

Essay

YOLOv5-Ytiny: A Miniature Aggregate Detection and Classification Model

Sheng Yuan, Yuying Du *, Mingtang Liu, Shuang Yue, Bin Li and Hao Zhang

College of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450000, China; yuansheng@ncwu.edu.cn (S.Y.); liumingtang@ncwu.edu.cn (M.L.); syncwu@yeah.net (S.Y.); mlking44@yeah.net (B.L.); 201616710@stu.ncwu.edu.cn (H.Z.)

* Correspondence: dyyncwu@yeah.net

Abstract: Aggregate classification is the prerequisite for making concrete. Traditional aggregate identification methods have the disadvantages of low accuracy and a slow speed. To solve these problems, a miniature aggregate detection and classification model, based on the improved You Only Look Once (YOLO) algorithm, named YOLOv5-ytiny is proposed in this study. Firstly, the C3 structure in YOLOv5 is replaced with our proposed CI structure. Then, the redundant part of the Neck structure is pruned by us. Finally, the bounding box regression loss function GIoU is changed to the CIoU function. The proposed YOLOv5-ytiny model was compared with other object detection algorithms such as YOLOv4, YOLOv4-tiny, and SSD. The experimental results demonstrate that the YOLOv5-ytiny model reaches 9.17 FPS, 60% higher than the original YOLOv5 algorithm, and reaches 99.6% mAP (the mean average precision). Moreover, the YOLOv5-ytiny model has significant speed advantages over CPU-only computer devices. This method can not only accurately identify the aggregate but can also obtain the relative position of the aggregate, which can be effectively used for aggregate detection.



Citation: Yuan, S.; Du, Y.; Liu, M.; Yue, S.; Li, B.; Zhang, H.

YOLOv5-Ytiny: A Miniature Aggregate Detection and Classification Model. *Electronics* **2022**, *11*, 1743. <https://doi.org/10.3390/electronics11111743>

Academic Editor: Pedro Latorre-Carmona

Received: 27 April 2022

Accepted: 26 May 2022

Published: 30 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/>).

Keywords: object detection; aggregate; YOLO; classification; computer vision

1. Introduction

Aggregate classification is an important factor for determining the performance and quality of concrete. Concrete is composed of cement, sand, stones, and water. Aggregate generally accounts for 70% to 80% of concrete [1]. Many factors affect the strength of concrete, mainly including the cement strength and water–binder ratio, the aggregate gradation and particle shape, the curing temperature and humidity, the curing age, etc. [2]. The aggregate processing system is one of the most important auxiliary production systems used in the construction of large-scale water conservancy and hydropower projects [3]. Aggregate quality control is of great significance to promote the sound development of the engineering construction industry [4], it is also extremely important for improving the quality of a project and optimizing the cost of a project [5]. Different types of aggregate have different effects on the performance of concrete [6]. Regarding the particle size and shape of the aggregate, the current specifications for coarse aggregate needle-like particles are relatively broad [7], and good-quality aggregate needs to have a standardized particle size and shape [8]. Therefore, we must ensure the quality requirements of aggregate and select raw materials are reasonable to ensure the quality of concrete. It is particularly important to find a suitable aggregate classification and detection method.

In recent years, the level of aggregate classification and detection has greatly improved [9], and there are now a variety of sand particle size measurement methods. These include, for example, mesoscale modeling of concrete static and dynamic tensile fractures for real shape aggregates [10], the development of a particle size and shape measurement system for manufactured sand [11], the use of extreme gradient boosting-based pavement aggregate shape classification [12], the use of the wire mesh method to sort aggregate

gardens [13], a method for evaluating the strength of individual ballast aggregates by point load testing and establishing a classification method [14], the determination of particle size, and core and shell size determination of core-shell particle distribution by analytical ultracentrifugation [15], the use of the projected area of the particle to calculate the surface area, equivalent diameter, and sphericity [16], the use of imaging methods to obtain reliable particle size distribution [17], the use of a vibration dispersion system, feed blanking system, and backlight image acquisition system to construct a particle size and shape measurement device [18]. Isa et al. proposed an automatic intelligent aggregate classification system combined with robot vision [19]. Sun et al. proposed a coarse aggregate granularity classification method based on a deep residual network [20]. Moaveni et al. developed a new segmentation technology that can capture images quickly and reliably to analyze their size and shape [21]. Sincen et al. established a laser-based aggregate shape representation system [22], which classifies aggregate from the features extracted from the created 3D images.

However, these screening methods can only measure the size of sand particles offline. Although digital image processing methods use more mature technical means [23,24], the research of these methods mainly focuses on the evaluation index of the shape characteristics of the aggregate [25], which cannot achieve the efficient real-time detection of images. In reality, with regards to the detection background of aggregate, the size of the detection target, the day and night light, and the difference in detection distances, the transmission of these detection targets to the processing side may cause different interferences. In this case, it is necessary to first detect the target position and locate and frame the target to reduce signal interference as much as possible, and, at the same time, detect target objects under different characteristics. Therefore, the real-time detection of aggregate features under complex backgrounds is of great significance.

In summary, this work brings the following main contributions:

- The design of a new type of aggregate detection and classification model, which can accurately identify the types of aggregate under complex backgrounds, such as different light, different distances, and different states of dry and wet aggregate.
- The improvement of YOLOv5, replacing the C3 module in the model backbone network, tailoring the Neck structure, and realizing the compression of the model, so that the model can be quickly detected on a computer that does not support GPU. The loss function is improved making the object frame selection more accurate.
- In the original YOLOv5 model, the original three detection heads are simplified into two, which is more suitable for the detection of a single target (only one target is recognized in a picture), thus reducing the number of parameters and the number of calculations.

2. Related Work

There are many mature target detection algorithms, such as YOLOv4 [26], SSD [27], YOLOv4-tiny, and YOLOv5 [28]. Compared with these algorithms, YOLOv5 is lighter and more portable. YOLOv5 uses a backbone feature extraction network, acquires the depth features of the input image, uses feature fusion to further improve the effectiveness of features, effectively frames the detection target, and improves the precision of target detection [29]. At present, YOLO is also widely used as a popular target detection algorithm. Yan et al. proposed a real-time apple targets for picking detection method robot based on improved YOLOv5 [30]. Yao et al. proposed a ship detection method in optical remote sensing images based on deep convolutional neural networks [31]. Gu et al. proposed a YOLOv5-based method for the identification and analysis of emergency behavior of caged laying ducks [32]. Zhu et al. proposed traffic sign recognition based on deep learning [33]. Fan et al. proposed a strawberry ripeness recognition algorithm combining dark channel enhancement and YOLOv5 [34]. A cost-performance evaluation of livestock activity recognition services using aerial imagery was proposed by Lema et al. [35] Jhong et al. proposed a night object detection system based on a lightweight and deep network of

the internet of vehicles [36]. Wang et al. proposed a fast smoky vehicle detection method based on improved YOLOv5 [37]. Wu et al. studied the application of YOLOv5 in the detection of small targets in remote sensing images [38]. Song et al. proposed an improved YOLOv5-based object detection method for grabbing robots [39].

Although YOLO v5 is much lighter than other object detection algorithms, the network structure is complex, with many layers, a large number of nodes, and limited experimental equipment. If it was running on one CPU, it would take longer during the actual training and inference.

In order to solve these problems, this experiment established an aggregate classification detection model, YOLOv5-ytiny, based on YOLOv5 in a complex background, compressing the YOLOv5 model, extracting complex detection background features in different environments, improving the detection speed, and providing real-time judgment of the classification of aggregate.

3. Materials and Methods

3.1. Data Collection and Processing

In this experiment, a high-definition camera is used to collect images, and the real-time images obtained by the camera are transmitted to the client. The model classifies and recognizes the acquired images, and then displays the results to the client. Figure 1 is a schematic diagram of image acquisition.

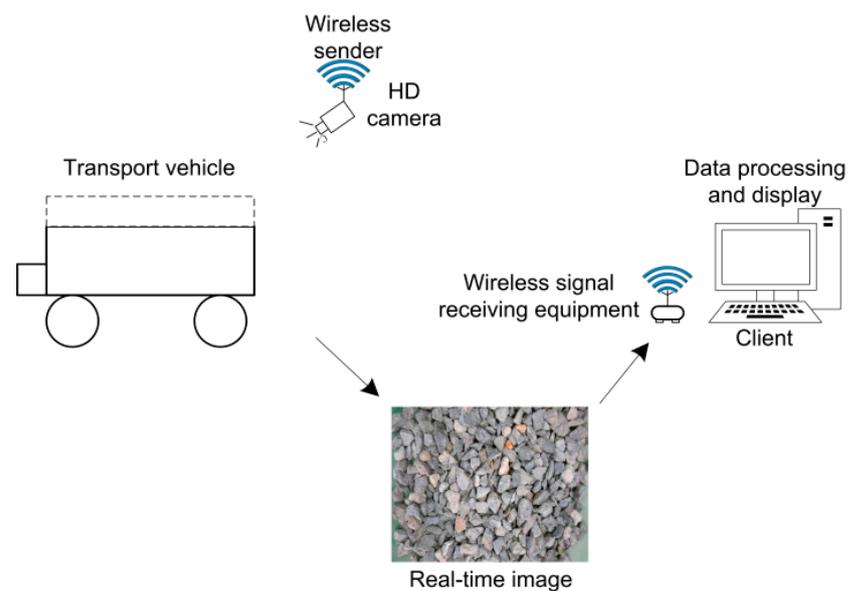


Figure 1. Image acquisition process: after the transport vehicle arrives, the aggregate image is collected, and the model returns the result after processing.

When capturing images, the camera is fixed 2 m above the aggregate box. Considering that the distance between the car and the camera is within 1–2 m during actual transportation, the effective detection distance we expect is also within 1–2 m. The image collection is shot under natural light and night lighting. The shooting result is saved as a 1920 pixel \times 1080 pixel RGB image. There are 4 types of aggregate, namely stones, small stones, machine-made sand, and surface sand. The particle size of the stones is in the range of 3–4 cm, the particle size of small stones is in the range of 2–3 cm, the particle size of machine-made sand is in the range of 1–2 cm, the particle size of surface sand is in the range of 0.1–0.5 cm. A total of 525 images in four types were taken, including different light conditions, dry and wet aggregate, and different shooting distances.

Taking the stones in Figure 2a and the small stones in Figure 2b as examples, the unit grayscale number of stones is distributed in the (130,180) pixel interval, and the unit grayscale number of small stones is distributed in the (120,180) pixel interval. The

grayscale distributions of these two types of aggregate have highly overlapping areas. It may be difficult to segment an image based on gray threshold, and it can be seen in Figure 3 that the stones and small stones are stacked. If an image is segmented using an image-based grayscale threshold segmentation method, it may not be possible to segment a single aggregate target because the grayscales connected to each other in the region are the same, which may easily cause the targets to stick together. On the other hand, the image of the aggregate with a short collection distance is clear, but the image is blurred when the distance is farther, and it is difficult to perform image processing. Therefore, this experiment uses the target detection algorithm YOLOv5 to extract the characteristics of aggregate in different backgrounds to realize the type recognition of aggregate.

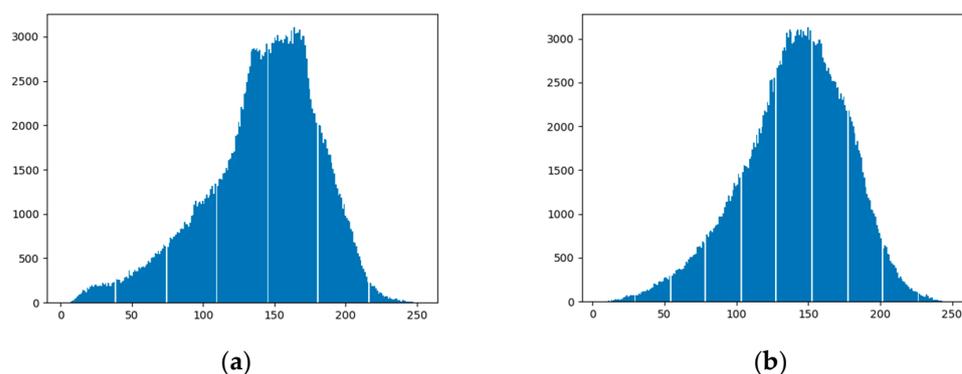


Figure 2. Comparison of grayscale histograms of stones and small stones. (a) Grayscale histogram of stones. (b) Grayscale histogram of small stones.



Figure 3. Stones and small stones. (a) Stones. (b) Small stones.

In this experiment, a total of 525 images were obtained. They were of stones, small stones, machine-made sand, and surface sand. In this experiment, the labels they delineated are sz, xsz, jzs, ms. We use LabelImg to label images, and the smallest bounding rectangle of the target was used as the real frame. In the final data set, 80% (420 sheets) were randomly selected as the training set and 20% (105 sheets) as the test set.

The four types of aggregate show different shapes and colors due to different dry and wet states and different brightness of light. The image collection environment was under cloudy, sunny, and night conditions. The aggregate states were dry, normal, and wet. The collection distance was 1.5 m.

The computer used in the research institute is Intel(R) Core(TM) i5-8250U, 1.80 GHz processor, running memory is 8 GB, storage memory is 512 GB, and the development environment is python 3.6.

3.2. Model Establishment and Experimental Process

Aggregate Classification Model Based on Improved YOLOv5

The technical route of the aggregate classification model is shown in Figure 4. The artificially-labeled aggregate data are input into the YOLOv5 model for training and fine-tuning, to realize the real-time recognition of the target. The improved YOLOv5-ytiny model is used for the classification of aggregate under complex backgrounds based on the YOLOv5 model. YOLOv5-ytiny replaces the C3 module of the backbone structure of the network structure, cuts the Neck structure to achieve compression, reduces the network prediction header, reduces the image size, and adjusts the network width. It simplifies the structure and parameters of the model while ensuring precision.

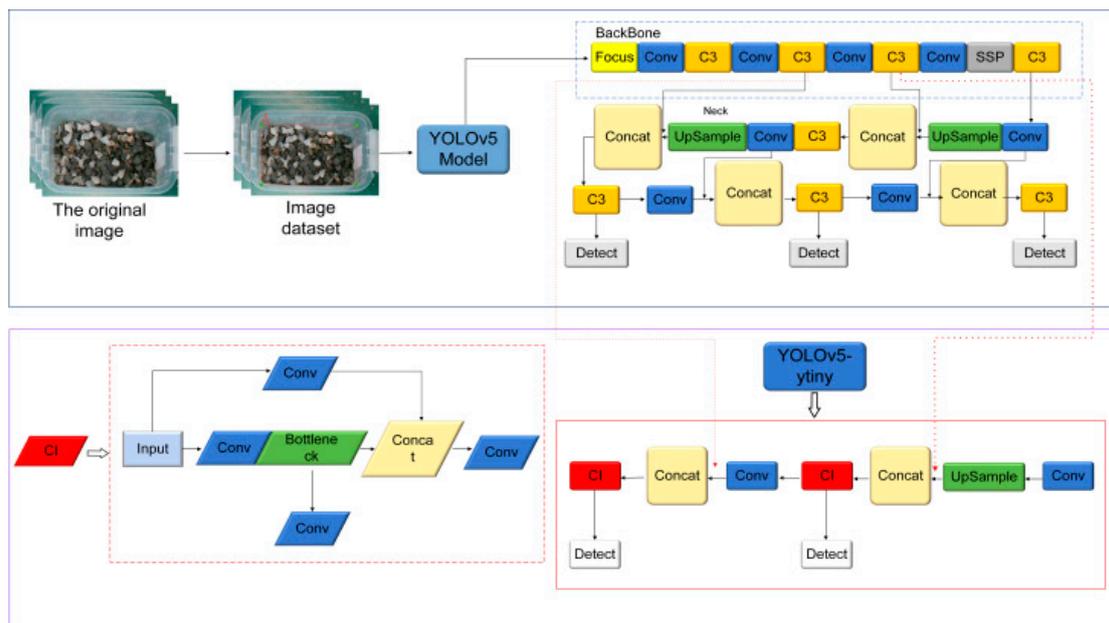


Figure 4. Improvements to YOLOv5 model.

The YOLOv5 algorithm has four network structures, namely YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. These four network structures are different in width and depth, but they are the same in principle. These four network structures can be flexibly selected according to need. The greater the depth of the structure selection, the higher the precision, but the training speed and inference speed also decrease. Aggregate is not a complex target. We hope to increase the speed of reasoning. Therefore, the selected structure is YOLOv5s, and improvements are made on this basis.

The YOLOv5 model mainly consists of the following five modules: (1) The Focus module slices the input image, which can achieve the effect of down-sampling the image without losing information. (2) The Conv module, three functions are encapsulated in this basic convolution module, including convolution (Conv2d) layer, BN (Batch Normalization) batch normalization layer, and SiLU (Swish) activation function, which realizes that the input features are passed through the convolution layer and the activation function. Through the normalization layer, the output layer is obtained. (3) The Bottleneck module, is mainly used to reduce the number of parameters, thereby reducing the amount of calculation, and after dimensionality reduction, data training and feature extraction can be performed more effectively and intuitively. (4) The C3 module, in the new version of YOLOv5, the author converts the Bottleneck CSP (bottleneck layer) module to the C3 module. Its structure and function are the same as the CSP architecture, but the selection of the correction unit is different. It contains three standard convolutional layers and multiple bottleneck modules. (5) SPP module, spatial pyramid pooling. The main purpose of this module is to fuse more features of different resolutions to obtain more information.

3.3. YOLOv5-Ytiny

Although the established YOLOv5 model can realize the detection and classification of aggregate, the structure and parameters of the model are still relatively large, and the calculation takes a long time. On the other hand, the detection and classification of aggregate are generally in the process of vehicle transportation, and the detection results need to be displayed in real-time. Therefore, to improve the detection speed of the model and reduce the amount of calculation, the model is optimized and compressed to form the YOLOv5-ytiny model.

Replace the C3 module of YOLOv5 with the CI module. As shown in Figure 5, the C3 module of YOLOv5 has a shortcut structure, which connects two adjacent layers of networks, and there are n residual blocks. For aggregate targets, the data set belongs to relatively simple target recognition. Multiple residual modules in the C3 module may be a waste of resources. Thus, replace it with a CI module. The structure of CI is shown in Figure 5.

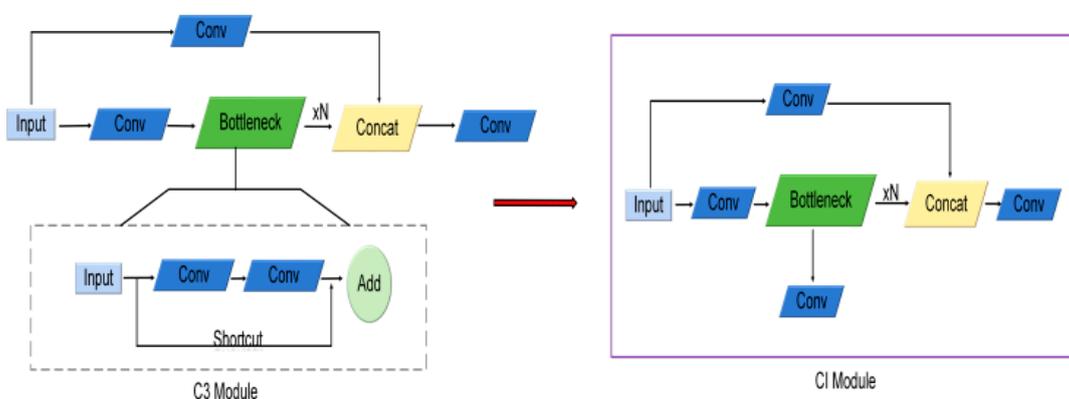


Figure 5. Replace C3 module with CI module.

After the C3 module is replaced with the CI module, the corresponding network structure in the original Backbone module and Neck module will also be changed. At the same time, all the layers of the C3 module in the original model are changed to one layer to reduce the overall depth of the network. In order to achieve model compression, after replacing the C3 module, cut the Neck module to remove the relatively redundant part of the network, and then delete part of the structure to reduce the depth of the network, reduce the amount of model calculations. The modified Neck module is shown in Figure 6.

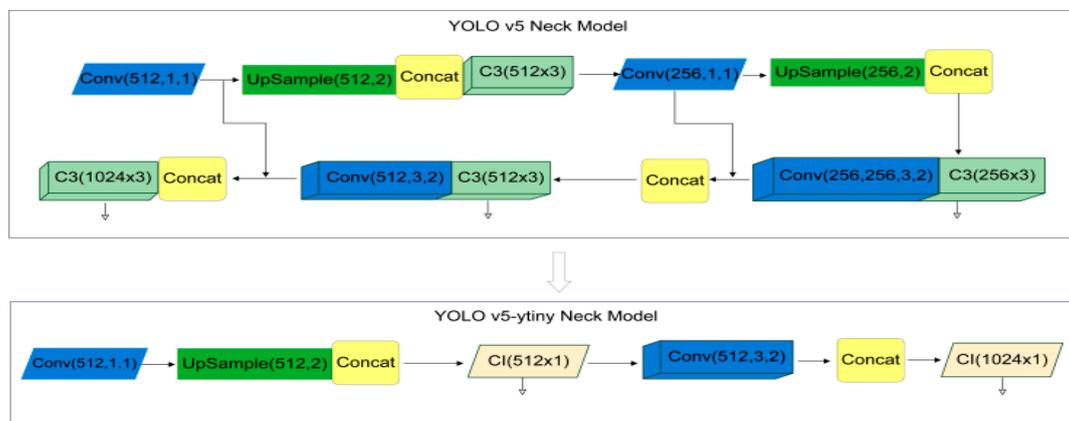


Figure 6. Modify the Neck module to compress the network structure.

There are three modules in the original YOLOv5 model, namely the Backbone module, the Neck module, and the detection head. After replacing the C3 module with the CI

module, in order to reduce the amount of calculation parameters, the detection head was changed from three to two. The original Neck module is composed of multiple convolutional layers (Conv), up-sampling, and tensor stitching (Concat). For a single and simple target, only a part of the combination in the Neck layer may be good. The original repeated multi-layer Neck structure may cause data redundancy, increase the amount of calculation, so we tailor the Neck structure to compress the network. We use fine-tuning-iterative training, a small number of times to cut the network structure, the number of times of training, adjusting it by judging the convergence effect, and the final network structure is shown in Figure 6. It can be seen that YOLOv5-tiny eliminates part of the repetitive hierarchy, retains the main network structure, and finally compresses it into two detection heads.

3.4. Improvement of Loss Function

YOLOv5s uses *GIoU* Loss as the bounding box regression loss function to judge the distance between the predicted box and the ground truth box. The formula is as follows.

$$IoU = \frac{A \cap B}{A \cup B} \quad (1)$$

$$GIoU = IoU - \frac{A^c - u}{A^c} \quad (2)$$

$$L_{GIoU} = 1 - GIoU \quad (3)$$

In the above formula, A is the predicted box, B is the ground truth box, IoU represents the intersection ratio of predicted box and ground truth box, A^c represents the intersection of predicted box and ground truth box, u represents the smallest circumscribed rectangle of predicted box and ground truth box, and L_{GIoU} is the *GIoU* Loss.

The original YOLOv5 model uses *GIoU* Loss as a position loss function to evaluate the distance between the predicted box and the ground truth box, but *GIoU* Loss cannot solve the situation where the prediction frame is inside the target frame and the size of the prediction frame is the same. In addition, the bounding box regression is not accurate enough, the convergence speed is slow, and only the overlap area relationship is considered. *CIoU* Loss takes into account the scale information of the aspect ratio of the bounding box, and measures it from the three perspectives of overlapping area, center point distance, and aspect ratio, which makes the prediction box regression more effective.

$$L_{loc} = 1 - IoU(B, B_{gt}) + \frac{d^2}{c^2} + av \quad (4)$$

$$a = \frac{v}{1 - IoU + v} \quad (5)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} \right) - \arctan \frac{w}{h} \quad (6)$$

In the above formula, w and h are the width and height of the prediction box, respectively, and w^{gt} and h^{gt} are the width and height of the ground truth box.

Compared with the *GIoU* Loss used in YOLOv5s, *CIoU* Loss takes into account the overlapping area, center point distance, and aspect ratio for measurement, so that the network can ensure faster convergence of the prediction frame during training and obtain higher regression positioning accuracy; this paper uses *CIoU* Loss as the loss function of the aggregate classification detection model.

4. Experimental Results and Analysis

4.1. Experimental Results

The YOLOv5-tiny model detects aggregate as shown in Figure 7. Taking small stones as an example, they are tested at different distances, different illuminations, and different

dry humidities of aggregate. It can be seen that Figure 7a–c are, respectively, the detection results on cloudy days, sunny days, and at night at a distance of 1.5 m. Figure 7d–f are the test results at distances of 1 m, 1.5 m, and 2 m, respectively. Figure 7g–i show the identification of small stones under dry, normal, and wet conditions. The accurate classification of aggregate under different backgrounds is realized. After verification, the effective recognition range of the YOLOv5-tytiny model is between 1 m and 2 m.

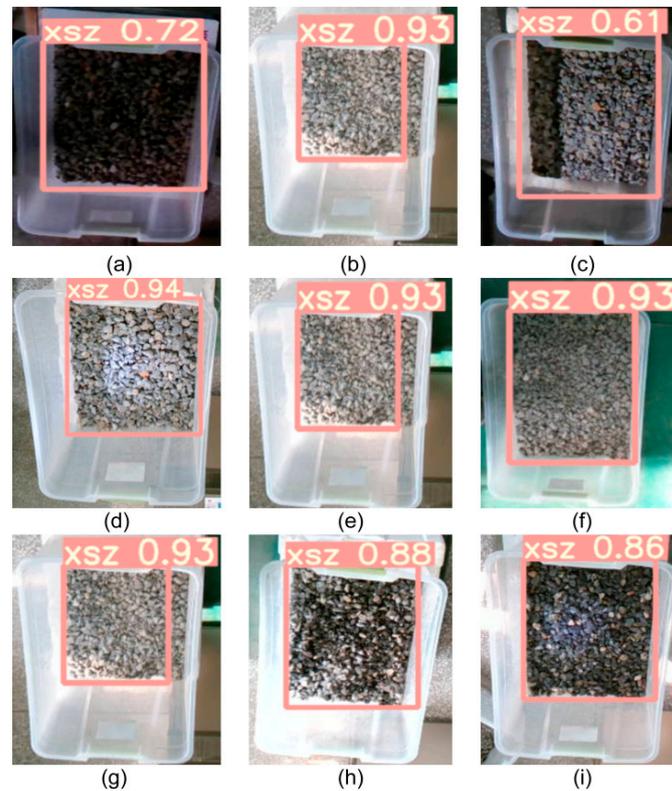


Figure 7. Small stones under different distances, light, and dry humidity. (a) is the small stones on a cloudy day, (b) is the small stones on a sunny day, (c) is the small stones of the night, (d) is the small stones at a distance of 1 meter, (e) is the small stones at a distance of 1.5 meters, (f) is the small stones at a distance of 2 meters, (g) is the small stones in the dry state, (h) is the small stones in the normal state, (i) is the small stones in the wet state.

Table 1 shows the confidence levels of the four types of aggregate under different conditions. The inspection was carried out under different light conditions on sunny and cloudy days, and at night, and at different distances, namely 1 m, 1.5 m, and 2 m. When inspecting the aggregate, the dry and wet state of the aggregate was also different. Tests were carried out in dry, normal, and wet conditions.

Table 1. Detection confidence in different sates.

Category Confidence/% State	Different Lighting			Different Distance			Different Wet And Dry		
	Cloudy Day	Sunny	Night	1 m	1.5 m	2 m	Dry	Normal	Moist
Stones	92	96	94	94	96	68	96	95	95
Small stones	72	93	61	94	93	93	93	88	86
Machine-made sand	86	80	78	90	80	83	80	83	89
Surface sand	95	95	86	94	95	70	95	96	95

In total, there were 300 iterations of the YOLOv5 model and YOLOv5-tytiny respectively. Figure 8 shows the change trend of the classification loss function. We can see

that the original YOLOv5 loss function drops rapidly in the early stages of the iteration, indicating that the model is quickly fitting. The convergence speed of YOLOv5-ytiny in the early stage is slower than that of the original model. By 200 iterations, the loss values of the two models are basically the same, indicating that the convergence effect of YOLOv5-ytiny is good, and the learning efficiency of the model is high. As the iterations continue, about 140 times, the model loss value decreases slowly. When the iteration reaches 220 times, the loss value fluctuates at 0.001 and the model reaches a stable state.

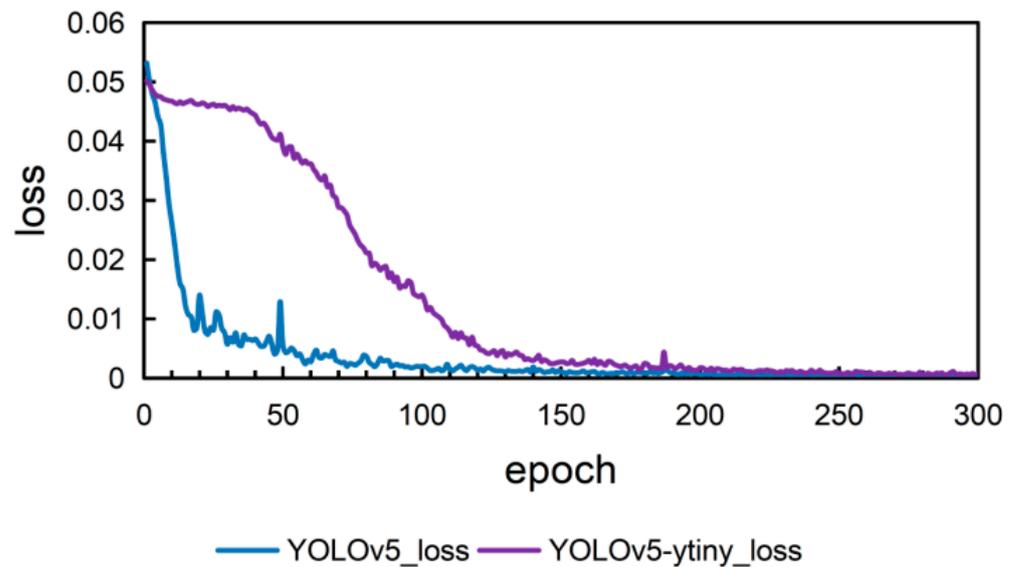


Figure 8. Comparison of the epoch process of the two models. Note: the experiment mainly seeks to compress the model and reduce the space occupied by the model, so it is normal to have a certain impact on precision and function loss.

4.2. Evaluation and Analysis of Model Performance

This paper selects the commonly used evaluation indicators of the target detection model: precision, recall, balanced F score (F1-score), mean average precision (*mAP*) and FPS (frames per second) for three evaluations of the model with three indicators.

The formula is as follows

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$F_1 - score = 2 * \frac{P * R}{P + R} \quad (9)$$

$$AP = \frac{\sum P}{Num(TotalObjects)} \quad (10)$$

$$mAP = \frac{\sum AP}{N(class)} \quad (11)$$

In the above formula, *TP* is the number of positive examples that are correctly classified. *FP* is the number of positive examples that are incorrectly classified. *FN* is the number of negative examples that are incorrectly classified, *TN* is the number of negative examples that are correctly classified. *AP* is the average precision, *mAP* is the mean of each type of *AP*. *Num* is the target number of each category. *N* is the total category.

4.3. Comparison with Original Model

The comparison of the evaluation indicators between the original YOLOv5 model and the YOLOv5-ytiny model is shown in Figure 9. This shows the precision, recall, and

F1-score of the YOLOv5-tiny and the original model YOLOv5. The precision of the YOLOv5-tiny is 96.5%, which is 0.2% lower than the original model, the recall rate is 98.5%, 0.4% higher than the original YOLOv5 model, and the F1-score is 97.5%, which is 0.1% higher than the original model.

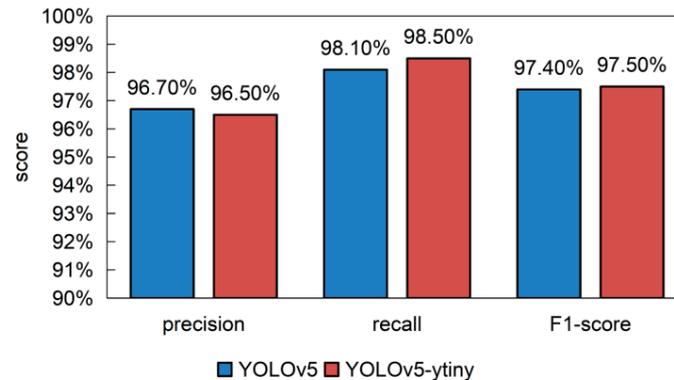


Figure 9. Comparison of model evaluation indicators.

Below, the YOLOv5-tiny model is compared with some parameters of the original YOLOv5 model. As shown in Table 2, the total parameters of the YOLOv5-tiny model are 19,755,583 fewer than the original YOLOv5 model. The mAP is 99.6%, which is consistent with the original YOLOv5 model. Precision is reduced by 0.2% compared to the original model, so there is no significant drop. YOLOv5-tiny's storage space is 3.04 MB, which is 10.66 MB smaller than the original YOLOv5, and the calculation time is 0.04 s, which is 60% faster than YOLOv5. The data in Table 2 show that the precision of the YOLOv5-tiny model is consistent with the original YOLOv5 mAP, with a slight decrease in precision, and the calculation speed is greatly improved compared with the original YOLOv5 model.

Table 2. Improved model parameters.

Parameter	YOLO v5	YOLOv5-Ytiny
Total parameter quantity/piece	21,375,645	1,620,062
mAP/%	99.6	99.6
Precision/%	96.7	96.5
Model storage space/MB	13.7	3.04
Computing time/s	0.1	0.04

4.4. Comparison with Other Target Detection Models

In the field of target detection, SSD, YOLOv4, and YOLOv4-tiny have high detection precision. In order to verify the effectiveness of this method, the training set of this paper was used to train these three models and the YOLOv5 model, respectively, and the test data were set to evaluate the performance of these four algorithms and obtain the precision, recall, and F1-score of the four algorithms. The comparison results are shown in Figure 10.

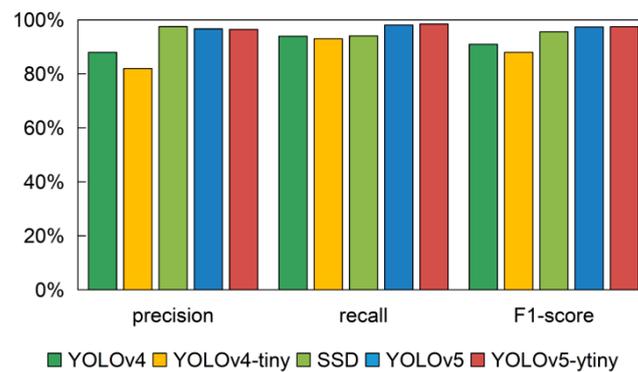


Figure 10. Comparison of various algorithm evaluation indicators.

We can see from Figure 10 that among the five algorithms, the SSD algorithm has the highest precision rate. The recall rate and F1-score of the algorithm in this paper are the highest. The F1-score is defined as the harmonic average of the precision and recall rates. It is a measure of classification problems. In some machine learning competitions for multi-classification problems, the F1-score is often used as the final evaluation method. YOLOv5-ytiny has superiority.

Table 3 compares the comprehensive evaluation indicators of the five algorithms. The precision, mAP, model storage space, and FPS of these five algorithms are compared, respectively. The comparison results are shown in Table 3.

Table 3. Comparison of comprehensive evaluation of algorithms.

Network	Precision/%	mAP/%	Model Storage Space/MB	FPS (Frame/s)
YOLO v4	88	99.62	244.24	0.4
YOLO v4-tiny	82	91.78	22.47	0.4
SSD	97.54	96.94	92.6	0.65
YOLO v5	96.7	99.6	13.7	9.17
YOLO v5-ytiny	96.5	99.6	3.04	22.73

Compared with the other four models, the detection speed of YOLOv5-ytiny is faster. The mAP of the improved YOLOv5-ytiny model is the same as that of the original YOLOv5 model, and the model storage space is reduced by 78%. In terms of the detection speed, the method in this paper has the fastest detection speed. The precision of the SSD and YOLOv5 models is slightly higher than that of YOLOv5-ytiny. The precision of the YOLOv5-ytiny model is 8.5% and 14.5% higher than that of YOLOv4 and YOLOv4-tiny, respectively. In terms of the models' storage space and detection speed, YOLOv5-ytiny has an absolute advantage, and while improving the detection speed, the mAP of the YOLOv5-ytiny model is consistent with the original YOLOv5 model, and the precision of the model has not dropped significantly.

In summary, compared with the other four models, YOLOv5-ytiny has a smaller model storage space and a faster detection speed. The detection speed is higher than that of YOLOv4, YOLOv4-tiny, SSD, and YOLOv5, 22.33 f/s, 22.33 f/s, 22.08 f/s, and 13.56 f/s, respectively. The detection precision of YOLOv5-ytiny is high, the detection speed is fast, and the improved space is small, which proves that the aggregate detection classification model YOLOv5-ytiny, based on the improved YOLOv5, has good practicability.

4.5. Practical Application

The experiment was conducted in cooperation with Zhengzhou Sanhe Hydraulic Machinery Co., Ltd. The experimental method was applied to the concrete batching plant for the preparation of concrete raw materials. Consistent with the model used in this experiment, the labels were inconsistent; we divided the results into four states, namely null (representing the unloaded state), complete unloading state, melon stones, and stones12.

The experimental results are shown in Figure 11a–d. The number on the image label is the probability of recognition.

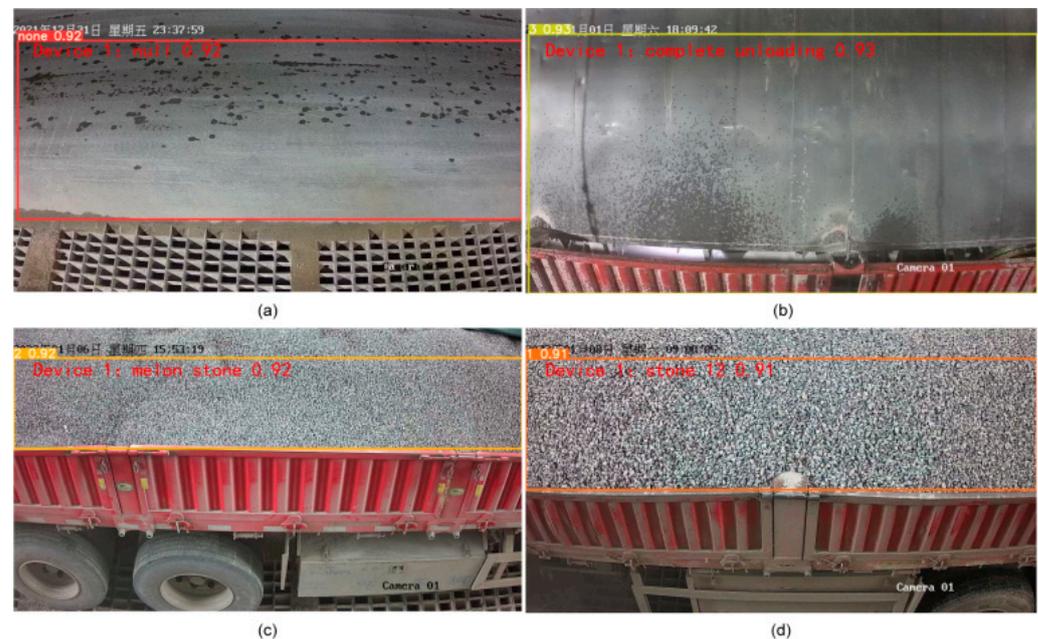


Figure 11. Examples of on-site aggregate inspection applications. (a) is the unloaded state, (b) is the state of completion of unloading, (c) represents that the aggregate is melon stones, (d) represents that the aggregate is 12 stones.

5. Discussion

Although tremendous progress has been made in the field of object detection recently, it remains a difficult task to detect and identify objects accurately and quickly. Yan et al. [30] named the YOLOv5 as the most powerful object detection algorithm in present times. In the current study, the overall performance of YOLOv5 was better than YOLOv4 and YOLOv3. This finding is in line with some previous researches, as we found several studies comparing YOLOv5 to previous versions of YOLO, such as YOLOv4 or YOLOv3. According to a study by Nepal [40], YOLOv5 is more accurate and faster than YOLOv4. YOLOv5 was compared to YOLOv3 and YOLOv4 for picking apples by robots, and the mAP was increased by 14.95% and 4.74%, respectively [30]. Similar results and comparisons with other YOLO models were demonstrated by [32] while using YOLOv5 to detect the behavior of cage-reared laying ducks. The recall (73%) and precision (62%) of YOLOv5 was better compared to YOLOv3-tiny (57% and 45%, respectively) for ship detection in satellite remote sensing images [31]. In experiments on grape variety detection, YOLOv5 had higher F1 scores than YOLOv4-tiny [41]. In our experiment, YOLOv5 also showed better results than YOLOv4 and YOLOv4-tiny. On the other hand, we encountered various studies that showed that YOLO outperforms SSD in object detection deep learning methods. In traffic sign recognition, Zhu et al. [33] used the same data set, and the results showed that the mAP of YOLOv5 was 7.56% higher than that of SSD, and YOLOv5 was also better than SSD in terms of recognition speed. In addition, YOLOv5 was found to have better recognition accuracy than SSD [34] when detecting strawberry ripeness. In this experiment, YOLOv5 also shows better results in inference speed and mAP compared to SSD. In many studies, YOLOv5 outperforms SSD in terms of speed and accuracy [35]. In this experiment, the YOLOv5-ytiny model based on an improved YOLOv5 has advantages in both speed and mAP. Furthermore, given the previous discussion, we believe that choosing to improve YOLOv5 for aggregate identification is a wise move.

6. Conclusions

The aggregate detection classification model YOLOv5-tiny is based on the improvement of YOLOv5. In order to adapt to the complex environmental factors in the detection process, we trained four aggregates under different types of light, different wet and dry conditions, and different detection distances to achieve the real-time classification of aggregate. YOLOv5-tiny used CIoU as the loss function of the frame regression to improve the precision of the frame regression. We modified the network structure of the Backbone C3 of YOLOv5. Under the premise of ensuring the mean average precision and precision, reducing the number of YOLOv5 detection heads to simplify the model reduces the amount of calculation and improves the detection speed of the model. The experiment shows that the model storage space is reduced by 10.66 MB compared with YOLOv5, and the detection speed is 60% higher than the original YOLOv5 model.

By comparing the experimental results of the proposed YOLOv5-tiny model with the object detection networks of SSD, YOLOv4, and YOLOv4-tiny, it can be proved that the strategy proposed in this study can effectively improve the detection precision. Meanwhile, the detection speed of 22.73 FPS enables the YOLOv5-tiny model to be applied to the industrial production of real-time aggregate classification.

In the experiment, on a computer device that supports a CPU, if the YOLOv5 model is used, the model occupies a large space and the inference speed is slow. After the proposed YOLOv5-tiny is compressed, the model space is 3.04 MB, and the inference speed can reach 22.73 FPS, which can meet the actual requirements.

Author Contributions: S.Y. (Sheng Yuan), methodology; Y.D., writing—original draft preparation; M.L., supervision; S.Y. (Shuang Yue), validation; B.L., software; H.Z., resources. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Major Science and Technology Project of Henan Province, Zhengzhou major scientific and technological innovation special project, Key scientific research project plan of colleges and universities in Henan Province, grant numbers are 201110210300, 2019CXZX0050, and 21A510007. The APC was funded by Major Science and Technology Project of Henan Province, Zhengzhou major scientific and technological innovation special project, Key scientific research project plan of colleges and universities in Henan Province.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Materials used for experiments are supported by Zhengzhou Sanhe Hydraulic Machinery Co., Ltd.

Conflicts of Interest: The authors declare no conflict of interest. A role of the funders in the design of the study is to provide data in the writing of the manuscript, or in the decision to publish the results must be declared in this section.

References

1. Yilmaz, M.; Tugrul, A. The effects of different sandstone aggregates on concrete strength. *Constr. Build. Mater.* **2012**, *35*, 294–303. [[CrossRef](#)]
2. Jiangwei, B.; Wenbing, Z.; Zhenzhong, S.; Song, L.; Zhanglan, C. Analysis and optimization of mechanical properties of recycled concrete based on aggregate characteristics. *Sci. Eng. Compos. Mater.* **2021**, *28*, 516–527. [[CrossRef](#)]
3. Yan, S.; Xia, B.; Sza, B.; Zi, C.; My, D.; Cl, C. Mechanical properties, durability, and itz characteristics of full-grade dam concrete prepared by aggregates with surface rust stains. *Constr. Build. Mater.* **2021**, *305*, 124798.
4. Aksakal, E.L.; Angin, I.; Sari, S. A new approach for calculating aggregate stability: Mean weight aggregate stability (MWAS). *Catena* **2020**, *194*, 104708. [[CrossRef](#)]
5. Wang, D.; Liu, G.; Li, K.; Wang, T.; Shrestha, A.; Martek, I.; Tao, X. Layout optimization model for the production planning of precast concrete building components. *Sustainability* **2018**, *10*, 1807. [[CrossRef](#)]
6. Sicakova, A. Effect of Aggregate Size on Recycled Aggregate Concrete under Equivalent Mortar Volume Mix Design. *Appl. Sci.* **2021**, *11*, 11274.

7. Ding, X.; Ma, T.; Gao, W. Morphological characterization and mechanical analysis for coarse aggregate skeleton of asphalt mixture based on discrete-element modeling. *Constr. Build. Mater.* **2017**, *154*, 1048–1061. [[CrossRef](#)]
8. Zhan, D.A.; Pi, B.; Xu, W. Evaluation of the contact characteristics of graded aggregate using coarse aggregate composite geometric indexes—ScienceDirect. *Constr. Build. Mater.* **2020**, *247*, 118608. [[CrossRef](#)]
9. Hong, L.; Gu, X.L.; Lin, F. Effects of coarse aggregate form, angularity, and surface texture on concrete mechanical performance. *J. Mater. Civ. Eng.* **2019**, *31*, 10. [[CrossRef](#)]
10. Sadjad, N.; Mingzhong, Z. Meso-scale modelling of static and dynamic tensile fracture of concrete accounting for real-shape aggregates. *Cem. Concr. Compos.* **2021**, *116*, 103889. [[CrossRef](#)]
11. Mohd, S.S.; Roslan, Z.A.; Nor, A.Z.; Mohammad, F.; Ahmad, F.; Ahmad, J. Revisiting the automated grain sizing technique (AGS) for characterizing grain size distribution. *Int. J. River Basin Manag.* **2021**, *37*, 974–980.
12. Pei, L.; Sun, Z.; Yu, T.; Li, W.; Hao, X.; Hu, Y. Pavement aggregate shape classification based on extreme gradient boosting-scienceDirect. *Constr. Build. Mater.* **2020**, *256*, 119356. [[CrossRef](#)]
13. Park, S.S.; Lee, J.S.; Lee, D.E. Aggregate Roundness Classification Using a Wire Mesh Method. *Materials* **2020**, *13*, 3682. [[CrossRef](#)] [[PubMed](#)]
14. Mehdi, K. Assessment of strength of individual ballast aggregate by conducting point load test and establishment of classification method. *Int. J. Rock Mech. Min. Sci.* **2021**, *141*, 104711. [[CrossRef](#)]
15. Schmidt, T.; Linders, J.; Mayer, C.; Cölfen, H. Determination of Particle Size, Core and Shell Size Distributions of Core–Shell Particles by Analytical Ultracentrifugation. *Part. Part. Syst. Character.* **2021**, *38*, 2100079. [[CrossRef](#)]
16. Bagheri, G.H.; Bonadonna, C.; Manzella, I.; Vonlanthen, P. On the characterization of size and shape of irregular particles. *Powder Technol.* **2015**, *270*, 141–153. [[CrossRef](#)]
17. Agimelen, O.S.; Jawor, B.A.; McGinty, J.; Dziejewicz, J.; Tachtatzis, C.; Cleary, A.; Mulholland, A.J. Integration of in situ imaging and chord length distribution measurements for estimation of particle size and shape. *Chem. Eng. Sci.* **2016**, *144*, 87–100. [[CrossRef](#)]
18. Yang, J.H.; Fang, H.Y.; Chen, S.J. Development of particle size and shape measuring system for machine-made sand. *Part. Sci. Technol.* **2019**, *37*, 974–980. [[CrossRef](#)]
19. Isa, N.A.M.; Sani, Z.M.; AlBatah, M.S. Automated Intelligent real-time system for aggregate classification. *Int. J. Miner. Process.* **2011**, *100*, 41–50. [[CrossRef](#)]
20. Sun, Z.; Li, Y.; Pei, L.; Li, W.; Hao, X. Classification of Coarse Aggregate Particle Size Based on Deep Residual Network. *Symmetry* **2022**, *14*, 349. [[CrossRef](#)]
21. Moaveni, M.; Wang, S.; Hart, J.M.; Tutumluer, E.; Ahuja, N. Evaluation of aggregate size and shape by means of segmentation techniques and aggregate image processing algorithms. *Transp. Res. Rec.* **2013**, *2335*, 50–59. [[CrossRef](#)]
22. Sinecen, M.; Topal, A.; Makinaci, M.; Baradan, B. Neural network classification of aggregate by means of line laser based 3D acquisition. *Expert Syst.* **2013**, *30*, 333–340. [[CrossRef](#)]
23. Moon, K.H.; Falchetto, A.C.; Wistuba, M.P.; Jeong, J.H. Analyzing aggregate size distribution of asphalt mixtures using simple 2D digital image processing techniques. *Arab. J. Sci. Eng.* **2015**, *40*, 1309–1326. [[CrossRef](#)]
24. Ozturk, H.I.; Rashidzade, I. A photogrammetry based method for determination of 3D morphological indices of coarse aggregate. *Constr. Build. Mater.* **2020**, *262*, 120794. [[CrossRef](#)]
25. Pan, T.; Tutumluer, E. Evaluation of Visual Based Aggregate Shape Classifications Using the University of Illinois Aggregate Image Analyzer (UIAIA). In Proceedings of the GeoShanghai International Conference 2006, Shanghai, China, 11 May 2006; pp. 203–211. [[CrossRef](#)]
26. Wang, C.Y.; Liao, H.Y.M.; Yeh, I.H. Exploring the power of lightweight YOLOv4. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 11 October 2021; pp. 779–788.
27. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single Shot Multibox Detector. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 21–37.
28. Jocher, G.; Stoken, A.; Borovec, J.; Chaurasia, A.; Xie, T. Ultralytics/yolov5: V5. 0-YOLOv5-P6 1280 models AWS Supervise. ly and YouTube integrations. *Zenodo* **2021**, 4679653. [[CrossRef](#)]
29. Yao, J.; Qi, J.; Zhang, J.; Shao, H.; Yang, J.; Li, X. A Real-time detection algorithm for kiwifruit defects based on YOLOv5. *Electronics* **2021**, *10*, 1711. [[CrossRef](#)]
30. Yan, B.; Fan, P.; Lei, X.; Liu, Z.; Yang, F. A Real-Time Apple Targets Detection Method for Picking Robot Based on Improved YOLOv5. *Remote Sens.* **2021**, *13*, 1619. [[CrossRef](#)]
31. Yao, Y.; Jiang, Z.; Zhang, H. Ship detection in optical remote sensing images based on deep convolutional neural networks. *J. Appl. Remote Sens.* **2017**, *11*, 042611. [[CrossRef](#)]
32. Gu, Y.; Wang, S.; Yan, Y.; Tang, S.; Zhao, S. Identification and Analysis of Emergency Behavior of Cage-Reared Laying Ducks Based on YoloV5. *Agriculture* **2022**, *12*, 485. [[CrossRef](#)]
33. Zhu, Y.; Yan, W.Q. Traffic sign recognition based on deep learning. *Multimed. Tools Appl.* **2022**, *81*, 17779–17791. [[CrossRef](#)]
34. Fan, Y.; Zhang, S.; Feng, K.; Qian, K.; Wang, Y.; Qin, S. Strawberry Maturity Recognition Algorithm Combining Dark Channel Enhancement and YOLOv5. *Sensors* **2022**, *22*, 419. [[CrossRef](#)] [[PubMed](#)]
35. Lema, D.G.; Pedrayes, O.D.; Usamentiaga, R.; García, D.F.; Alonso, Á. Cost-Performance Evaluation of a Recognition Service of Livestock Activity Using Aerial Images. *Remote Sens.* **2021**, *13*, 2318. [[CrossRef](#)]

36. Jhong, S.Y.; Chen, Y.Y.; Hsia, C.H.; Lin, S.C.; Hsu, K.H.; Lai, C.F. Nighttime object detection system with lightweight deep network for internet of vehicles. *J. Real-Time Image Proc.* **2021**, *18*, 1141–1155. [[CrossRef](#)]
37. Wang, C.; Wang, H.; Yu, F.; Xia, W. A High-Precision Fast Smoky Vehicle Detection Method Based on Improved Yolov5 Network. In Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Industrial Design (AIID) IEEE, Bandung, Indonesia, 28 May 2021; pp. 255–259.
38. Wu, W.; Liu, H.; Li, L.; Long, Y.; Wang, X.; Wang, Z.; Chang, Y. Application of local fully Convolutional Neural Network combined with YOLOv5 algorithm in small target detection of remote sensing image. *PLoS ONE.* **2021**, *16*, e0259283. [[CrossRef](#)] [[PubMed](#)]
39. Song, Q.; Li, S.; Bai, Q.; Yang, J.; Zhang, X.; Li, Z.; Duan, Z. Object Detection Method for Grasping Robot Based on Improved YOLOv5. *Micromachines* **2021**, *12*, 1273. [[CrossRef](#)] [[PubMed](#)]
40. Nepal, U.; Eslamiat, H. Comparing YOLOv3, YOLOv4 and YOLOv5 for Autonomous Landing Spot Detection in Faulty UAVs. *Sensors* **2022**, *22*, 464. [[CrossRef](#)] [[PubMed](#)]
41. Sozzi, M.; Cantalamessa, S.; Cogato, A.; Kayad, A.; Marinello, F. Automatic Bunch Detection in White Grape Varieties Using YOLOv3, YOLOv4, and YOLOv5 Deep Learning Algorithms. *Agronomy* **2022**, *12*, 319. [[CrossRef](#)]