

Article

A Study on the Improvement of YOLOv5 and the Quality Detection Method for Cork Discs

Liguo Qu ^{1,2,*}, Guohao Chen ³, Ke Liu ¹ and Xin Zhang ¹

¹ School of Physics and Electronic Information, Anhui Normal University, Wuhu 241002, China; keliu@ahnu.edu.cn (K.L.); 2121011165@ahnu.edu.cn (X.Z.)

² Anhui Intelligent Robot Information Fusion and Control Engineering Research Center, Wuhu 241002, China

³ Wuhan Mingke rail Transit Equipment Co., Ltd., Wuhan 430074, China; chen2711@ahnu.edu.cn

* Correspondence: lgqu@ahnu.edu.cn

Abstract: Combining machine vision and deep learning, optical detection technology can achieve intelligent inspection. To address the issues of low efficiency and poor consistency in the quality classification of cork discs used for making badminton heads, research on optimizing the YOLOv5 image-processing algorithm was conducted and applied to cork disc quality detection. Real-time images of cork discs were captured using industrial cameras, and a dataset was independently constructed. A GAN-based defect synthesis algorithm was employed to resolve the lack of defect samples. An attention mechanism was embedded in the YOLOv5 backbone network to enhance feature representation. The number of anchors in the YOLOv5 detection layer was reduced to address similar sample sizes, a center-matching strategy was designed to balance positive samples, and a shortest-distance label assignment algorithm was developed to eliminate ambiguities, improving accuracy and reducing postprocessing complexity. Detection results were integrated into quality classification. Experiments on the NVIDIA RTX3080 GPU demonstrated that the optimized algorithm improved the original YOLOv5 F1 score by 2.4% and mF1 score by 9.0%, achieving a quality classification F1 score of 95.1%, a processing speed of 178.5 FPS, and an mAP of 81.5%. Comparative experiments showed that the improved algorithm achieved the best detection accuracy on the cork disc dataset while maintaining high processing speed.

Keywords: deep learning; YOLOv5; defect detecting; cork disc; optical detection technology



Citation: Qu, L.; Chen, G.; Liu, K.; Zhang, X. A Study on the Improvement of YOLOv5 and the Quality Detection Method for Cork Discs. *Photonics* **2024**, *11*, 825. <https://doi.org/10.3390/photonics11090825>

Received: 5 August 2024

Revised: 28 August 2024

Accepted: 29 August 2024

Published: 1 September 2024



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

1. Introduction

Natural cork, a product derived from the bark of Mediterranean oak trees, is extensively utilized due to its excellent elasticity, sealing properties, and abrasion resistance. It is widely applied in the manufacturing of premium badminton shuttlecocks and cork stoppers for sealing wine bottles, possessing significant commercial value [1]. In the production of badminton shuttlecock heads, entire sheets of cork bark are cut into small cork discs. The quality screening of these cork discs is currently one of the crucial processes in the manufacturing of badminton shuttlecock heads. However, natural cork, influenced by environmental factors during its growth, exhibits complex surface defects and varied textures, making it impossible to find two samples with identical defect patterns. Traditional manual visual inspection methods are highly subjective, influenced by factors such as physical fatigue and work experience, resulting in inconsistent sorting quality. Therefore, achieving stable quality control presents substantial challenges.

Research on the quality screening of cork discs based on optical detection technology was initiated in the 1990s. Cork images were captured using cameras to conduct automatic visual inspection of cork-related products. In 1997, Chang et al. [2] designed a feature extraction method involving morphological filtering, contour extraction, and tracking. A complex neural network was used as a classifier in the cork stopper quality classification system, achieving the classification of eight different quality grades of cork stoppers. In

2000, Gonzalez-Adrados et al. [3] proposed a cork board quality grading system, which was based on the analysis of dozens of data features from cross-sectional and tangential images of cork boards, identifying three different types of defects. Discriminant analysis was further employed for quality grading, with classification results surpassing manual classification. Costa et al. [4–6] analyzed the contribution of each porosity feature to cork stopper grading and developed a cork stopper quality classification system based on canonical discriminant analysis and stepwise discriminant analysis techniques, achieving a 14% error rate for the surface classification of seven standard commercial quality grades of cork stoppers. In 2009, Georgieva et al. [7] studied an intelligent machine vision system for the classification of seven different types of cork bricks. Feature generation was performed using Laws' masks, and the feature vectors were processed using linear discriminant analysis and principal component analysis for cork brick classification. In 2010, Paniagua et al. [8] constructed a cork stopper classification vision system that used a static threshold method to determine defect areas and morphological calculations to measure defect sizes, classifying cork stoppers using a neuro-fuzzy classifier. In 2015, Vanda Oliveira et al. [9] characterized the surface porosity of cork stoppers of three grades using image analysis methods and established a predictive classification model using stepwise discriminant analysis, achieving a classification accuracy of 75%. Other advanced techniques have also been applied to cork classification, such as neutron radiography and tomography for analyzing internal defects of cork, and volatile organic compound (VOC) analysis for natural cork stoppers with different porosity levels [10]. These methods generally involve two main steps: first, feature extraction from cork images using texture feature generation and extraction techniques, and then classification using neural networks.

In recent years, automatic optical detection technology based on deep learning has been widely and importantly applied in intelligent manufacturing for tasks such as positioning detection and surface quality inspection of components [11,12]. Examples include rubber threading line detection [13], metal surface defect detection [14–16], fabric defect detection [17], and sanitary ceramic defect detection [18], providing new solutions for cork disc quality screening. Machine vision combined with automated equipment enables automatic detection systems, which offer advantages such as accuracy, efficiency, and continuous operation. Since the introduction of AlexNet [19] in 2012, numerous excellent deep-learning algorithms have emerged, including R-CNN [20–22], ResNet [23], SSD [24], RetinaNet [25], CenterNet [26], and ConvNeXt [27]. In industrial applications, the real-time object detection algorithm YOLO (You Only Look Once) has become one of the most popular deep-learning detection frameworks. YOLO [28–30], proposed by Redmon, has since been integrated with the latest object detection algorithms by various researchers, leading to the development of efficient detection algorithms such as YOLOv4 [31], YOLOv5 [32], YOLOX [33], YOLOv6 [34], and YOLOv7 [35]. Among these, YOLOv5 has garnered extensive attention due to its optimal balance between accuracy and speed. YOLOv5 maintains high accuracy while offering rapid detection speeds, and it allows for model customization according to user-specific requirements. With a compact model size, it is particularly suited for edge computing and deployment environments with limited resources. Therefore, the application of YOLOv5 to the quality inspection of cork discs can better facilitate the automatic extraction of cork defect features and the implementation of a low-cost system. This study addresses the issues of low efficiency and poor consistency in cork disc quality screening by investigating an improved YOLOv5-based method for cork disc quality detection, achieving rapid and precise cork disc quality inspection.

2. Materials

2.1. Dataset Construction

The cork discs in the dataset were sourced from a badminton production factory. Images of each cork disc's front and back sides were captured using industrial cameras. As shown in Figure 1a, an example image of a cork disc is presented. After preliminary processing, the oak bark is segmented into cork discs with a diameter of approximately

27 mm and a height of about 5 mm. Each shuttlecock head is formed by adhering three cork discs together. The top cork disc, which requires the highest quality, has holes punched around it to insert feathers. Cork discs of slightly lower quality are placed in the middle or bottom layers.

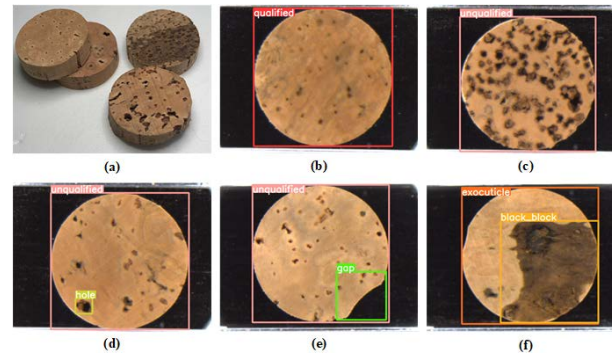


Figure 1. Cork discs and classification. (a) Example images of cork disc. (b) Qualified cork disc. (c) Unqualified cork disc with many holes. (d) Unqualified cork disc with a large hole in the edge. (e) Unqualified cork disc with a notches in the edge. (f) Unqualified cork disc with black spots.

In this study, both quality classification and defect detection tasks for cork discs are conducted simultaneously. Quality classification is divided into three categories: qualified, unqualified, and outer bark; defect detection is divided into three categories: holes, notches, and black spots. In Figure 1c, holes are annotated. Pores in the oak bark appear as black spots of varying sizes on the disc's surface, but only holes larger than 2 mm are annotated as hole defects, as small holes do not affect the shuttlecock head's quality. In Figure 1c, notches are annotated, indicating missing portions at the edge of the cork disc, primarily caused by natural factors or during cutting. In Figure 1f, black spots are annotated, originating from the oak bark's outer cuticle, which is hard, non-elastic, and of no use. Regarding these defects, cork discs can be annotated with different quality classifications. In Figure 1b, a qualified disc is annotated, indicating a surface with minimal defects, suitable for the upper layer of the shuttlecock head to insert feathers. In Figure 1c,e, unqualified discs are annotated, indicating they cannot be used for the upper layer but can be used for the middle and lower layers. In Figure 1f, the outer bark is annotated, which is of no use. Figure 1d appears to have few surface defects but is annotated as unqualified due to a hole defect located in the 1–4 mm edge region (feather insertion area). Although large holes in the center area may still be classified as qualified and used for the upper layer, the classification of hole defects requires further determination of hole size and distribution for precise quality classification.

To train the model, a total of 8570 cork images were collected. After preprocessing, each sample had a pixel size of 480×480 . The images were annotated using the Labelling tool, and the annotation results were saved in XML format. All images were randomly divided at a ratio of approximately 8:1:1, constructing an original cork disc dataset that includes 6770 training images, 900 validation images, and 900 test images.

2.2. Data Augmentation

In the cork disc dataset, the proportion of defective samples is small, resulting in an imbalance in the training data and poor generalization ability of the model. Collecting a large number of such samples is challenging; therefore, a defect synthesis algorithm based on Generative Adversarial Networks (GANs) is proposed to address the issue of insufficient defective samples. Due to the varying shapes of naturally occurring hole defects, a GAN is employed to generate new hole defects to simulate this variability. The GAN consists of

two modules: a generator and a discriminator. Its main objective is to obtain the optimal solution to the following objective function [36], as shown in Equation (1).

$$\min_G \max_D V(D, G) = E_{x \sim p_{data}(x)} [\log D(x)] + E_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

In Equation (1), x represents the training images, G represents the generator, D represents the discriminator, and z represents the generated images. During the adversarial process, the generator creates fake data samples and attempts to deceive the discriminator, while the discriminator tries to distinguish between real and fake samples. When the discriminator cannot determine the source of the images, it can be considered that the generator is capable of producing images that follow the same distribution as the training set.

Fake “holes” similar to those in the original dataset are generated using a Generative Adversarial Network (GAN). These fake holes are integrated as new defects into cork discs to create new samples. Figure 2 illustrates the proposed data augmentation (DA) algorithm process, which includes the following steps:

- (1) Selection of Background Images: “Qualified” samples from the original dataset are selected as background images. To prevent data repetition, each background image undergoes horizontal or vertical flipping.
- (2) Transformation of Defect Images: Defect images generated by the GAN are randomly selected and subjected to affine transformations, including random horizontal flipping, vertical flipping, scaling, and rotation.
- (3) Binarization and ROI Extraction: The transformed defect images are binarized using the OTUS method to obtain the Region of Interest (ROI) mask, identifying the defect ROI.
- (4) Background Region Extraction: A center point is randomly selected within a specific area of the cork disc background image (1–4 mm from the outer diameter where holes are punched for feather insertion). A background region of the same size as the transformed defect image is cropped around this center point.
- (5) Defect Integration: The defect ROI is fused with the background region and overlaid onto the cork disc background image.

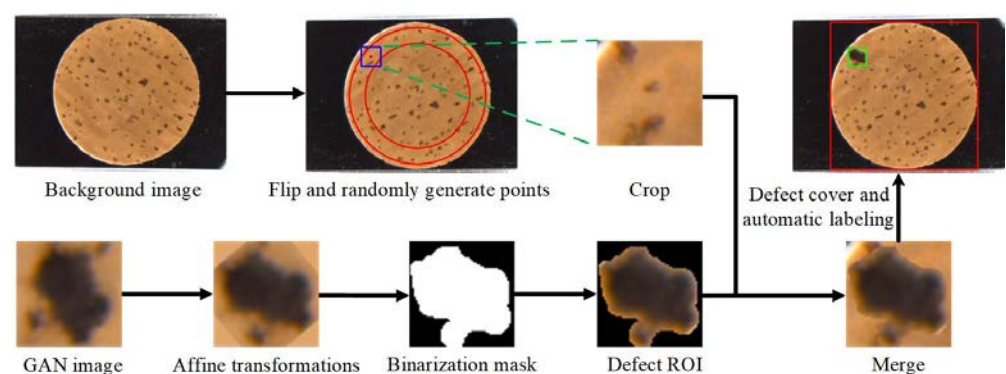


Figure 2. Sample defect synthesis algorithm based on GAN.

The GAN-based defect synthesis algorithm can generate a large number of hole defects. However, augmenting notch defects is challenging due to their occurrence only at the edges of cork discs and their directional nature; thus, random angle rotation is applied only to samples with notches. For black spot defects, due to their distinctive features and high detection accuracy, data augmentation is not performed. For hole and notch defects, 500 new samples are generated for each type. These samples are randomly divided into training, validation, and test sets at approximately an 8:1:1 ratio. After data augmentation, the dataset contains 7570 training images, 1000 validation images, and 1000 test images.

3. Quality Detection Method of Cork Discs Based on Improved YOLOv5

3.1. Overall Architecture

The overall architecture of the cork disc quality detection model is illustrated in Figure 3. The model architecture consists of three parts: the backbone network, the neck, and the detection head. It is an improvement upon YOLOv5, with the primary modifications including the integration of a Convolutional Block Attention Module (CBAM) into the backbone network to further enhance the model's feature representation capability. Given the uniform size of the cork discs, the number of prediction anchors in the detection layer has been reduced. A Center Match (CM) strategy is introduced to expand the range of positive sample selection, thereby balancing the number of positive samples. A shortest-distance label assignment (SDLA) strategy is proposed to address the issue of ambiguous sample regression. Additionally, a Detection Result Processing (DRP) algorithm is designed to further improve accuracy by aligning with the quality screening requirements for badminton cork discs.

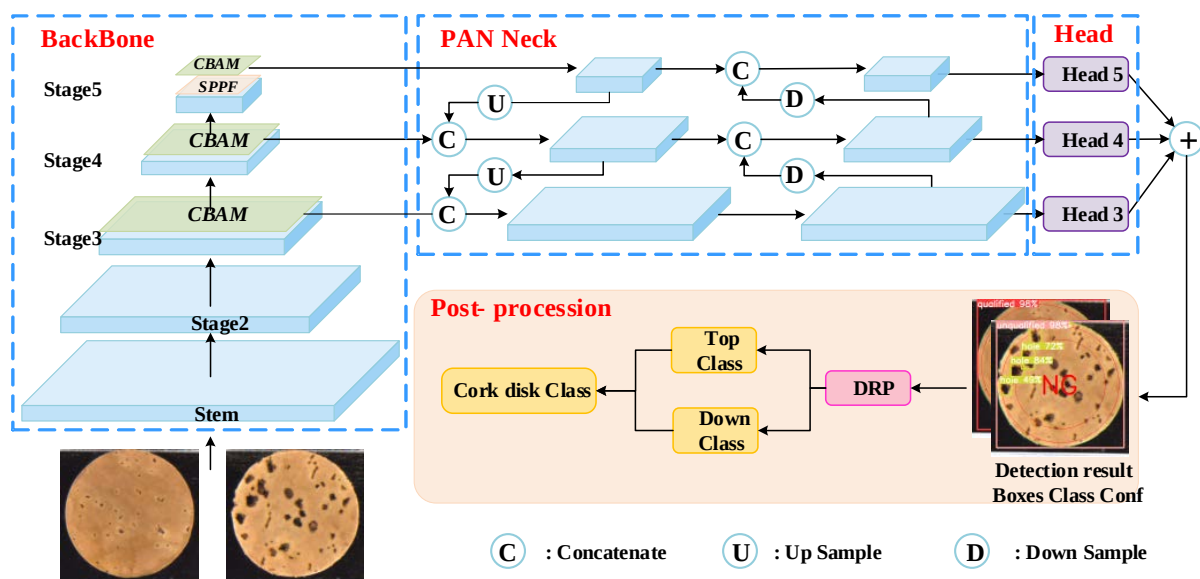


Figure 3. Architecture of the quality inspection model for cork discs.

In the backbone network section, the 6×6 convolution in the Stem module provides a larger receptive field spatially, enabling the acquisition of richer image features. Down-sampling convolutions expand the channel dimensions while reducing the feature map scale, segmenting image features into different stages. The feature extraction structures at each stage combine the advantages of cross-stage local network structures and bottleneck structures [23], reducing computational cost while enhancing the backbone's feature extraction capability. The CBAM is embedded in the backend of Stage 3, Stage 4, and Stage 5 of the backbone network, as a large amount of high-level semantic information is present in the deeper layers of the backbone network. CBAM improves the network's capability to express features.

The neck of YOLOv5 consists of an improved Path Aggregation Network (PAN) structure, which aggregates multi-scale features by connecting low-level physical features with high-level semantic features through bottom-up paths, thereby constructing pyramid feature maps and providing feature inputs for the detection head. The detection head performs predictions on the three scale feature maps generated by the neck, using a lightweight 1×1 convolutional layer. For quality detection, both sides of the cork disc are simultaneously captured and preprocessed to a size of 480×480 . The two images are stacked into a batch of 2 and input into the neural network. Subsequently, the improved YOLOv5 performs data inference and detects both images simultaneously.

3.2. CBAM

To further enhance the feature expression capability of the model, the CBAM is embedded in the backbone network based on YOLOv5. The CBAM is a lightweight convolutional attention module that operates on both channel and spatial information. It consists of two submodules: the channel attention module (CAM) and the spatial attention module (SAM), as illustrated in Figure 4. The two modules are connected in series and introduce parallel branches relative to the input feature path. This configuration allows the generation of attention feature map information sequentially along both the channel and spatial dimensions, thereby enabling adaptive feature refinement.

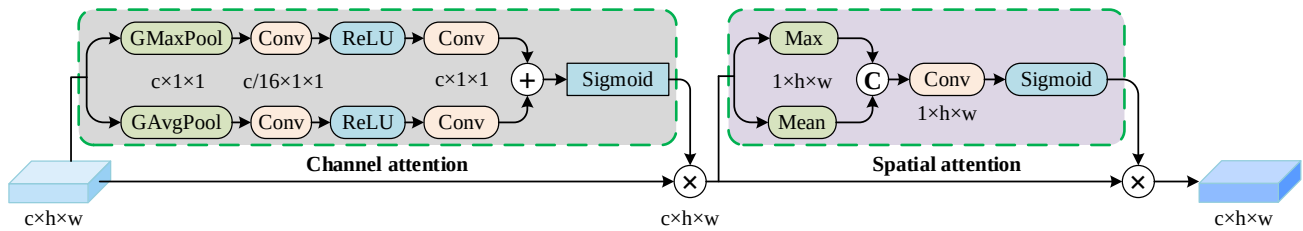


Figure 4. CBAM structure diagram.

The computation of the CBAM is described by Equation (2).

$$\begin{aligned} F' &= M_c(F) \otimes F \\ F'' &= M_s(F') \otimes F' \end{aligned} \quad (2)$$

In the formula, $F \in \mathbb{R}^{C \times H \times W}$ represents the input features, while $M_c \in \mathbb{R}^{C \times 1 \times 1}$ and $M_s \in \mathbb{R}^{1 \times H \times W}$, respectively, denote the channel attention features and spatial attention features. The symbol $\otimes$ signifies element-wise multiplication. The computation within the channel attention module is illustrated as shown in Equation (3).

$$M_c = \sigma(\text{Conv}_2(\text{ReLU}(\text{Conv}_1(F_{\max}^c))) + \text{Conv}_2(\text{ReLU}(\text{Conv}_1(F_{\text{avg}}^c)))) \quad (3)$$

In the formula, $F_{\max}^c \in \mathbb{R}^{C \times 1 \times 1}$ and $F_{\text{avg}}^c \in \mathbb{R}^{C \times 1 \times 1}$ represent the channel weights obtained through the application of global maximum pooling and global average pooling operations across all channels, respectively. σ denotes the sigmoid activation function, and $\text{Conv}_1 \in \mathbb{R}^{C \times \frac{C}{r} \times 1 \times 1}$, $\text{Conv}_2 \in \mathbb{R}^{\frac{C}{r} \times C \times 1 \times 1}$, r is a dimensionality reduction factor employed to decrease computational load, with a default setting of 16, while the minimum value of $\frac{C}{r}$ is set to 8. The computation of the spatial attention mechanism module is expressed as shown in Equation (4).

$$M_s = \sigma(\text{Conv}^{7 \times 7}(\text{Cat}(F_{\max}^s, F_{\text{avg}}^s))) \quad (4)$$

In the equation, $F_{\max}^s \in \mathbb{R}^{1 \times H \times W}$ and $F_{\text{avg}}^s \in \mathbb{R}^{1 \times H \times W}$ represent two-dimensional spatial mappings obtained from two types of pooling operations. Cat signifies the concatenation operation, and $\text{Conv}^{7 \times 7} \in \mathbb{R}^{2 \times 1 \times H \times W}$ represents a convolution operation with a 7×7 kernel.

The CBAM is embedded at the backend of Stage 3, Stage 4, and Stage 5 because deep layers of the backbone network are replete with abundant high-level semantic information. Utilizing attention modules for information fusion in these areas allows the model to differentiate features more effectively.

3.3. CM Strategy

YOLOv5 is an anchor-based detector that sets three different scales of anchors for predictions at each detection layer. While multi-scale prediction can offer performance improvements, it results in slower prediction speed and higher postprocessing complexity. Additionally, given that the target scales of cork disc objects in the dataset are nearly

uniform, the benefits of multi-scale prediction are minimal. Therefore, the number of anchors is set to one to reduce the number of predictions and enhance processing speed. The reduction in the number of anchors leads to a decrease in the number of positive samples used for training. To address this, a CM strategy is employed to expand the range of positive sample selection and balance the number of positive samples.

As shown in Figure 5a, the positive sample selection strategy of YOLOv5 matches predictions from three grids with the ground truth targets. During the selection process, when the center of a ground truth target falls within a particular grid, that grid is chosen. When the number of anchors is reduced to one, the number of selected positive samples decreases from nine to three, resulting in a significant reduction in the number of positive samples, which in turn lowers the algorithm's convergence rate and accuracy. To balance the number of positive samples, a Center Match (CM) strategy is proposed: predictions from the grid containing the center of the ground truth target and its eight neighboring grids are selected as positive samples, as shown in Figure 5b. After replacing the original positive sample selection strategy with the Center Match strategy, the transformation formula for the prediction results relative to the ground truth bounding box center coordinates changes. The calculations for YOLOv5 with the Center Match strategy are given by Formula (5).

$$\begin{aligned} B_{x,y} &= (2 \times \sigma(P_{x,y}) - 0.5) + G_{x,y} \quad \text{for YOLOv5} \\ B_{x,y} &= (3 \times \sigma(P_{x,y}) - 1.0) + G_{x,y} \quad \text{for CM} \end{aligned} \quad (5)$$

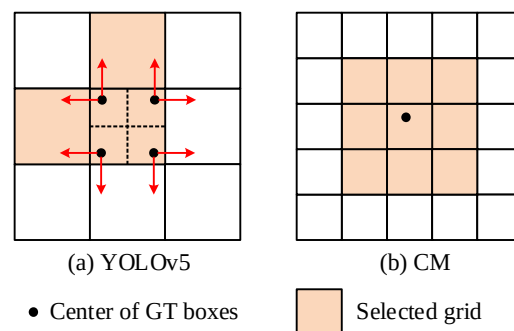


Figure 5. Schematic diagram of YOLOv5 and CM positive sample selection strategy. The grid is divided into four regions by using dotted cross lines. If the center point is predicted within a certain region, the two adjacent grids are selected according to the direction of the arrows.

In the formula, σ represents the sigmoid function, $\sigma(P_{x,y})$ denotes the predicted offsets for x and y , and $G_{x,y}$ signifies the coordinates of the top-left corner of the grid point in the x - y plane.

Compared to the original method of selecting positive samples, the CM strategy maintains the same quantity of positive samples while reducing the number of model prediction outputs, enhancing the model's prediction speed and reducing postprocessing complexity. Moreover, the positive samples involved in loss calculation come from the target's central area, where prediction results are generally of higher quality. Calculating loss with higher-quality prediction results enriches the gradient information.

3.4. SDLA Strategy

When selecting positive samples, sufficient proximity between targets can lead to a challenging ambiguity: predictions from certain grids may be selected as positive samples for multiple ground truth targets, resulting in confusion about which target the prediction should be regressed to. Such overlapping samples are referred to as ambiguous samples, as illustrated by the green striped grids in Figure 6.

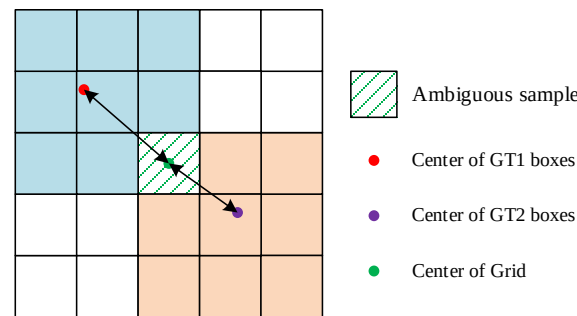


Figure 6. Schematic diagram of the SDLA strategy. The line with the arrow is the Euclidean distance from the anchor center to the truth targets.

YOLOv5 does not address such cases. It is noted that selecting positive samples within a smaller spatial range can significantly reduce the occurrence of ambiguous samples. The CM strategy, however, expands the positive sample selection space, increasing the likelihood of ambiguous samples, which impacts accuracy to some extent. Therefore, a simple and effective SDLA strategy is proposed to eliminate this ambiguity. The process is detailed in Figure 6. When an ambiguous sample is assigned to two ground truth targets, GT1 and GT2, the center point of the grid cell is first calculated, equal to the top-left corner coordinates plus 0.5 times the downsampling stride. The Euclidean distance from this center point to the centers of the two ground truth targets is then computed, and the target with the shortest distance is chosen as the regression target. As shown in Figure 6, the ambiguous sample is ultimately assigned to GT2. The SDLA strategy can effectively and simply eliminate erroneous gradient information during backpropagation.

3.5. DRP Algorithm

In the cork disc quality detection results, predictions for multiple categories are included. However, only one classification prediction for the cork disc is desired. Therefore, during non-maximum suppression of redundant detection boxes, the predictions for the three quality categories of the cork discs are processed together, retaining only the category with the highest confidence score. This approach suppresses multiple prediction categories and reduces the complexity of subsequent processing. However, for corks similar to those shown in Figure 1d, where the probability of model inference errors is higher, the Detection Result Processing (DRP) algorithm is employed to correct the final output category labels based on the relationship between cork disc categories and surface defects, thereby further enhancing classification accuracy.

The logic of the Detection Result Processing (DRP) algorithm is illustrated in Figure 7. Based on quality classification, the results are further refined by incorporating defect detection outcomes. The “black spot” defect is considered the most severe; cork discs with this defect are classified as the “outer skin” category. Therefore, when the category is “outer skin”, the result is directly output without any further processing. For other categories, it is necessary to check for the presence of the “black spot” defect. If it is present, the output category will be changed to “outer skin”.

A more complex scenario arises when the cork disc is classified as “qualified” but contains defects. In this case, the following steps are sequentially assessed: First, determine if the “black spot” defect is present. If it is, the category will be modified to “outer skin”. If not, check for the presence of “notch” defects. If present, the category is changed to “unqualified”.

Finally, for hole defects, classification will be based on three criteria: (1) whether the holes are located within a specific area, (2) whether the largest hole is greater than 4 mm in size, and (3) whether the total number of holes exceeds 3. If any of these conditions are met, the category will be updated to “unqualified”.

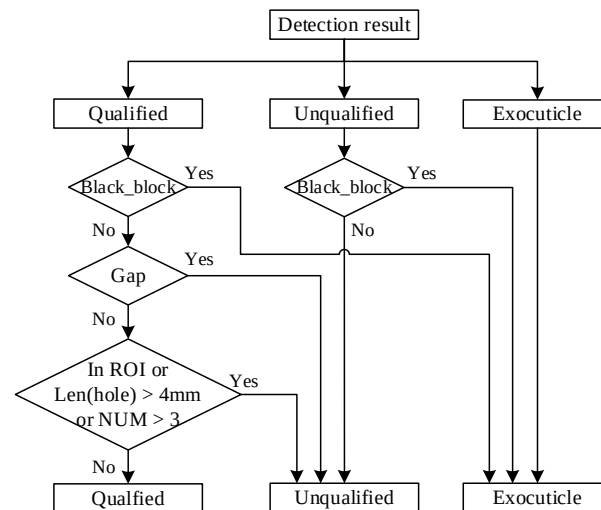


Figure 7. DRP logic diagram.

4. Experiment and Result Analysis

4.1. Experimental Environment and Evaluation Index

The hardware configuration for model training and testing includes an Intel Core™ i7-11700K processor and an NVIDIA RTX3080 GPU with 10 GB of memory. The above hardware platform comes from Lenovo Group Co., Ltd. in Beijing, China. The computer's operating system is Ubuntu 18.04, and the software environment is configured with Python 3.8, Pytorch 1.8, Opencv 4.1.2, and CUDA v11.1.

During model training, the number of training epochs is set to 300, with a batch size of 32. K-means clustering is used to recalculate anchors suitable for the cork disc dataset. For the original YOLOv5 model, anchors are set as follows: [[49, 49], [61, 76], [114, 46], [66, 148], [154, 98], [180, 205], [332, 226], [229, 328], [443, 441]]. After applying the CM strategy, anchors are set as follows: [[66, 66], [240, 237], [442, 441]]. Other hyperparameters are evolved using a genetic algorithm (GA) and selected as follows: the initial learning rate is 0.01, the momentum is 0.946, the weight decay is 0.00047, the bounding box regression loss gain is 0.053, the classification loss gain is 0.86, the classification binary cross-entropy loss positive weight is 0.82, the objectness loss gain is 0.566, and the objectness binary cross-entropy loss weight is 0.949.

For model evaluation, the confidence threshold is set to 0.45 and the IoU threshold to 0.7. The evaluation metric chosen for the detection model is the *F1* score, as shown in Equation (6), because it considers both the *precision* and *recall* of the classification model, providing a more accurate reflection of the model's classification capability.

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (6)$$

It is noteworthy that the model task includes both quality classification and defect detection. Therefore, in addition to using the mean *F1* score (mF1) metric, we also consider the classification detection *F1* score (CDF1) and the defect detection *F1* score (Defect_F1). The CDF1 score determines the accuracy of the quality classification, while the defect detection results are used by the DRP algorithm to correct the quality classification.

4.2. Ablation Experiment

To verify the effectiveness of the algorithm improvements, ablation experiments were conducted on different modifications made to the YOLOv5 base network.

4.2.1. Effects of Model Size and Pre-Training Weights

YOLOv5 provides models of various scales along with corresponding pre-trained weights to meet the needs of different application scenarios. Generally, larger models offer better performance but slower inference speeds. To determine the most suitable model scale, experiments evaluated three models (n, s, and m) on the cork disc dataset with and without pre-trained weights, assessing mF1 scores and inference speeds. The results are shown in Figure 8. In Figure 8, the blue solid line and red dashed line represent the detection results of models with and without pre-trained weights, respectively. It is evident that using pre-trained weights improves detection performance across various model scales, as it better initializes model parameters. Furthermore, while larger models provide limited improvement in mF1 scores, inference time increases rapidly. Among them, the YOLOv5s model, with an inference speed of 4.5 ms and a detection mF1 score of 85.7%, achieves the best balance between speed and accuracy. Therefore, the YOLOv5s model is selected as the baseline model, and pre-trained weights are used to initialize parameters in subsequent training.

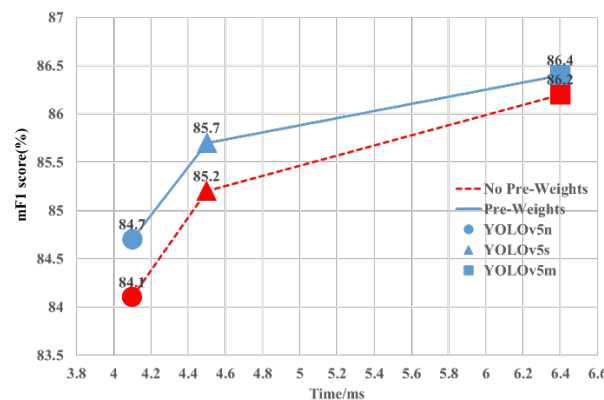


Figure 8. Detection speed and mF1 score of different YOLOv5 models. The circle represents YOLOv5n, the triangle represents YOLOv5s, and the square represents YOLOv5m, regardless of color.

4.2.2. Impact of Adding Defect Training

The core task of model inference is the quality classification of cork discs. The model was designed as both a dedicated quality classification model and a unified model (performing both quality classification and defect detection). In the dataset with data augmentation, the dedicated quality classification model achieved a CDF1 score of 91.1%, whereas the unified model achieved a CDF1 score of 93.8%, representing an improvement of 2.7%. This significant improvement is attributed to the inherent relationship between cork disc categories and defects. The neural network, through extensive data learning, gradually recognizes the relationship between them and considers the spatial activation features of defects when making classification predictions. Further explanation using the visual feature maps shown in Figure 9 reveals that these feature maps are derived from the Head 5 layer of the detection model. In Figure 9, the detection model without defect targets can only learn global features, while the feature maps with defect detection exhibit strong spatial feature activation at defect targets (e.g., holes in the lower-left corner), leading to more accurate classification judgments.

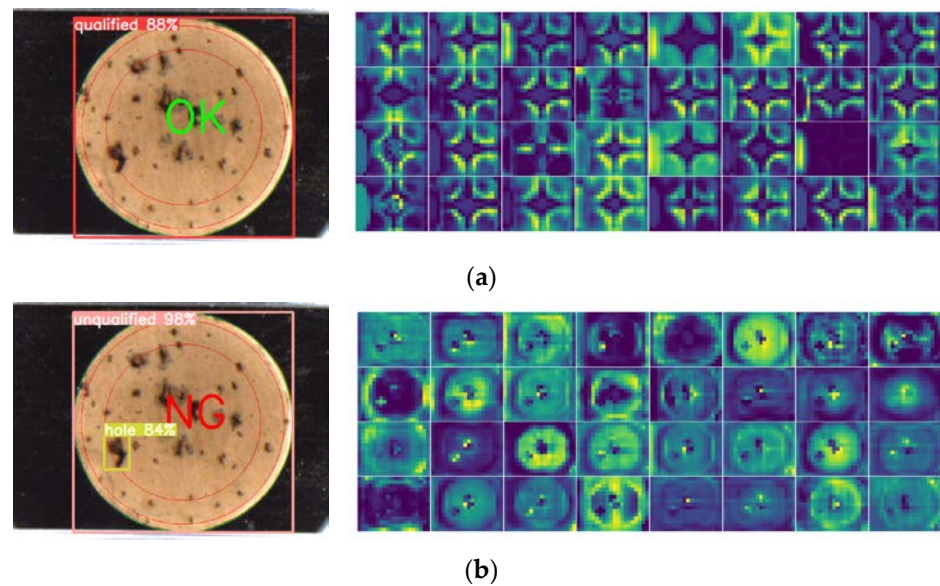


Figure 9. Removal and addition of visual feature maps of defect target models. (a) The model output feature maps with removed defect sample training. (b) The model output feature maps with added defect sample training. The area between the two red circles represents the 1-4mm region of interest.

4.2.3. Impact of DA and DRP

The amount of training data directly affects the final performance of the detection model. Data augmentation methods were employed to add 500 “hole” and “gap” defects to the dataset to mitigate the issue of imbalanced defect samples.

As shown in Table 1, experimental results indicate that when the number of defect samples is low, the Defect_F1 score for the three types of defects is only 62.7%. After using DA to increase the number of defect samples, the Defect_F1 score significantly improved by 14.1%, making the model’s defect detection more robust. Additionally, the CDF1 score also increased slightly by 1.1%. This improvement is partly due to the model’s enhanced ability to extract defect features, which ultimately impacts the overall classification judgment. Although the unified model improves generalization and efficiency through shared lower-level features and joint training, this self-learned cognition is not perfect, and errors may still occur in some cases. Therefore, using the DRP algorithm to correct the quality classification based on defect detection results further enhances classification accuracy, with the CDF1 score increasing by 0.7% from 93.8%.

Table 1. The effect of DA data enhancement was compared.

Model	CDF1	Defect_F1	mF1
YOLOv5	92.7%	62.7%	77.7%
YOLOv5 + DA	93.8%	76.8%	85.3%
YOLOv5 + DA + DRP	94.5%	76.8%	85.7%

4.2.4. Impact of CM and SDLA

The performance of both the CM and SDLA optimization methods was evaluated in the experiment. These methods are “bag-of-freebies” techniques focused on optimizing the training process [31], aimed at improving object detection accuracy by increasing training costs without adding to the inference cost.

Table 2 compares the impact of the two methods on model accuracy and training time. Table 3 analyzes the proportion of ambiguous samples among all positive samples when different positive sample selection strategies are applied, with data from YOLOv5 tested using an anchor count of one.

Table 2. The impact of CM and SDLA strategies on accuracy and training time.

Model	CDF1	mF1	Train (Time/h)
YOLOv5 + DA + DRP	94.5%	85.7%	5.61
YOLOv5 + DA + DRP + CM	94.6%	86.4%	5.61
YOLOv5 + DA + DRP + CM + SDLA	94.9%	86.4%	6.50

Table 3. The proportion of fuzzy samples to all positive samples under different methods.

Method	Amb Samp	Pos Samp
YOLOv5	599	142,875
CM	3824	428,625
CM + SDLA	0	428,625

“Amb samp” denotes ambiguous samples, and “Pos samp” denotes positive samples.

Table 2 shows that the CM strategy slightly increased the CDF1 score by 0.1% and the mF1 score by 0.7% without any additional training cost. However, as indicated in Table 3, the CM strategy expanded the positive sample selection range, causing the number of ambiguous samples in the training process to increase from 599 to 3824, effectively doubling the probability of ambiguous samples, which destabilized the training process.

The SDLA strategy, which applies a simple shortest spatial distance principle to handle ambiguous samples, reduced the number of ambiguous samples to zero, thereby eliminating erroneous gradient information during training. This led to a 0.3% increase in the CDF1 score. However, the SDLA strategy requires processing on individual images, extending the training time by 0.89 h. Additionally, repeated training of models using the SDLA strategy yields fully consistent training results. This indicates that the SDLA strategy thoroughly eliminates ambiguous samples, ensuring complete consistency in data gradient information under the same training images and parameters.

4.2.5. Impact of CBAM

The experimental results of inserting attention modules at different positions in the YOLOv5 + DA + DRP + CM + SDLA model backbone network are shown in Table 4. CBAM_n represents the placement of *n* CBAMs in the backbone network following a top-down order (Stage 5, Stage 4, Stage 3, Stage 2). Latency refers to the inference delay caused by CBAM_n, excluding preprocessing and postprocessing times.

Table 4. Impact of CBAM on performance.

CBAM	Latency	CDF1	mF1
CBAM × 1	5.0 ms	94.1%	86.3%
CBAM × 2	5.3 ms	93.3%	84.9%
CBAM × 3	5.6 ms	95.1%	86.7%
CBAM × 4	5.9 ms	94.4%	86.6%

The experiment demonstrates that as the number of inserted attention modules increases, the model’s inference delay also increases, with each additional CBAM contributing an extra 0.3 ms to the inference delay. According to the results in Table 4, embedding three CBAM attention modules at the end of Stage 5, Stage 4, and Stage 3 achieves the best balance, with an inference time of 5.6 ms and an mF1 score of 86.7%.

4.2.6. Accuracy Analysis

Table 5 summarizes the impact of different optimization methods on detection accuracy. Sequential application of DA, the DRP algorithm, the CM strategy, the SDLA strategy, and the CBAM attention mechanism to the YOLOv5s baseline model led to incremental

improvements in model performance. On the cork disc dataset, a total improvement of 2.4% in CDF1 score and 9.0% in mF1 score was achieved, resulting in a final optimized detection algorithm with a CDF1 score of 95.1% and an mF1 score of 86.7%.

Table 5. The impact of different optimization methods on model precision.

Model	CDF1	mF1
(1) YOLOv5	92.7%	77.7%
(2) YOLOv5 + DA	93.8%	85.3%
(3) YOLOv5 + DA + DRP	94.5%	85.7%
(4) YOLOv5 + DA + DRP + CM	94.6%	86.4%
(5) YOLOv5 + DA + DRP + CM + SDLA	94.9%	86.4%
(6) YOLOv5 + DA + DRP + CM + SDLA + CBAM	95.1%	86.7%

Additionally, further analysis revealed that the DRP algorithm effectively improved classification accuracy without affecting inference time, though it did increase postprocessing time by an average of 1.5 ms per detection. The CM strategy reduced inference time by 2.3 ms and postprocessing time by 1.1 ms, as fewer predictions resulted in faster prediction speeds and lower postprocessing complexity. The SDLA strategy enhanced the model's classification accuracy without affecting inference speed, as it only increased training costs. The CBAM improved detection accuracy but increased inference time due to the additional parameters and computational load introduced to the model backbone network. Despite the increased inference time, CBAM's contribution to accuracy improvement was more significant.

The results of the optimized model (Model 6 in Table 5) compared to the original YOLOv5 model are shown in Figure 10. For five typical samples selected from the actual test results, the improved model demonstrated enhanced detection accuracy, particularly for samples 3, 4, and 5, where the original YOLOv5 model made incorrect quality classifications. The improved YOLOv5 model correctly performed quality classification and accurately identified defect features.

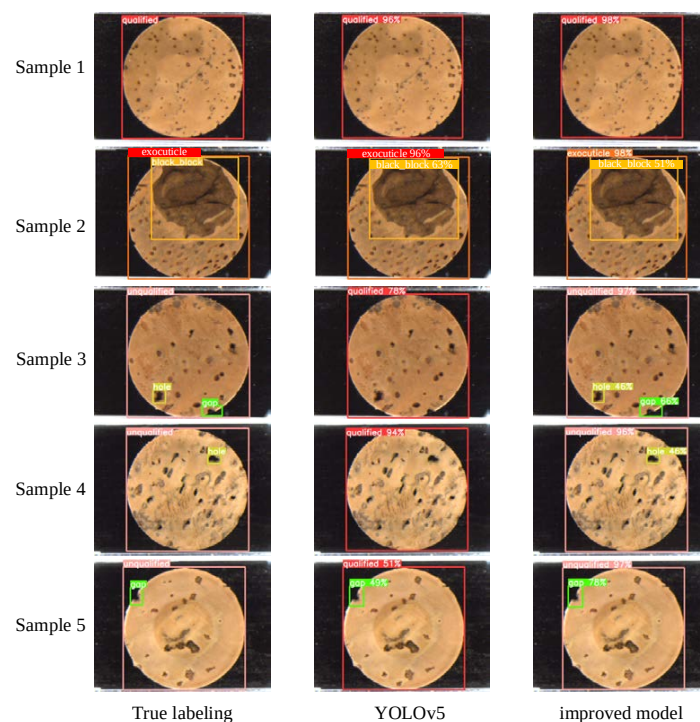


Figure 10. Comparison of detection results between YOLOv5 and the improved model.

4.3. Comparison Experiments

To validate the advancement of the improved algorithm, a comparative experiment was conducted with six popular object detection network models: Faster RCNN, RetinaNet, CenterNet, YOLOX, YOLOv4, and YOLOv7. Identical training parameters were set for all models, and their performance was tested on the same cork disc test dataset, as shown in Table 6. The single-stage detector YOLO models were able to process images at speeds exceeding 100 FPS and achieved an mF1 score of no less than 80% on the cork disc test dataset. In contrast, the classic models Faster RCNN and RetinaNet exhibited lower performance in both accuracy and speed, while the keypoint-based CenterNet demonstrated faster inference speed but lower detection accuracy.

Table 6. Comparison of different object detection models.

Model	Platform	Backbone	Type	mF1	mAP0.5	mAP	FPS
Faster RCNN	MMDetection	ResNet50	Anchor-based	77.2%	79.9%	71.9%	57.7
RetinaNet	MMDetection	ResNet18	Anchor-based	78.6%	78.6%	70.9%	88.6
CenterNet	MMDetection	ResNet18	Keypoint-based	74.9%	78.2%	69.9%	194.3
YOLOX	MMDetection	Darknet	Anchor-free	85.2%	79.7%	72.6%	102.8
YOLOv4	Darknet	Darknet	Anchor-based	75.0%	82.2%	58.7%	326.7
YOLOv7	YOLOv7	E-ELAN	Anchor-based	82.9%	77.4%	72.0%	227.2
Ours	YOLOv5	Darknet	Anchor-based	86.7%	89.4%	81.5%	178.5

In the YOLO series models, YOLOv7 claims to surpass YOLOv5 in performance. However, the results in Table 6 demonstrate that our improved YOLOv5 achieves the best balance between accuracy and speed, with an mF1 score of 86.7% and an mAP of 81.5%, surpassing all other detection models listed in terms of cork disc classification accuracy. Additionally, it processes at a speed of 178.5 FPS, which, while slightly lower than YOLOv7, still meets practical requirements.

5. Conclusions

Through in-depth research and experimental validation, a method for cork disc quality detection based on the improved YOLOv5 has been successfully developed. A proprietary cork disc dataset was constructed, including 8570 original images and 1000 augmented images, created using data augmentation algorithms. Tailored to the characteristics of cork disc detection and combined with the YOLOv5 model, a series of deep-learning optimization methods were proposed to enhance detection efficiency and accuracy. These include expanding the dataset using data augmentation, incorporating attention mechanism modules to strengthen feature representation, designing a center-matching strategy to balance the number of positive samples, designing a shortest-distance label assignment strategy to eliminate ambiguous samples, and further improving detection accuracy through result-processing algorithms.

Finally, ablation and comparative experiments were conducted on an NVIDIA RTX 3080 GPU platform. The ablation experiments demonstrated that the proposed optimization methods effectively improved model detection accuracy. The optimized model achieved a 2.4% increase in CDF1 score and a 9.0% increase in mF1 score compared to the original YOLOv5 model on the cork disc dataset. The final optimized model reached a CDF1 score of 95.1%, a processing speed of 178.5 FPS, and an mAP of 81.5%. Compared to mainstream algorithms like Faster RCNN, RetinaNet, and CenterNet, the improved algorithm achieved the best detection performance on the cork disc dataset while maintaining high processing speed. Future work will deploy this algorithm on embedded development platforms and explore its application in other optical inspection fields.

Author Contributions: Conceptualization, L.Q.; methodology, G.C. and L.Q.; software, X.Z.; formal analysis, K.L.; investigation, K.L.; data curation, G.C. and X.Z.; writing—original draft preparation,

G.C.; writing—review and editing, L.Q.; funding acquisition, L.Q. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Wuhu Science and Technology Project (No. 2022jc07) and the Graduate Student Innovation and Entrepreneurship Practice Project of Anhui Provincial Department of Education (No. 2022xcxsj058, 2023xcxsj043).

Institutional Review Board Statement: The study did not require ethical approval.

Informed Consent Statement: The study did not involve humans.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: Author Guohao Chen was employed by the Wuhan Mingke rail Transit Equipment Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Díaz-Maroto, M.C.; López-Viñas, M.; Loarce, L.; Sanza, M.d.; Nevares, I.; Alañón, M.E.; Pérez-Coello, M.S. Quality control of natural cork stoppers by image analysis and oxygen transmission rate. *Holzforschung* **2022**, *76*, 863–873. [\[CrossRef\]](#)
2. Chang, S.H.; Han, G.H.; Valverde, J.M.; Griswold, N.C.; Duque-Carrillo, J.F.; Sanchez-Sinencio, E. Cork quality classification system using a unified image processing and fuzzy-neural network methodology. *IEEE Trans. Neural Netw.* **1997**, *8*, 964–974. [\[PubMed\]](#)
3. Gonzalez-Adrados, J.R.; Lopes, F.; Pereira, H. Quality grading of cork planks with classification models based on defect characterization. *Holz als Roh-und Werkstoff* **2000**, *58*, 39–45. [\[CrossRef\]](#)
4. Costa, A.; Pereira, H. Quality characterization of wine cork stoppers using computer vision. *J. Int. Sci. Vigne Vin.* **2005**, *39*, 209–218. [\[CrossRef\]](#)
5. Costa, A.; Pereira, H. Decision rules for computer-vision quality classification of wine natural cork stoppers. *Am. J. Enol. Vitic.* **2006**, *57*, 210–219. [\[CrossRef\]](#)
6. Costa, A.; Pereira, H. Computer vision applied to cork stoppers inspection. In *Cork Oak Woodlands and Cork Industry: Present, Past and Future*; Zapata, S., Ed.; Museu del Surode Palafrugell Publications: Barcelona, Spain, 2009.
7. Georgieva, A.; Jordanov, I. Intelligent visual recognition and classification of cork tiles with neural networks. *IEEE Trans. Neural Netw.* **2009**, *20*, 675–685. [\[CrossRef\]](#)
8. Paniagua, B.; Vega-Rodríguez, M.A.; Gomez-Pulido, J.A.; Sanchez-Perez, J.M. Improving the industrial classification of cork stoppers by using image processing and Neuro-Fuzzy computing. *J. Intell. Manuf.* **2010**, *21*, 745–760. [\[CrossRef\]](#)
9. Oliveira, V.; Knapic, S.; Pereira, H. Classification modeling based on surface porosity for the grading of natural cork stoppers for quality wines. *Food Bioprod. Process.* **2015**, *93*, 69–76. [\[CrossRef\]](#)
10. Furtado, I.; Oliveira, A.S.; Amaro, F.; Lopes, P.; Cabral, M.; Bastos, M.d.L.; de Pinho, P.G.; Pinto, J. Volatile profile of cork as a tool for classification of natural cork stoppers. *Talanta* **2021**, *223*, 121698. [\[CrossRef\]](#) [\[PubMed\]](#)
11. Tang, H.; Zhu, H.; Fei, L.; Wang, T.; Cao, Y.; Xie, C. Low-Illumination Image Enhancement Based on Deep Learning Techniques: A Brief Review. *Photonics* **2023**, *10*, 198. [\[CrossRef\]](#)
12. Guan, J.; Li, J.; Yang, X.; Chen, X.; Xi, J. Defect detection method for specular surfaces based on deflectometry and deep learning. *Opt. Eng.* **2022**, *61*, 061407. [\[CrossRef\]](#)
13. Wongtanawijit, R.; Khaorapapong, T. Rubber tapping line detection in near-range images via customized YOLO and U-Net branches with parallel aggregation heads convolutional neural network. *Neural Comput. Appl.* **2022**, *34*, 20611–20627. [\[CrossRef\]](#)
14. He, Y.; Song, K.; Meng, Q.; Yan, Y. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 1493–1504. [\[CrossRef\]](#)
15. Sharma, M.; Lim, J.; Lee, H. The amalgamation of the object detection and semantic segmentation for steel surface defect detection. *Applied. Sci.* **2022**, *12*, 6004. [\[CrossRef\]](#)
16. Wang, W.Y.; Mi, C.F.; Wu, Z.H.; Lu, K.; Pan, B.; Li, D.; Zhang, J.; Chen, P.; Wang, B. A real-time steel surface defect detection approach with high accuracy. *IEEE Trans. Instrum. Meas.* **2022**, *71*, 1–10. [\[CrossRef\]](#)
17. Mei, S.; Wang, Y.; Wen, G. Automatic fabric defect detection with a multi-scale convolutional denoising autoencoder network model. *Sensors* **2018**, *18*, 1064. [\[CrossRef\]](#)
18. Ren, X.Y.; Lin, W.Y.; Yang, X.Q.; Yu, X.; Gao, H. Data augmentation in defect detection of sanitary ceramics in small and non-i.i.d datasets. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *34*, 8669–8678. [\[CrossRef\]](#)
19. Krizhevsky, A.; Sutskever, I.; Hinton, G. Imagenet classification with deep convolutional neural networks. *Commun. ACM* **2017**, *60*, 84–90. [\[CrossRef\]](#)
20. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. *IEEE Comput. Soc.* **2014**, 580–587.
21. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.

22. Ren, S.Q.; He, K.M.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 1137–1149. [[CrossRef](#)]
23. He, K.M.; Zhang, X.Y.; Ren, S.Q.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
24. 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 (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
25. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.M.; Dollar, P. Focal loss for dense object detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)]
26. Duank, W.; Bai, S.; Xie, L.; Qi, H.; Huang, Q.; Tian, Q. CenterNet++ for Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2024**, *46*, 3509–3521. [[CrossRef](#)]
27. Liu, Z.; Mao, H.Z.; Wu, C.Y. A ConvNet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.
28. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 5571–5579.
29. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
30. Edmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
31. Bochkovskiy, A.; Wang, C.Y.; Liao, H.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
32. Maktoof, M.A.J.; Al Attar, I.T.A.; Ibraheem, I.N. Comparison YOLOv5 Family for Human Crowd Detection. *Int. J. Online Biomed. Eng.* **2023**, *19*, 94–108. [[CrossRef](#)]
33. Ge, Z.; Liu, S.T.; Wang, F.; Li, Z.; Sun, J. YOLOX: Exceeding YOLO series in 2021. *arXiv* **2021**, arXiv:2107.08430.
34. Li, C.Y.; Li, L.L.; Jiang, H.L.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications. *arXiv* **2022**, arXiv:2209.02976.
35. Wang, C.Y.; Bochkovskiy, A.; Liao, H.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv* **2022**, arXiv:2207.02696.
36. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Commun. ACM* **2020**, *63*, 139–144. [[CrossRef](#)]

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