageous in practical applications.

Author Contributions: Conceptualization, M.L., S.L. and W.H.; methodology, W.H.; software, J.Z.; validation, M.L., H.W. and K.Z.; formal analysis, M.L. and W.H.; investigation, J.Z.; resources, M.L. and Y.H.; data curation, M.L. and Y.H.; writing—original draft preparation, W.H., M.L. and J.Z.; writing—review and editing, S.L., M.L. and W.H.; visualization, H.W. and K.Z.; supervision, M.L.; project administration, M.L. and W.H.; funding acquisition, W.H. and S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by Sichuan Science and Technology Program (2023YFSY0026, 2023YFH0004).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: In this paper, we use publicly available datasets, including Ge Shiming's MAFA dataset (http://www.escience.cn/people/geshiming/mafa.html accessed on 10 June 2023) and the open-source WIDER FACE dataset of the Multimedia Laboratory, Department of Information Engineering, The Chinese University of Hong Kong (shuoyang1213.me/WIDERFACE/ accessed on 10 June 2023).

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

References

1. Liao, M.; Liu, H.; Wang, X.; Hu, X.; Huang, Y.; Liu, X.; Brenan, K.; Mecha, J.; Nirmalan, M.; Lu, J.R. A technical review of face mask wearing in preventing respiratory COVID-19 transmission. *Curr. Opin. Colloid Interface Sci.* **2021**, *52*, 101417. [CrossRef] [PubMed]
2. Hechun, W.; Xiaohong, Z. Survey of deep learning based object detection. In Proceedings of the 2nd International Conference on Big Data Technologies, Jinan, China, 28–30 August 2019; pp. 149–153.
3. Tong, K.; Wu, Y.; Zhou, F. Recent advances in small object detection based on deep learning: A review. *Image Vis. Comput.* **2020**, *97*, 103910. [CrossRef]
4. Fan, D.-P.; Ji, G.-P.; Cheng, M.-M.; Shao, L. Concealed object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 6024–6042. [CrossRef] [PubMed]
5. Li, K.; Wan, G.; Cheng, G.; Meng, L.; Han, J. Object detection in optical remote sensing images: A survey and a new benchmark. *ISPRS J. Photogramm. Remote Sens.* **2020**, *159*, 296–307. [CrossRef]
6. Qi, G.; Zhang, Y.; Wang, K.; Mazur, N.; Liu, y.; Malaviya, D. Small object detection method based on adaptive spatial parallel convolution and fast multi-scale fusion. *Remote Sens.* **2022**, *14*, 420. [CrossRef]
7. Wang, J.; Wang, Y.; Wu, Y.; Zhang, K.; Wang, Q. FRPNet: A feature-reflowing pyramid network for object detection of remote sensing images. *IEEE Geosci. Remote Sens. Lett.* **2020**, *19*, 1–5. [CrossRef]
8. Liu, Y.; Li, Q.; Yuan, Y.; Du, Q.; Wang, Q. ABNet: Adaptive balanced network for multiscale object detection in remote sensing imagery. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 1–14. [CrossRef]
9. Wan, S.; Goudos, S. Faster R-CNN for multi-class fruit detection using a robotic vision system. *Comput. Netw.* **2020**, *168*, 107036. [CrossRef]
10. Huang, Z.; Wang, J.; Fu, X.; Yu, T.; Guo, Y.; Wang, R. DC-SPP-YOLO: Dense connection and spatial pyramid pooling based YOLO for object detection. *Inf. Sci.* **2020**, *522*, 241–258. [CrossRef]
11. Jiang, P.; Chen, Y.; Liu, B.; He, D.; Liang, C. Real-time detection of apple leaf diseases using deep learning approach based on improved convolutional neural networks. *IEEE Access* **2019**, *7*, 59069–59080. [CrossRef]
12. Chen, J.; Liu, H.; Zhang, Y.; Zhang, D.; Ouyang, H.; Chen, X. A Multiscale Lightweight and Efficient Model Based on YOLOv7: Applied to Citrus Orchard. *Plants* **2022**, *11*, 3260. [CrossRef] [PubMed]
13. Liu, S.; Wang, Y.; Yu, Q.; Liu, H.; Peng, Z. CEAM-YOLOv7: Improved YOLOv7 Based on Channel Expansion and Attention Mechanism for Driver Distraction Behavior Detection. *IEEE Access* **2022**, *10*, 129116–129124. [CrossRef]
14. Liu, Y.; Wang, X. SAR Ship Detection Based on Improved YOLOv7-Tiny. In Proceedings of the 2022 IEEE 8th International Conference on Computer and Communications (ICCC), Chengdu, China, 9–12 December 2022; pp. 2166–2170.
15. Chang, Y.-L.; Anagaw, A.; Chang, L.; Wang, Y.C.; Hsiao, C.-Y.; Lee, W.-H. Ship detection based on YOLOv2 for SAR imagery. *Remote Sens.* **2019**, *11*, 786. [CrossRef]
16. Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [CrossRef]
17. Cai, Y.; Luan, T.; Gao, H.; Wang, H.; Chen, L.; Li, Y.; Sotelo, M.A.; Li, Z. YOLOv4-5D: An effective and efficient object detector for autonomous driving. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. [CrossRef]
18. Xu, R.; Lin, H.; Lu, K.; Cao, L.; Liu, Y. A forest fire detection system based on ensemble learning. *Forests* **2021**, *12*, 217. [CrossRef]
19. Jabeen, K.; Khan, M.A.; Alhaisoni, M.; Tariq, U.; Zhang, Y.-D.; Hamza, A.; Mickus, A.; Damaševičius, R. Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion. *Sensors* **2022**, *22*, 807. [CrossRef]
20. Zhang, T.; Zhang, X.; Ke, X. Quad-FPN: A novel quad feature pyramid network for SAR ship detection. *Remote Sens.* **2021**, *13*, 2771. [CrossRef]
21. Wang, C.; Zhang, B.; Cao, Y.; Sun, M.; He, K.; Cao, Z.; Wang, M. Mask Detection Method Based on YOLO-GBC Network. *Electronics* **2023**, *12*, 408. [CrossRef]
22. Kumar, A.; Kalia, A.; Kalia, A. ETL-YOLO v4: A face mask detection algorithm in era of COVID-19 pandemic. *Optik* **2022**, *259*, 169051. [CrossRef]
23. Yu, J.; Zhang, W. Face mask wearing detection algorithm based on improved YOLO-v4. *Sensors* **2021**, *21*, 3263. [CrossRef] [PubMed]
24. Wu, P.; Li, H.; Zeng, N.; Li, F. FMD-Yolo: An efficient face mask detection method for COVID-19 prevention and control in public. *Image Vis. Comput.* **2022**, *117*, 104341. [CrossRef]
25. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
26. Srinivasu, P.N.; SivaSai, J.G.; Ijaz, M.F.; Bhoi, A.K.; Kim, W.; Kang, J.J. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors* **2021**, *21*, 2852. [CrossRef]
27. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. GhostNet: More Features from Cheap Operations. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
28. Wu, H.; Luo, X.; Zhou, M.C. Advancing non-negative latent factorization of tensors with diversified regularization schemes. *IEEE Trans. Serv. Comput.* **2020**, *15*, 1334–1344. [CrossRef]
29. Wang, S.-H.; Muhammad, K.; Hong, J.; Sangaiah, A.K.; Zhang, Y.-D. Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization. *Neural Comput. Appl.* **2020**, *32*, 665–680. [CrossRef]

30. Liu, L.; Qiang, B.; Wang, Y.; Yang, X.; Tian, J.; Zhang, S. Object Detection Algorithm Based on Coordinate Attention and Context Feature Enhancement. In Proceedings of the 2022 11th International Conference on Computing and Pattern Recognition, Beijing, China, 17–19 November 2022; pp. 95–101.
31. Rezatofighi, 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.
32. Khanna, S.; Cao, J.; Bai, Q.; Xu, G. PRICAI 2022: Trends in Artificial Intelligence. In Proceedings of the 19th Pacific Rim International Conference on Artificial Intelligence, PRICAI 2022, Shanghai, China, 10–13 November 2022.
33. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
34. Lo, E. Target Detection Algorithms in Hyperspectral Imaging Based on Discriminant Analysis. *J. Image Graph.* **2019**, *7*, 140–144. [CrossRef]
35. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.