

Article

DGS-YOLOv8: A Method for Ginseng Appearance Quality Detection

Lijuan Zhang ^{1,2} , Haohai You ¹, Zhanchen Wei ¹, Zhiyi Li ³, Haojie Jia ¹, Shengpeng Yu ¹, Chunxi Zhao ⁴, Yan Lv ^{1,*}  and Dongming Li ^{1,2} 

¹ College of Information Technology, Jilin Agricultural University, Changchun 130118, China; zhanglijuan@cwzu.edu.cn (L.Z.); youhaohai@126.com (H.Y.); weizhanchen@126.com (Z.W.); jhj483790074@163.com (H.J.); shengpengyu@126.com (S.Y.); ldm0214@163.com (D.L.)

² College of Internet of Things Engineering, Wuxi University, Wuxi 214105, China

³ College of Instrument Science & Electrical Engineering, Jilin University, Changchun 130012, China; zyli6522@mails.jlu.edu.cn

⁴ Information Center, Jilin Agricultural University, Changchun 130118, China; zcx@jlau.edu.cn

* Correspondence: lvyanjlu@163.com

Abstract: In recent years, the research and application of ginseng, a famous and valuable medicinal herb, has received extensive attention at home and abroad. However, with the gradual increase in the demand for ginseng, discrepancies are inevitable when using the traditional manual method for grading the appearance and quality of ginseng. Addressing these challenges was the primary focus of this study. This study obtained a batch of ginseng samples and enhanced the dataset by data augmentation, based on which we refined the YOLOv8 network in three key dimensions: firstly, we used the C2f-DCNv2 module and the SimAM attention mechanism to augment the model's effectiveness in recognizing ginseng appearance features, followed by the use of the Slim-Neck combination (GSConv + VoVGSCSP) to lighten the model. These improvements constitute our proposed DGS-YOLOv8 model, which achieved an impressive mAP50 of 95.3% for ginseng appearance quality detection. The improved model not only has a reduced number of parameters and smaller size but also improves 6.86%, 2.73%, and 3.82% in precision, mAP50, and mAP50-95 over the YOLOv8n model, which comprehensively outperforms the other related models. With its potential demonstrated in this experiment, this technology can be deployed in large-scale production lines to benefit the food and traditional Chinese medicine industries. In summary, the DGS-YOLOv8 model has the advantages of high detection accuracy, small model space occupation, easy deployment, and robustness.

Keywords: ginseng appearance quality; machine vision; image analysis; attention mechanism; deep learning



Citation: Zhang, L.; You, H.; Wei, Z.; Li, Z.; Jia, H.; Yu, S.; Zhao, C.; Lv, Y.; Li, D. DGS-YOLOv8: A Method for Ginseng Appearance Quality Detection. *Agriculture* **2024**, *14*, 1353. <https://doi.org/10.3390/agriculture14081353>

Academic Editors: Weizhen Liang and Jingqiu Chen

Received: 29 June 2024

Revised: 8 August 2024

Accepted: 12 August 2024

Published: 13 August 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

Ginseng (*Panax ginseng* C.A. Meyer), a highly valued Chinese medicinal ingredient with a history spanning over 2000 years, has been deeply rooted in the healthcare traditions of Asian countries. Its therapeutic properties are documented in the well-known herbal compendium Shennong's Classic of the Materia Medica [1]. Ginseng boasts many medicinal benefits, including nourishing the five principal organs, soothing the spirit, calming the soul, alleviating palpitations, expelling malevolent influences, enhancing vision, promoting happiness, bolstering intellectual insight, and extending longevity [2]. The therapeutic efficacy of ginseng is primarily attributed to its saponins, a group of naturally occurring active compounds with diverse biological and pharmacological effects. Ginsenosides, which are structurally varied, predominantly include 20(S)-ginsenosides Rg1, Rb1, and Rc, offering extensive health benefits to humans [3–5]. Optimal growth conditions for ginseng are found under the canopies of deciduous broad-leaved forests or mixed coniferous broad-leaved

forests at several hundred-meter elevations. Ginseng from Jilin Province is renowned for its superior quality. As of 2023, Jilin Province has dedicated substantial acreage to cultivating garden and forest ginseng, resulting in significant yields of fresh ginseng. White ginseng, identified by its slightly off-white to yellowish hue, is produced from fresh ginseng that has been naturally dried after four to six years of growth. Most white ginseng undergoes further processing through slicing [6]. The classification of ginseng relies heavily on the extraction of color and textural features, which are inherently complex and challenging to differentiate. The shape of the root remains a crucial factor in determining ginseng quality [7]. Traditionally, ginseng appearance and quality classification have been conducted manually. Despite its historical significance, this method has several drawbacks, including subjectivity and susceptibility to human error [8]. Manual classification is prone to personal biases, incurs significant labor costs, and is ill-suited for handling large-scale samples and frequent classifications. This approach is inefficient and error-prone when dealing with vast amounts of data, highlighting the need for more advanced and reliable techniques for ginseng classification. Furthermore, ginseng's delicate and brittle nature makes manual identification prone to damage, potentially leading to inconsistent standards and unreliable outcomes. Therefore, the development of intelligent ginseng identification technology is crucial. This advanced approach not only circumvents the pitfalls associated with manual handling but also markedly enhances the precision and dependability of classification. Consequently, it leads to a more streamlined and accurate quality assessment process.

In recent years, computer vision technology has played an essential role in identifying Chinese herbal medicines [9] based on deep convolutional neural networks (CNN), which have been proven to grade Chinese herbal medicines. For instance, Li Dongming [10] and colleagues utilized an improved IResNet model to identify the origin of *Saposhnikovia divaricate* by enhancing both the early and late stages of the model. Kim et al. [11] compared and analyzed the performance of four different models using five different preprocessing methods on the grading accuracy of red ginseng images, and finally, DenseNet121 with CLAHE preprocessing performed the best with an accuracy rate of 95.11%. Li Dongming et al. [12] achieved the same objective by substituting the conventional activation function with Leaky ReLU, incorporating an ECA module, and refining the ResNet50 model through data augmentation on a self-constructed dataset. This resulted in an efficient, rapid, and accurate algorithm for grading the appearance and quality of ginseng. Moreover, Li Dongming et al. [13] introduced a ginseng grading model based on the enhanced ConvNeXt framework. Experimental results demonstrated that this method achieved accuracy improvements of 2.46% and 4.32% over the current state-of-the-art networks, Vision Transformer and Swin Transformer. Moreover, traditional augmented Convolutional Neural Networks (CNNs) and analogous techniques are limited in their ability to perform in-depth analysis of the extracted target features, resulting in many parameters and high computational complexity. While these methods yield excellent recognition outcomes, they demand substantial computational resources and storage space, constraining their utility on resource-constrained mobile devices. Target detection is an important research direction in computer vision [14]. Researchers have widely utilized the YOLO family of models for multi-scale plant disease identification and recognition. Notably, Chen et al. [15] enhanced the TSP-YOLO model for real-time monitoring of kale seedling emergence, achieving a 14.5% increase in mean Average Precision (mAP50-95) over the basic model. Their counting method demonstrated exceptional performance in both speed and accuracy. Similarly, Yang et al. [16] developed a strawberry maturity detection and grading model by integrating the YOLOv8s model with the LW-Swin Transformer module, resulting in the LS-YOLOv8s, which exhibits enhanced detection accuracy and efficiency compared to YOLOv8m. Liu et al. [17] proposed a novel lightweight apple detection algorithm called Faster-YOLO-AP based on YOLOv8. Parameters and floating-point operations (FLOPs) are reduced to 0.66 M and 2.29 G, respectively, with mAP50-95 of 84.12%. On edge computing devices, the Faster-YOLO-AP model shows superior performance in terms of speed and

accuracy compared to other lightweight models. Ma et al. [18] proposed an improved YOLOv5n model, CTR_YOLOv5n, for recognizing common corn leaf spots, gray spots, and rust diseases. The average recognition accuracy of the algorithmic model can reach 95.2%, which is 2.8% higher than the original model, and the memory size is reduced to 5.1 MB to fulfill the lightweight requirement. However, the YOLO series models have not yet been applied to detect the appearance and quality of ginseng. This paper presents a lightweight model based on an enhanced YOLOv8n named DGS-YOLOv8 to address these challenges. This model is specifically tailored to achieve superior detection capabilities for ginseng appearance quality classification, addressing current limitations and optimizing performance for use on mobile devices.

We can summarize the main contributions of this study as follows:

(1) Adopting a novel cross-phase partially connected module, termed C2f-DCNv2, integrates the advanced DCNv2 (Deformable Convolutional Networks v2) with the C2f architecture within the backbone region. Ginseng exhibits intricate shapes and poses, which can be challenging for traditional convolutional methods to capture. The dynamic mask feature of C2f-DCNv2 excels in handling these subtle deformations, allowing the network to delineate the minute surface features of ginseng accurately. This enhancement significantly boosts the precision of classification and detection tasks.

(2) Introduction of a streamlined neck structure, the Slim-Neck (GSCnv + VoVGSCSP): This replaces the C2f and standard convolution in the neck layer of YOLOv8. This modification effectively reduces computational complexity and network infrastructure while still preserving a robust level of accuracy.

(3) Integration of SimAM (Simple Attention Module) at the P5 layer: With its parameter-free design and minimal computational requirements, SimAM effectively highlights the critical features of ginseng, such as texture and morphology, without imposing a substantial computational load.

2. Material Handling and Methods

2.1. Ginseng Image Dataset Acquisition

The demand for processed fresh ginseng varies, with white ginseng experiencing exceptionally high popularity. White ginseng, derived from fresh ginseng cultivated for several years, undergoes fundamental processing procedures such as cleaning, drying, or sun-drying. White ginseng from Fusong County in Jilin Province was chosen as the research material for this study. The researchers meticulously selected and prepared these samples to ensure the integrity and reliability of the experiment. The sample collection process began with gathering samples from several pre-sorted sample boxes. These samples were initially categorized and then reclassified and authenticated by local experts according to the grading criteria outlined in the Released Edition of Ginseng of Jilin Province Roadside Medicinal Herbs, details are shown in Table 1. This classification system enabled the ginseng samples to be divided into three distinct categories, ensuring accuracy and consistency in the classification process. For capturing the visual data, the research team used a high-quality compact folding studio box (Sutefoto, Guangzhou, China) and a mobile phone camera (Apple, Cupertino, CA, USA). The camera was positioned on top of the studio box, directly overhead and perpendicular to the ginseng samples, at a height of 35 cm. This setup ensured that all samples were photographed from a consistent angle, minimizing variations in image quality due to perspective differences. To ensure comprehensive and diverse data, samples were captured from multiple angles against various backgrounds, including white, black, and dark brown woodgrain. These background choices were intended to enhance image contrast and clarity, making the details of the samples more pronounced. Figure 1 illustrates the arrangement and usage of the data collection equipment, detailing the placement of the studio box, the camera setup, and the positioning of the samples. Throughout the imaging process, all samples were photographed under consistent conditions to maintain data uniformity and comparability. Each image was captured in high

resolution to ensure detailed recording of sample characteristics, providing a reliable basis for subsequent analysis and classification.

Table 1. Ginseng grading criteria.

Projects	Principal Ginseng	First-Class Ginseng	Second-Class Ginseng
Main Root	There are 2~3 evident branched roots, and the thickness is more uniform	Cylindrical-like	One to four branches, coarser and finer Rutabaga and ginseng with incomplete fibrous roots
Branch Root			
Rutabaga	Complete with reed head and ginseng fibrous roots	The reed head and ginseng fibrous roots are more complete	
Groove	Clear grooves	Not unmistakable, distinct groove	Without grooves
Diameter Length	≥3.5	3.0–3.49	2.5–2.99
Surface	Yellowish-white or grayish-yellow, no water rust, no draw grooves	Yellowish-white or grayish-yellow, light water rust, or with pumping grooves	Yellowish-white or grayish-yellow, slightly more water rust, with pumping grooves
Cross-section		Yellowish-white in section, powdery, with resinous tract visible	
Texture		Harder, powdery, non-hollow	
Damage, Scars	No significant injury	Minor injury	More serious
Insects, Mildew, Impurities	None	Mild	Presence
Section	Section neat, clear	Segment is obvious	Segments are not obvious
Springtails	Square or rectangular	Made conical or cylindrical	Irregular shape
Weight	500 g/root or more	250–500 g/root	100–250 g/root

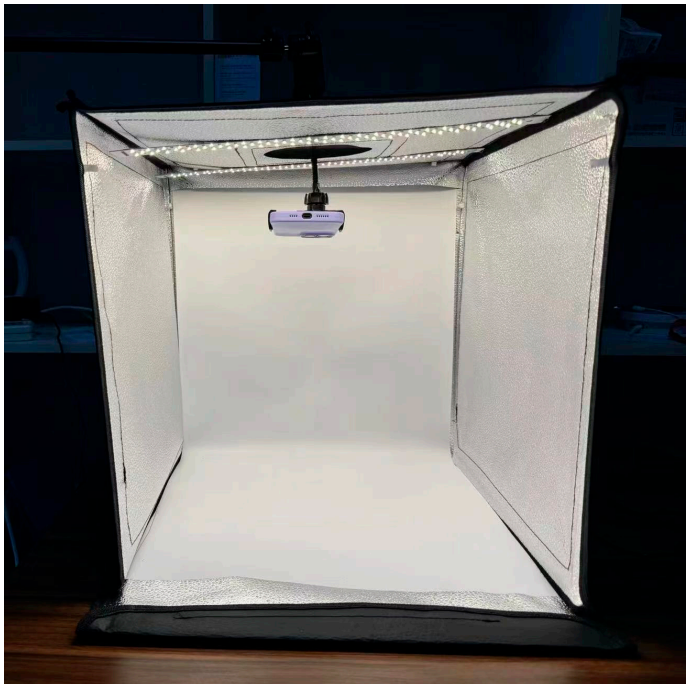


Figure 1. A ginseng image acquisition device.

2.2. Sample Pre-Processing and Creation of Datasets

The scale of datasets is intrinsically linked to the precision of deep learning-based computer vision models. A larger dataset enhances the target deep network’s feature extraction and learning capabilities. It also mitigates experimental bias, better mimics intricate real-world scenarios, and reduces the reliance on deep learning image detection and classification models on particular image features. By expanding the dataset, we aim to bolster the robustness and generalization capabilities of the network [19–21]. The original dataset of 1343 images was divided into training and validation sets in an 8:2 ratio to achieve this objective. Various augmentation techniques, such as horizontal flipping, rotation, and noise addition, expanded the dataset to 4029 images, distribution as shown in Table 2. These augmentations simulated potential appearances of ginseng, including conditions like red rust, excessive surface dryness, and characteristic color changes. These enhancements ensured a comprehensive representation of the various scenarios ginseng might present, thereby improving the model’s generalization and adaptability to complex conditions. Care was taken to ensure that the validation set did not contain duplicate entries. For evaluation purposes, we used a separate set of images from three ginseng categories photographed in 2022, with the number of images matching that of the validation set. During the data annotation phase, we employed the Labelling image annotation toolbox and the VOC dataset format to annotate the image data. For the representation of ginseng grades, we assigned three distinct labels: ‘particular grade’, ‘first grade’, and ‘second grade’, each labeled numerically as 0, 1, and 2, respectively. This structured annotation process facilitates our computer vision model’s accurate training and assessment.

Table 2. Dataset classification.

Level	Number of Original Training Sets	Number of Enhanced Training Sets	Number of Original Validation Sets	Number of Enhanced Training Sets
Principal	339	1017	85	255
First-class	380	1140	100	300
Second-class	355	1065	84	252

The Figure 2 illustrates the dataset distribution, highlighting the count of specific grades, explicitly focusing on first-grade and second-grade ginsengs. Furthermore, it provides an in-depth look at the detailed distribution post-enhancement. We can furnish ample and representative data for the training model through this meticulous labeling and strategic division of the dataset. This, in turn, enhances the model’s performance and bolsters its generalization capabilities.

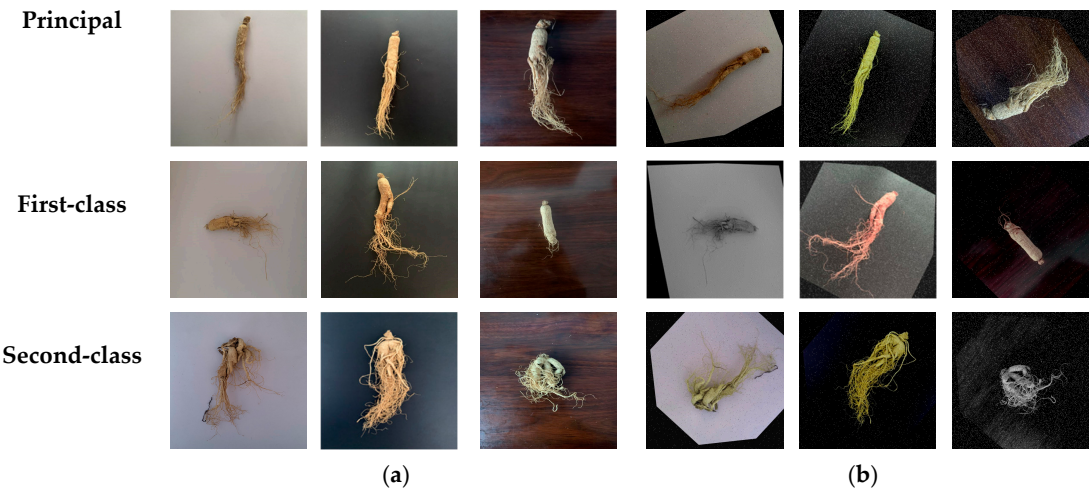


Figure 2. Ginseng dataset. (a) Original dataset; (b) dataset after data enhancement.

2.3. YOLOv8 Algorithm and Improvements

2.3.1. YOLOv8 and DGS-YOLOv8 Network Architecture

The YOLO (You Only Look Once) algorithm family is renowned for its exceptional efficiency and precision [22]. YOLOv8, a state-of-the-art model introduced by Ultralytics in 2023, is an evolution of the YOLOv5 architecture [23], enhanced with fusion improvements. Its standout features include adopting the C2f module within the upgraded C3 module as the primary residual learning unit. The detection head ingeniously integrates anchor-free and decoupled-head techniques, while the loss function combines classification BCE, regression CIOU, and VFL strategies. The frame-matching method has been replaced with the innovative Task-Aligned Assigner approach, and the model benefits from the Mosaic technique during the last ten epochs [24]. YOLOv8 streamlines the conventional convolutional layer, optimizing the use of the Bottleneck module to enhance gradient branching capabilities.

Additionally, it introduces an image segmentation algorithm that marries a deep learning network model with an adaptive threshold function, thereby capturing a wealth of gradient streaming data. By integrating multiple Conv and C2f modules, the model efficiently processes input images to extract feature maps at various scales while incorporating the advantages of the ELAN structure from YOLOv7. The SPPF (Spatial Pyramid Pooling-Fast) module further refines and conveys the output feature maps to the neck layer. Figure 3 illustrates the detailed network architecture, showcasing the intricate design that sets YOLOv8 apart as a cutting-edge solution in object detection.

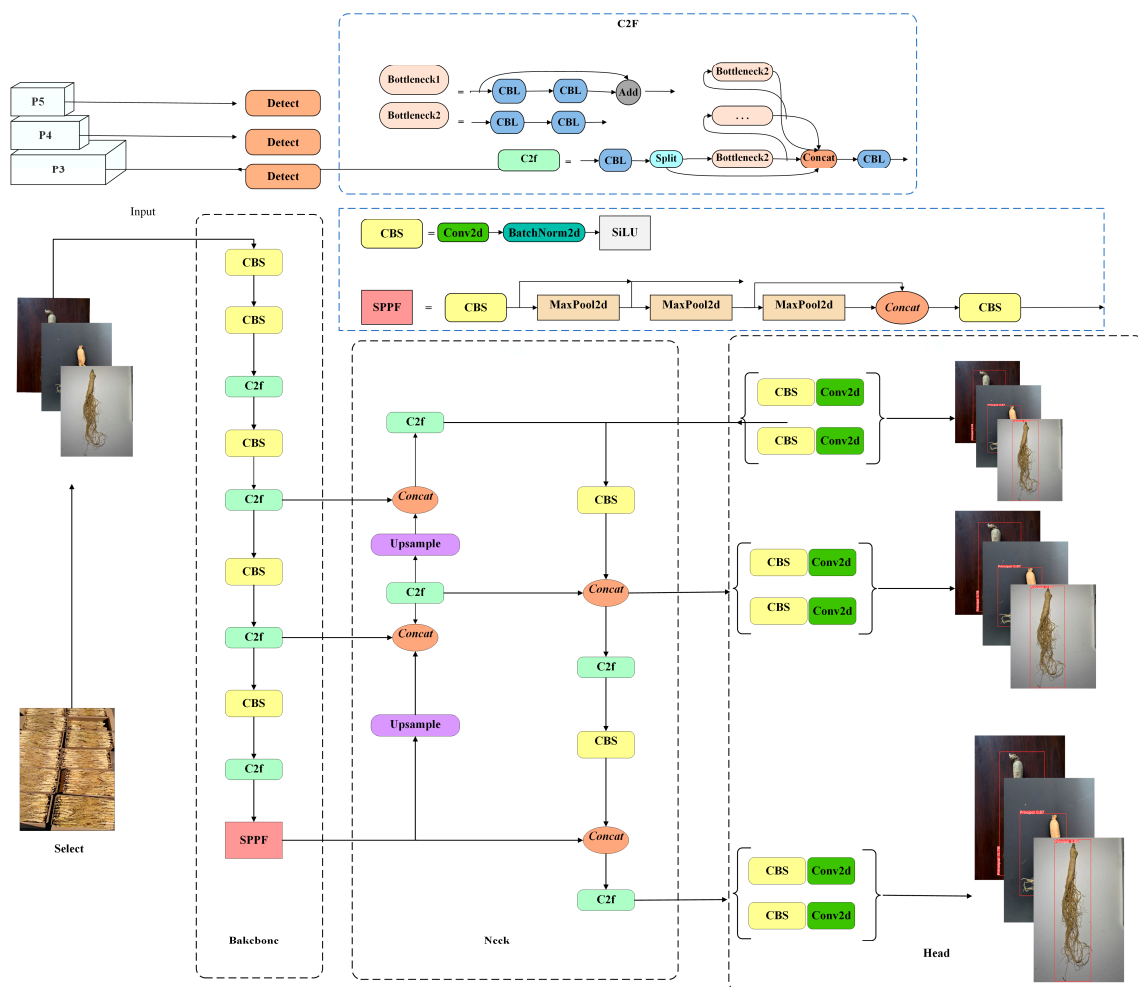


Figure 3. YOLOv8 model structure.

The enhanced DGS-YOLOv8 architecture integrates the advanced deformed convolutional network V2, equipped with the C2f module, to create a hybrid C2f-DCNV2. This innovative fusion replaces the original C2f module in layers 6 and 8 of the primary network's backbone. In the model's neck section, the traditional convolution paired with the C2f module is substituted with a more efficient lightweight convolution, incorporating GSConv and VoVGSCSP. The SimAM attentional mechanism is implemented in layer P5, preceding the network's head in the integrated DGS-YOLOv8 network structure. As illustrated in Figure 4, this integration highlights the optimization and complexity of the DGS-YOLOv8 architecture.

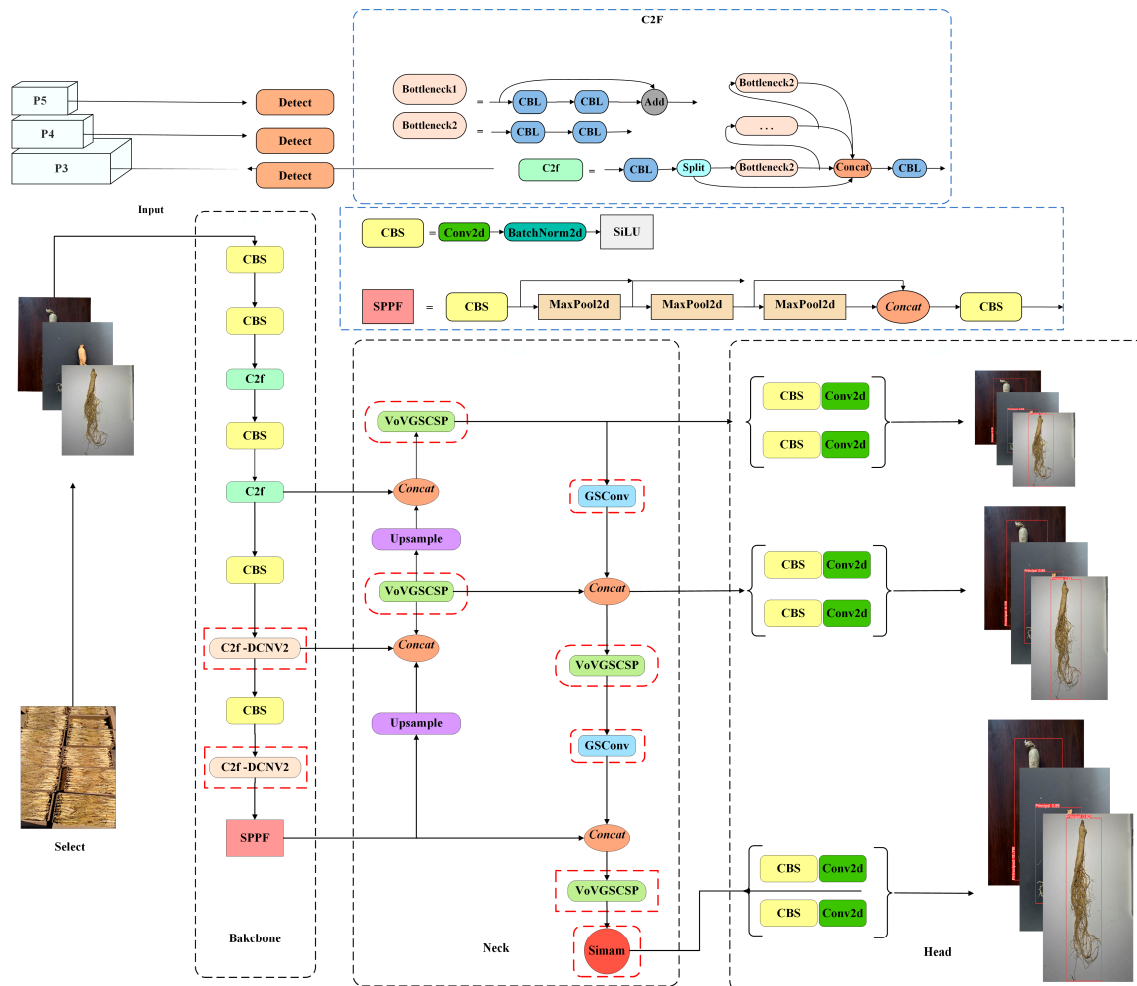


Figure 4. Improved YOLOv8 model (DGS-YOLOv8).

2.3.2. YOLOv8-C2f-DCNV2

The C2f-DCNV2 module in YOLOv8 has shown significant advantages in improved object detection performance [25]. The core advantage of the DCNV2 (Deformable Convolutional Network V2) module is its innovative deformable convolution mechanism, which provides unprecedented flexibility and enables the convolutional kernel to adjust its sampling position to optimize feature extraction capabilities adaptively. Traditional convolution operations have fixed sampling locations, which makes it difficult to cope with the diversity of target shapes and locations, especially when dealing with complex scenes. This limitation has prompted researchers to explore more flexible convolution methods to improve the effectiveness of feature extraction. The DCNV2 module effectively solves this problem by introducing a deformable convolutional layer. Specifically, the DCNV2 module uses deformable convolutional layers for feature extraction in its working mode. These layers dynamically generate offsets based on the input feature map, allowing the

sampling position to be adjusted to capture complex global features efficiently. In the initial feature extraction phase, deformable convolutional layers significantly improve the ability to capture details. Subsequently, the extracted features are further refined by an additional deformable convolutional layer, enhancing the ability to capture subtle changes and local features, resulting in a more nuanced understanding of the image. The DCNv2 module fuses diverse feature representations through simple addition, ensuring a comprehensive and robust feature set. This fusion strategy not only simplifies the calculation process but also enhances the ability to express features and further improves the accuracy and robustness of object detection. As shown in Figure 5, the complex structure of DCNv2 demonstrates its detailed composition and functions. The DCNv2 module can perform well in various complex scenarios through this structural design, providing strong technical support for YOLOv8.

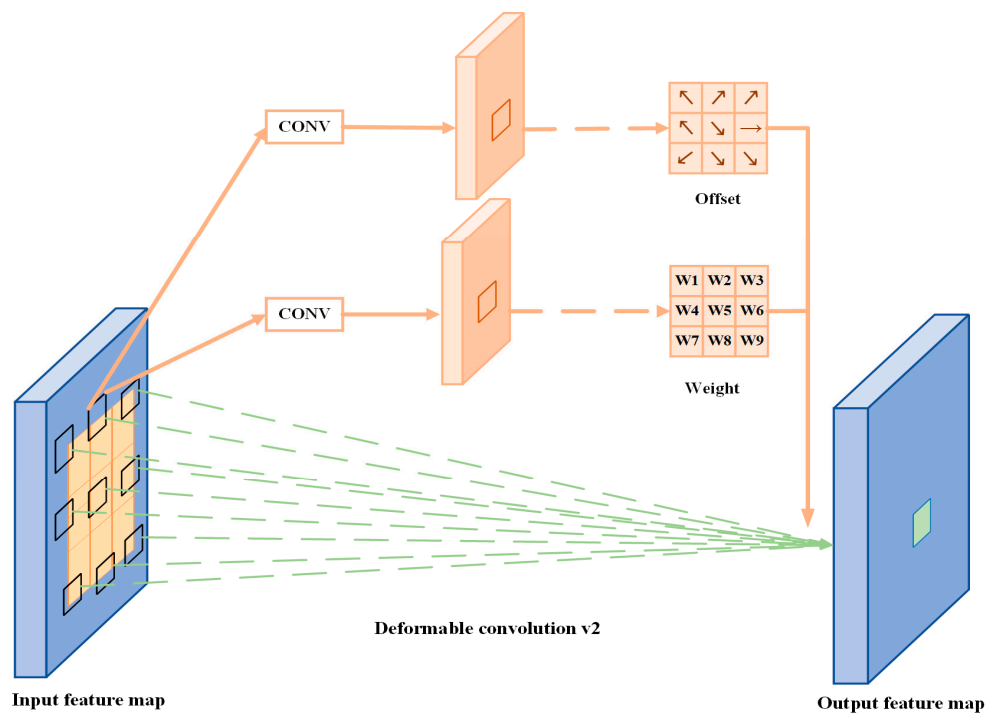


Figure 5. Illustration of 3×3 deformable convolution net v2.

The formula for DCNv2 is expressed as an equation.

$$y(p) = \sum_{k=1}^K w_k \cdot x(p + p_k + \Delta p_k) \cdot \Delta m_k \quad (1)$$

The input feature map x and the output feature map y can be imagined as an input image processed to obtain an output image. Pixel location p represents the position of a specific pixel in the image, while convolution kernel sampling points p_k are the locations where the convolution kernel samples the image. Each sampling point weighs and is used to determine the impact of that point on the final result. The learnable deviation Δp_k is a positional deviation that can be adjusted to allow the sampling point to move slightly over the picture to capture more detail. Modulation parameter Δm_k is used to adjust the influence of the sampled point further. Please think of this process as looking at a picture where each pixel point has a small window (convolution kernel) looking at the area around it, and this window will not only look at a fixed position but also move slightly to find more detail.

Based on the unique shape characteristics of ginseng, we developed the C2f-DCNv2 model by integrating the DCNv2 architecture with the C2f module, as shown in Figure 6.

The C2f-DCNv2 module comprises two DCNv2 layers and several Bottleneck Units. This design enables the C2f-DCNv2 model to effectively combine feature information across multiple scales, thereby enhancing its ability to represent complex shapes like ginseng. The first DCNv2 layer performs initial processing of the input features, which are then segmented for further handling by the Bottleneck Units. Each Bottleneck Unit consists of two DCNv2 layers and may include shortcut connections. Units with shortcuts allow the direct addition of input features to the output features, facilitating gradient flow and feature reuse, whereas units without shortcuts focus on nonlinear transformations. The outputs from all Bottleneck Units are integrated through feature concatenation and processed by the second DCNv2 layer, enhancing the overall feature representation. The configuration of multiple Bottleneck Units and DCNv2 layers allows C2f-DCNv2 to extract and fuse features at different scales, improving the model's capability to capture and represent complex shapes such as ginseng. The repeated processing by Bottleneck Units and the feature concatenation enhance the model's sensitivity to the intricate details of the target shape, making it more effective in capturing fine features.

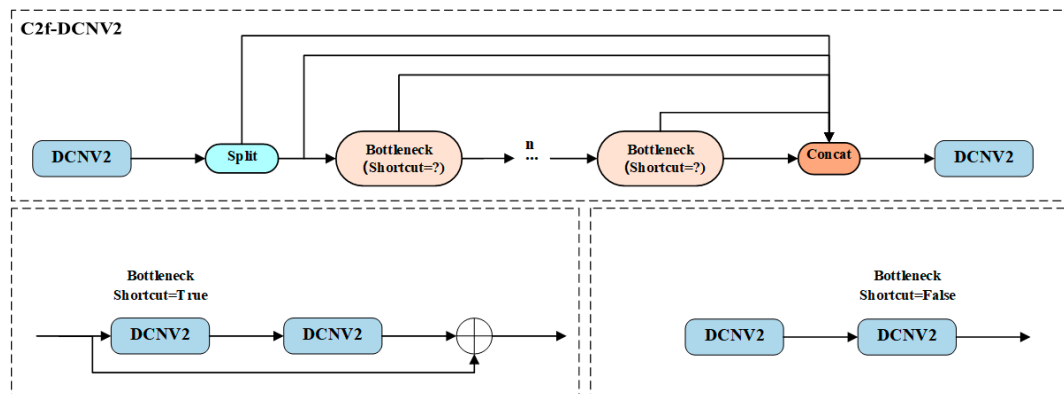


Figure 6. The structure of the C2f-DCN network.

2.3.3. YOLOv8-GSconv

In a recent advancement in deep learning, GSConv [26], a novel lightweight convolution method, was introduced in 2022. This method encompasses Conv, DWConv [27], Concat, and Shuffle modules, as illustrated in Figure 7. GSConv excels in improving model accuracy accelerating model convergence and detection processes, and its core design is to optimize the feature extraction mechanism. By reducing computational complexity and the number of parameters and enhancing the effectiveness of the attention module, GSConv enables the model to focus more on critical areas of the image. Although GSConv can replace the conventional convolution in the model, using the backbone part deepens the network hierarchy, increasing the data flow resistance and inference time. Therefore, the researchers mainly deployed GSConv in the head section where the feature map was already compact enough to avoid further conversion. The computational cost of GSConv processing feature maps is about 60% to 70% of that of standard convolution methods, effectively reducing redundancy and eliminating duplicate information. In addition, the researchers proposed a multi-scale image pyramid model, combined with a feature extraction algorithm, to convert the input 2D feature map into 3D tensors, refine these features through 3D convolution, and finally reconstruct them into a 2D feature map using GSConv technology. This marks an efficient graph. In order to further improve the model's performance, it is vital to choose the Slim-Neck module. By simplifying the network structure and reducing the number of parameters, the Slim-Neck module significantly reduces the computational cost and increases the speed of inference, which is especially critical for real-time applications and resource-constrained devices. By streamlining feature layers and optimizing information flow, Slim-Neck avoids unnecessarily complex calculations while ensuring the accuracy and effectiveness of feature extraction. It optimizes feature extraction while retaining

critical information and reducing computational redundancy, resulting in faster model convergence and higher detection accuracy. The results show that Slim-Neck outperforms the traditional deep network structure in various tasks, especially in scenarios that require fast response and efficient computing. Combined with Slim-Neck and GSConv technology, the model achieves a breakthrough in lightweight design while reaching new heights in accuracy and speed.

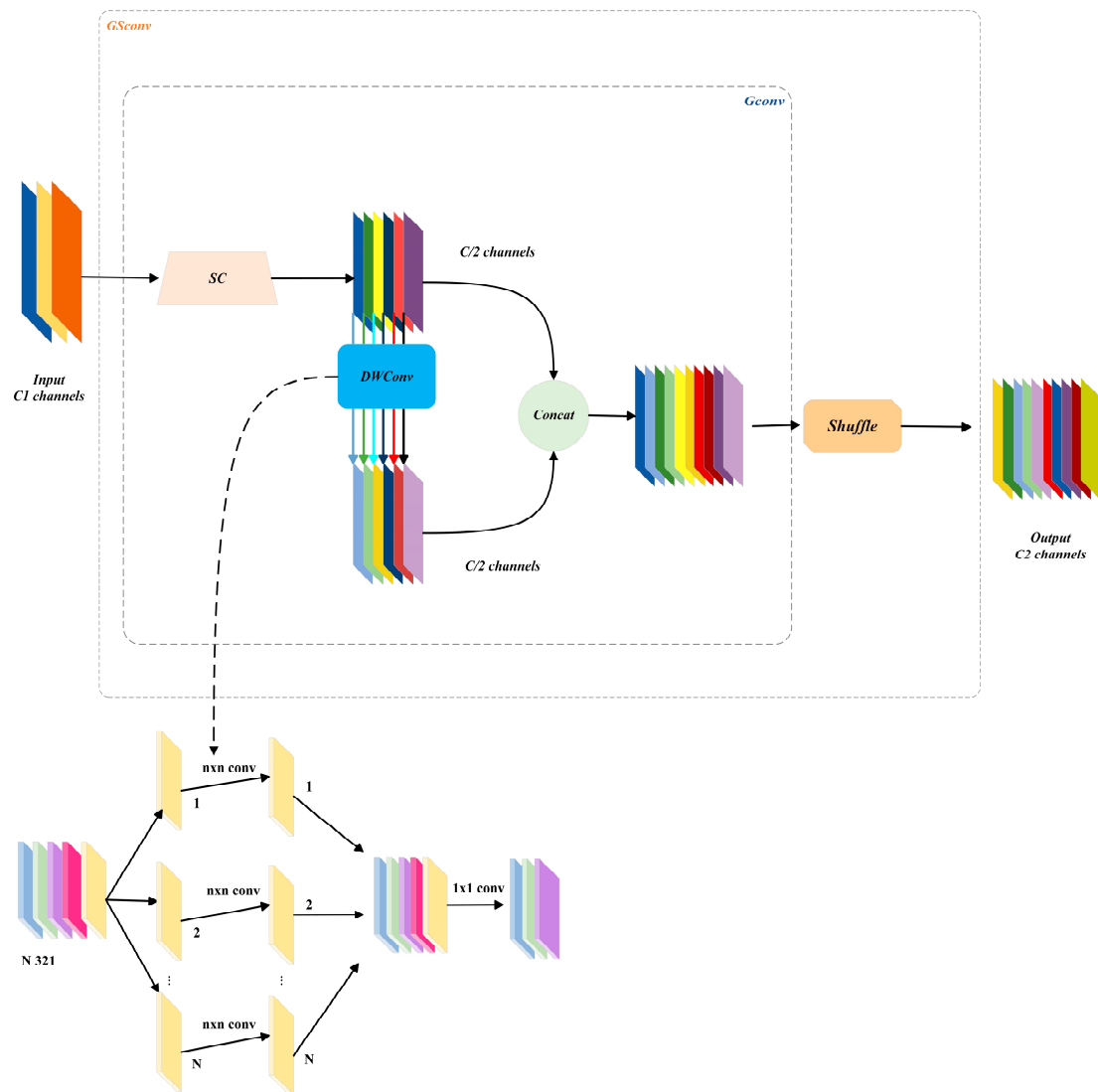


Figure 7. Structure of the GSConv module.

GSConv has been instrumental in mitigating redundant information within feature maps to enhance the efficiency of ginseng appearance detection models. However, there remains a challenge to further curtail inference time without compromising accuracy. In response, the GSbottleneck structure and the VoV-GSCSP network, built upon the GSConv framework, were introduced in this study. Figure 8 illustrates the architecture of the VoV-GSCSP backbone, which adeptly fosters the transmission of robust semantic features by employing a pair of GSConv operations for both up-sampling and down-sampling processes. In the neck section of the proposed model, the C2f module is introduced as a substitute for the VoV-GSCSP module. This strategic replacement diminishes the computational complexity of the model while concurrently preserving an adequate level of accuracy. This modification is critical to balancing efficiency and performance in ginseng appearance detection systems.

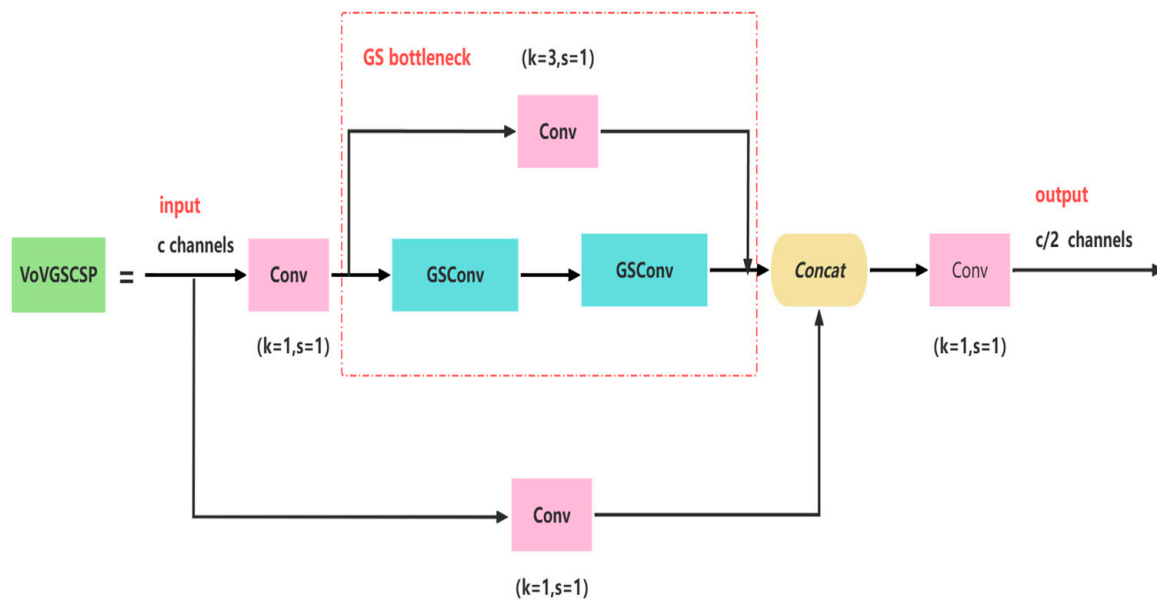


Figure 8. Structure of the VoVGSCSP module.

2.3.4. YOLOv8-SimAM (Simple Attention Module)

SimAM (Yang et al., 2021) is an innovative parameter-free attention module designed to enhance the robustness of neural networks when processing complex features in complex environments [28]. The module computes the importance of each neuron by synthesizing channel and spatial information to derive 3D attention weights. Unlike traditional 1D or 2D attentional mechanisms, SimAM focuses on global and local features, allowing it to accurately capture the nuances and complex structures of the input data. Figure 9 illustrates the components of SimAM, which simplifies the model architecture, reduces computational complexity, and improves performance through parameter-free construction. It quantifies the uniqueness and relevance of neurons in the feature map by calculating the energy function, thus accurately evaluating the contribution of each neuron. SimAM's contributions to performance improvement mainly include integrating global and local features, capturing the nuances and complex structures of input data through 3D attention weighting, making the model more accurate and efficient in handling complex tasks, parameter-free design simplifying the model architecture, reducing computational complexity, improves performance, and is suitable for resource-constrained environments, quantifying the uniqueness of neurons and their correlations through the energy function, optimizing the attention mechanism. The energy function is defined so that lower energy signifies greater importance, as articulated in Equation (2).

$$e_t^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \quad (2)$$

The $(t - \hat{\mu})^2$ formula indicates the distance of the target neuron from the average value of all neurons; the more significant this value, the greater the difference between that neuron and the others. $\hat{\sigma}^2$ is the variance of all neurons, which indicates how dispersed the neuron values are. If the values of the neurons in the channel are very concentrated (slight variance), then the variation in the energy values of each neuron will be more pronounced. λ is a tuning parameter used to balance the various calculations in the formula. Through these calculations, the formula yields an energy value of e_t^* , which indicates the importance of the target neuron in the channel, with a lower energy value indicating a more critical neuron.

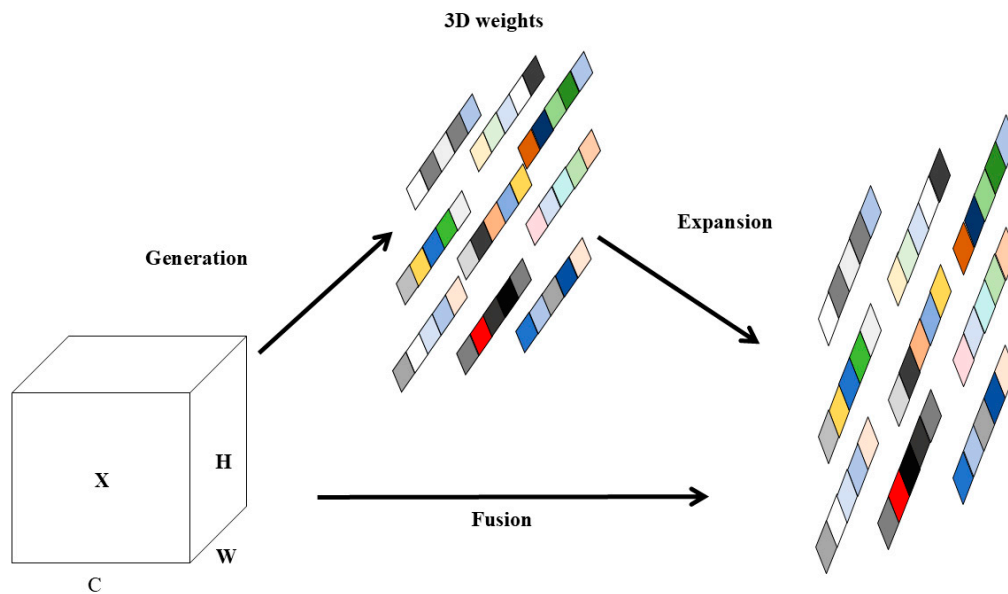


Figure 9. Structure of the SimAM attention mechanism.

The energy function of the target neuron e_t is as follows:

$$e_t = \frac{1}{M-1} \sum_{i=1}^{M-1} (1 - (w_t x_i + b_t))^2 + (1 - (w_t t + b_t))^2 + w_t^2 \quad (3)$$

Equation $e_t = \frac{1}{M-1} \sum_{i=1}^{M-1} (1 - (w_t x_i + b_t))^2 + (1 - (w_t t + b_t))^2 + w_t^2$ is used to assess the importance of the target neuron t , in the channel. Where x_i are all neurons in the channel except the target neuron t , M is the total number of all neurons in the channel, w_t , and b_t are the weight and deviation of the target neuron, respectively, and w_t is used to prevent overfitting and control the magnitude of the weight. First, the formula calculates the difference between the target neuron and other neurons and the weighted difference of the target neuron itself then averages these difference values. Thus, here is a description of the SimAM attention module:

$$X_{\text{att}} = \text{Sigmoid}\left(\frac{1}{e_t}\right) \odot X \quad (4)$$

The inverse of the target neuron's energy value e_t is $\frac{1}{e_t}$. The lower the energy value, the larger the inverse, indicating the more critical the neuron is. Then, the inverse value is limited to between 0 and 1 by the Sigmoid function to obtain the scaling factor $\text{Sigmoid}\left(\frac{1}{e_t}\right)$, and finally, the original feature map X is multiplied element by element with the scaling factor to obtain the weighted feature map X_{att} . The Sigmoid function controls the magnitude of the weights and limits the extremes within a reasonable range. This adjustment does not affect the relative importance of individual neurons because the Sigmoid function is monotonic.

2.4. Experimental Environment

In the experiments, Ubuntu 18.04 was used as the operating system, PyTorch was used as the deep learning framework, an experimental platform was set up, and Python 3.9.13 and torch-1.1.3 + cuda11.6 were used. The CPU model is Intel(R) Xeon(R) Silver 4214R 2.40 GHz. The graphics card model was NVIDIA GeForce RTX 3090 with 24,260 MiB of memory, was sourced from NVIDIA, based in Santa Clara, United States. The detailed hyperparameters of the experiment are shown in Table 3.

Table 3. Detailed hyperparameters of the experiment.

Parameters	Setup
Epochs	200
Batch Size	32
Optimizer	SGD
Initial Learning Rate	0.001
Final Learning Rate	0.001
Momentum	0.937
Weight-Decay	5×10^{-4}
Close Mosaic	Last ten epochs
Images	640
Workers	8
Mosaic	1.0

2.5. Evaluation Criteria

This study evaluated the performance of YOLOv8 and the improved model using Recall, Precision, AP, mAP, and F1 score composite metrics.

TP (True Positive) shows the number of accurate ginseng identifications detected by the YOLOv8 network model. FP (False Positive) indicates the number of inaccurate ginseng seedling identifications detected by the YOLOv8 network model. FN (False Negative) is the number of actual ginseng samples not identified by the model. *Recall* (R) shows how many actual ginseng samples were correctly identified. The ratio of true positive (TP) detections to the sum of true positives and false negatives (FN) reflects the model's ability to identify all relevant instances. Precision (P) indicates how many ginseng samples with optimistic predictions were correct. It is the ratio of accurate positive detections to the sum of true and false positives (FP). AP (Average Precision) for a specific class is calculated as the ratio of true positives to the adjusted sum of true positives and false negatives divided by the number of instances (N). This measures the precision across different recall levels. The average of the AP values for multiple categories is mAP (mean Average Precision), and a higher value indicates a higher average accuracy of the model's detection for each category. F1 score is the harmonic mean of Precision and Recall, providing a single metric that balances both. It is advantageous when there is an uneven class distribution.

The formula is as follows:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$AP = \int P(R) dR \times 100\% \quad (7)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (8)$$

$$F1 = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \quad (9)$$

3. Experiments and Analysis of Results

3.1. Experimental Comparison before and after Model Improvement

Figure 10 shows that DGS-YOLOv8 outperforms the original YOLOv8 model in three key performance metrics: accuracy, mAP50, and mAP50-95 over 200 training cycles. DGS-YOLOv8 consistently demonstrates higher accuracy than YOLOv8 throughout all training cycles and exhibits better stability, suggesting a superior ability to reduce false alarms. Additionally, DGS-YOLOv8 performs well across different IoU threshold ranges. In the mAP50 and mAP50-95 metrics, DGS-YOLOv8 shows more significant improvements in the

middle and late stages of training, achieving higher target detection accuracy than YOLOv8. These results indicate that DGS-YOLOv8 improves detection accuracy and significantly enhances the model's robustness.

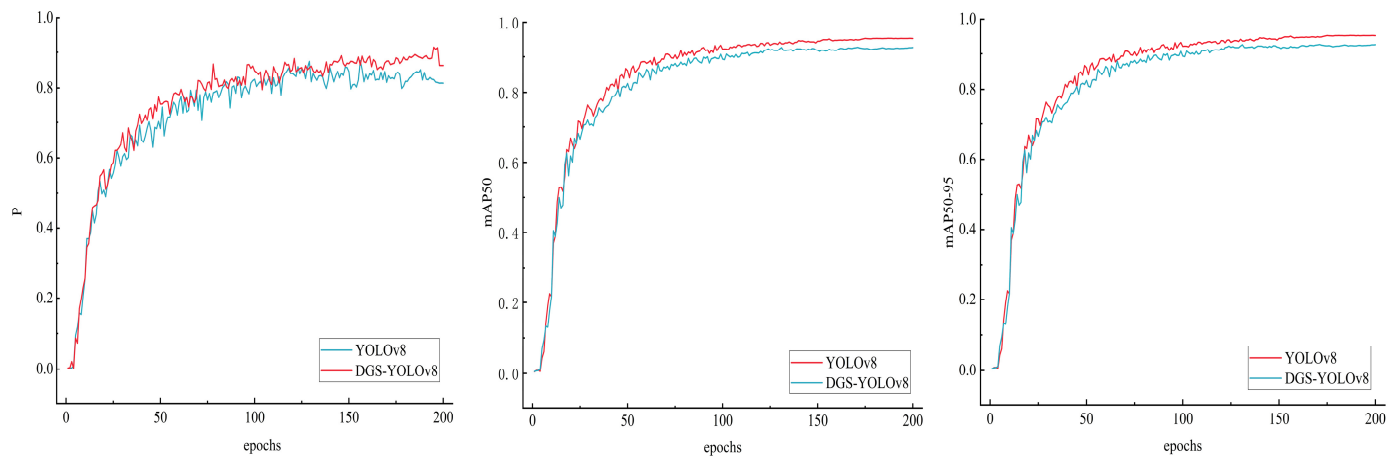


Figure 10. Comparison of indicators before and after model improvement.

Table 4 illustrates that after 200 training sessions, the improved DGS-YOLOv8 model outperforms the original YOLOv8 model across various evaluation categories. Specifically, DGS-YOLOv8 significantly improves the principal category, substantially increasing precision and recall. The F1 score, mAP50, and mAP50-95 improved by 2.72%, 2.90%, and 4.73%, respectively. In the First-class category, DGS-YOLOv8 also excels, with enhancements in precision and recall leading to notable increases in the F1 score, mAP50, and mAP50-95. The most remarkable improvement is seen in the Second-class category, where DGS-YOLOv8's precision increased by 10.01%, the F1 score rose by 5.08%, and mAP50 and mAP50-95 improved by 2.62% and 3.30%, respectively. Overall, DGS-YOLOv8 achieved a 6.86% improvement in precision, a 3.83% increase in the F1 score, a 2.73% rise in mAP50, and a 3.82% enhancement in mAP50-95. These findings underscore the substantial improvements in accuracy and mean Average Precision (mAP) across all categories for DGS-YOLOv8, with particularly impressive gains in the Second-class category. These significant advancements highlight the superior detection accuracy and precision rate offered by the DGS-YOLOv8 model.

Table 4. Improved classification results of ginseng.

Level	Model	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)	mAP50-95 (%)
Principal	YOLOv8	84.08	89.79	86.84	93.57	68.28
	DGS-YOLOv8	89.33	89.8	89.56	96.47	73.01
First-class	YOLOv8	80.43	90.63	85.23	92.29	70.35
	DGS-YOLOv8	85.75	91.85	88.70	94.96	73.77
Second-class	YOLOv8	80.18	85.12	82.7	91.85	72.48
	DGS-YOLOv8	90.19	85.49	87.78	94.47	75.78
ALL	YOLOv8	81.56	88.52	84.9	92.57	70.37
	DGS-YOLOv8	88.42	89.05	88.73	95.3	74.19

3.2. Ablation Experiment

The C2f-DCNv2, Slim-Neck, and SimAM modules are introduced to enhance the performance of the YOLOv8n model in recognizing the appearance quality of ginseng. The C2f-DCNv2 module combines the advantages of the C2f structure and deformable convolutional network (DCN). The C2f structure improves feature extraction by integrating features across different stages of the network, enhancing feature propagation and gradient flow. This allows for effective feature extraction at multiple scales, which is crucial for cap-

turing the diverse appearance features of ginseng. The DCN component adapts the shape and position of convolutional kernels to handle geometric deformations such as scaling and rotation, thus improving the model's ability to recognize ginseng despite variations in appearance. With this module's introduction, as shown in Table 5, the model's accuracy improved from 81.56% to 84.94%, the F1 score from 84.88% to 86.88%, and mAP50 from 92.57% to 93.59%. Although the number of parameters increased to 3.07 M and the model weight to 6.4 MB, these enhancements in performance highlight the module's effectiveness in improving feature extraction and handling geometric variations. The Slim-Neck module incorporates a lighter network structure that reduces the number of parameters while maintaining or enhancing feature extraction capabilities. This optimization results in a reduced parameter count of 2.80 M and a model weight of 5.9 MB, while achieving accuracy improvements to 87.79%, an F1 score of 87.68%, and mAP50 of 93.74%. The Slim-Neck module balances the reduction in parameters with maintained or improved performance, making the model suitable for lightweight applications without significant sacrifices in feature extraction ability. The SimAM module introduces an attention mechanism that enhances the model's focus on essential features. This mechanism improves the model's ability to recognize detailed ginseng features by prioritizing important regions in the feature maps, all without significantly increasing computational complexity. As a result, the model's accuracy increased to 86.19%, the F1 score reached 86.38%, and mAP50 rose to 92.83%, with parameters set at 3.01 M and the model weight at 6.3 MB. The SimAM module's attention mechanism enhances feature recognition while keeping the computational load manageable. Combining the C2f-DCNv2 and Slim-Neck modules, the model achieves significant improvements in accuracy and F1 score despite reducing the parameter count, indicating potential for lightweight applications. The combination of Slim-Neck and SimAM modules not only reduces the number of parameters but also enhances overall recognition performance by improving the model's focus on essential features. Integrating the C2f-DCNv2 and SimAM modules combines advanced feature extraction with attention mechanisms, leading to significant improvements in recognizing detailed features, although with a slight increase in the number of parameters. Integrating all three modules—C2f-DCNv2, Slim-Neck, and SimAM—demonstrates exceptional performance, with accuracy reaching 88.42%, an F1 score of 88.73%, mAP50 at 95.3%, and mAP50-95 at 74.19%. The combined model's parameters are 2.86 M, and the model weight is 6.0 MB. This integration effectively extracts features and manages the unique deformation of ginseng roots while achieving a good balance between performance and model complexity. The lightweight design of Slim-Neck and SimAM, along with the attention mechanism, results in an efficient and effective model for practical applications.

Table 5. Ablation experiments.

YOLOv8n	C2f-DCNv2	Slim-Neck	Simam	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)	mAP50-95 (%)	Parameters (M)	Weight (MB)
✓				81.56	88.52	84.88	92.57	70.37	3.01	6.3
✓	✓			84.94	88.76	86.88	93.59	72.61	3.07	6.4
✓		✓		87.79	87.53	87.68	93.74	71.8	2.80	5.9
✓			✓	86.19	86.61	86.38	92.83	70.86	3.01	6.3
✓	✓	✓		86.54	88.0	87.26	93.67	72.97	2.86	6.0
✓	✓		✓	84.80	87.48	86.10	92.82	71.85	3.07	6.4
✓		✓	✓	84.63	89.09	86.82	94.0	71.83	2.80	5.9
✓	✓	✓	✓	88.42	89.05	88.73	95.3	74.19	2.86	6.0

3.3. Comparison of Results of Different Attention Mechanisms

In this investigation, the influence of diverse attention mechanisms on the efficacy of neural network models is meticulously assessed by incorporating them into an enhanced version of the YOLOv8 architecture. The following attention mechanisms were integrated into the framework: Context-Based Multi-Head Attention (CBMA) [29], which synergistically merges channel and spatial attention to notably refine feature extraction; Coordinate

Attention (CA) [30], which adeptly incorporates coordinate information to adeptly capture long-range dependencies; Squeeze-and-Excitation (SE) [31], which modulates channel weights via global average pooling to enrich feature representation; Efficient Multi-Head Attention (EMA) [32], which employs an exponential moving average technique to progressively assimilate historical data for temporal attention smoothing; Global Attention Mechanism (GAM) [33], which leverages global contextual information to amalgamate feature interdependencies; Non-Local Attention Mechanism (NAM) [34], which enhances the stability of attention weight calculation through normalization procedures; and SIMAM, characterized by its succinct design, particularly suited for lightweight models to curtail computational expenditure. The subsequent Table 6 delineates the empirical outcomes, wherein all models equipped with an attention mechanism surpass the performance of the baseline model devoid of such enhancements across all evaluation metrics. Notably, SimAM emerges as the frontrunner across all assessment criteria, with its precision reaching 88.42%, recall at 89.05%, F1 score at 88.73%, mAP50 at 95.30%, and mAP50-95 at 74.19%, demonstrating its excellent ability to improve the overall performance of the model. CBMA also shows commendable performance, with a precision of 87.88%, recall of 87.87%, F1 score of 87.94%, and mAP50-95 values of 73.23% and 73.71%, respectively, highlighting its robustness in feature extraction and accuracy improvement. CA and NAM excel in capturing long-range dependencies and maintaining computational stability, with CA achieving 85.62% precision, 87.05% recall, 86.30% F1 score, and 71.63% mAP50-95, while NAM achieves 85.05% precision, 87.96% recall, 86.46% F1 score, and 72.22% mAP50-95. Additionally, EMA and GAM demonstrate a high degree of resilience in handling time series data and global feature integration. EMA achieves a precision of 83.96%, recall of 89.07%, F1 score of 86.48%, and mAP50-95 of 72.91%, while GAM achieves a precision of 86.22%, recall of 88.16%, F1 score of 87.12%, and mAP50-95 of 72.15%. These results underscore the effectiveness of SimAM in enhancing model performance, while CBMA, CA, NAM, EMA, and GAM also provide significant contributions to the robustness and accuracy of the model, each excelling in specific aspects of feature extraction and data handling.

Table 6. Comparison of results of different attention mechanisms.

Attention Mechanisms	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)	mAP50-95 (%)
NO	81.56	88.52	84.88	92.57	70.37
CBMA	87.88	87.87	87.94	93.97	73.23
CA	85.62	87.05	86.30	93.27	71.63
SE	87.71	87.53	87.64	94.38	73.71
EMA	83.96	89.07	86.48	93.89	72.91
GAM	86.22	88.16	87.12	93.68	72.22
NAM	85.05	87.96	86.46	95.3	74.19
SimAM	88.42	89.05	88.73	95.3	74.19

To aesthetically evaluate model efficacy, researchers employed the Grad-CAM (Gradient-weighted Class Activation Mapping) technique [35], delineating the critical regions of the model's focus within an image. By scrutinizing the pictures from the test set using the optimal weights derived from the training phase, insights into the model's performance could be gleaned. As depicted in Figure 11, the heat map produced by the YOLOv8 model exhibits subdued intensity in highlighting ginseng features. Conversely, the DGS-YOLOv8 model's heat map demonstrates enhanced attention to the target area, with clusters of high-activation regions standing out. Integrating the SimAM (Simplified Attention Module) attention mechanism markedly refines the model's ability to concentrate on the target domain, enhancing its discriminative power.

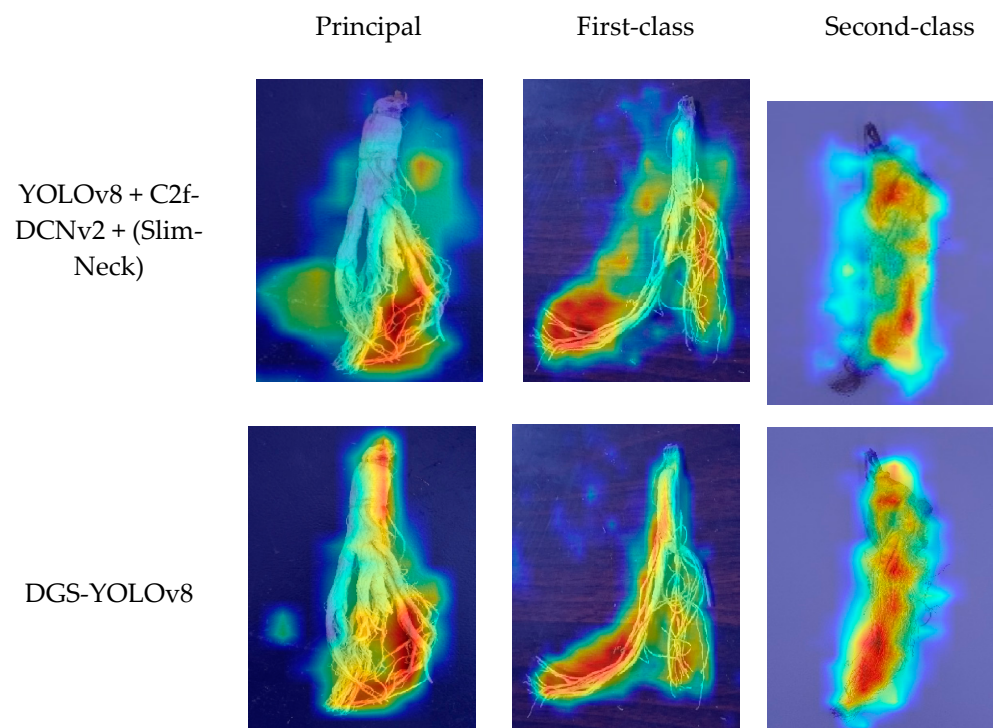


Figure 11. Visualization results of thermal features before and after the introduction of SimAM.

3.4. Optimal Location of Attention

Adopting the attention mechanism, a flexible module that can theoretically be embedded behind any feature processing layer brings multiple advantages. However, the specific performance benefits depend on the exact location where the attention module is added. To gauge the influence of the attention mechanism at different positions within the model, this research embedded the SimAM attention mechanism at various points within the model backbone and the small target detection layer. Subsequently, a series of comparative experiments were meticulously designed and executed to ascertain its efficacy. Table 7 delineates the empirical findings derived from these experiments.

Table 7. The increased impact of SimAM in different locations.

Experiment	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)	mAP50-95 (%)
A	87.32	88.78	88.04	94.11	71.89
B	84.67	88.7	86.64	93.94	72.4
C	84.8	87.7	86.23	92.98	72.13
D	84.93	90.41	87.58	94.19	73.2
E	83.27	90.57	86.77	94.45	73.28
F	88.42	89.05	88.73	95.3	74.19

(A: added to the front of the SPPF added to the trunk; B: added to the back of the SPPF added to the trunk; C: connected to the P4 layer of the up-sampling; D: added to the small object detection layer (P3); E: added to the small object detection layer (P4); F: added to the small object detection layer (P5)).

By integrating the attention mechanism, we gain access to a versatile module that can, in principle, be incorporated behind any feature-processing layer, yielding multiple benefits. The realization of these benefits in terms of performance is contingent upon the strategic placement of the attention module within the architecture. To gauge the influence of the attention mechanism at various positions within the model, the SimAM attention mechanism was embedded at different points within the model backbone and the small target detection layer. Subsequently, this research meticulously designed and executed a series of comparative experiments to ascertain its efficacy.

Figure 12 delineates the trajectory of mean Average Precision at 50% (mAP50) and mean Average Precision at 50% to 95% (mAP50-95) as a function of training epochs for the YOLOv8 model, incorporating the SimAM mechanism at various strategic locations. The graphical representation reveals that all experimental cohorts experience a marked performance surge during the initial training phase, which subsequently plateaus. Experiment F consistently exhibits superior mAP50 and mAP50-95 metrics, indicating that integrating the SimAM mechanism into the P5 layer significantly enhances the model's efficacy. These empirical findings underscore the substantial disparity in performance augmentation achieved by placing the SimAM module at different positions within the network. This insight offers a crucial perspective on the pivotal role of attention mechanisms in refining feature extraction processes.

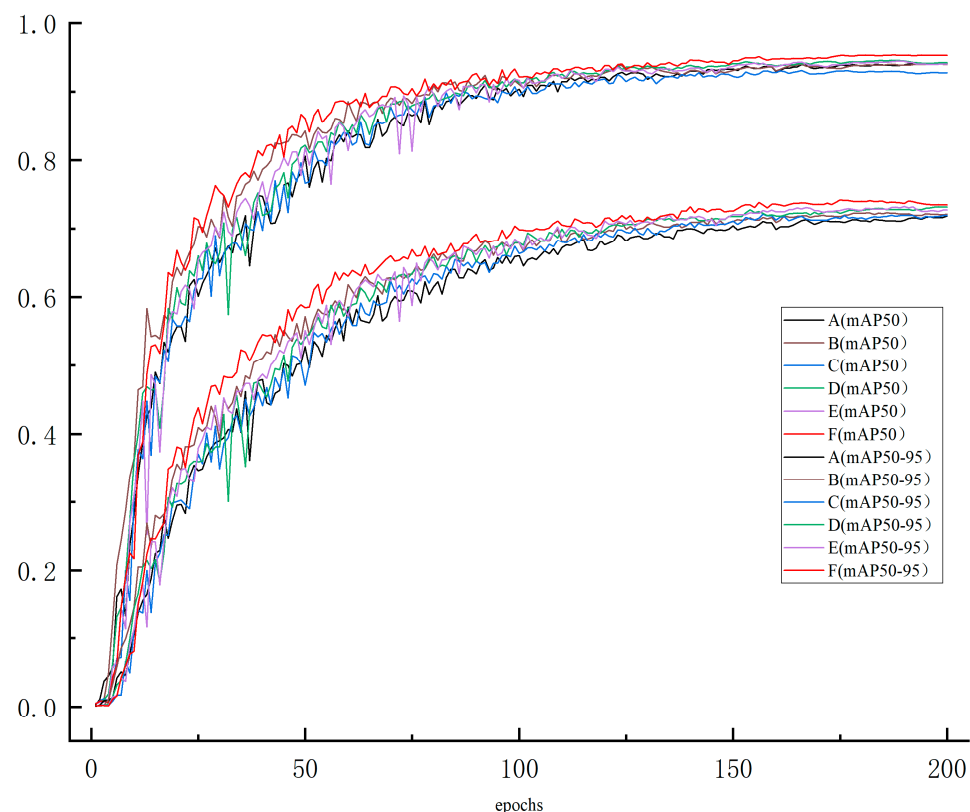


Figure 12. Comparison of mAP50 and mAP50-95 curves of DGS-YOLOv8 model with different locations of added Simam attention mechanism.

3.5. Comparison before and after Data Enhancement

To improve the recognition performance of the new network model, this paper implements several strategies to tackle the issue of homogeneous backgrounds in the ginseng dataset. The model was evaluated using both the original dataset (1343 images) and the augmented dataset (4029 images), while keeping all other parameters consistent. Data augmentation, described in Section 2.2, was utilized to enhance the model's generalization and robustness. Data augmentation resulted in substantial improvements in model performance. As shown in Figure 13, precision increased from 87.1% to 88.42%, recall from 86.33% to 89.05%, and the F1 score from 86.5% to 88.73%. Additionally, mean Average Precision at 50% IoU (mAP50) improved from 91.89% to 95.3%, and the mAP across IoU thresholds (mAP50-95) rose from 71.77% to 74.19%. These improvements highlight the effectiveness of data augmentation in enhancing both model accuracy and robustness.

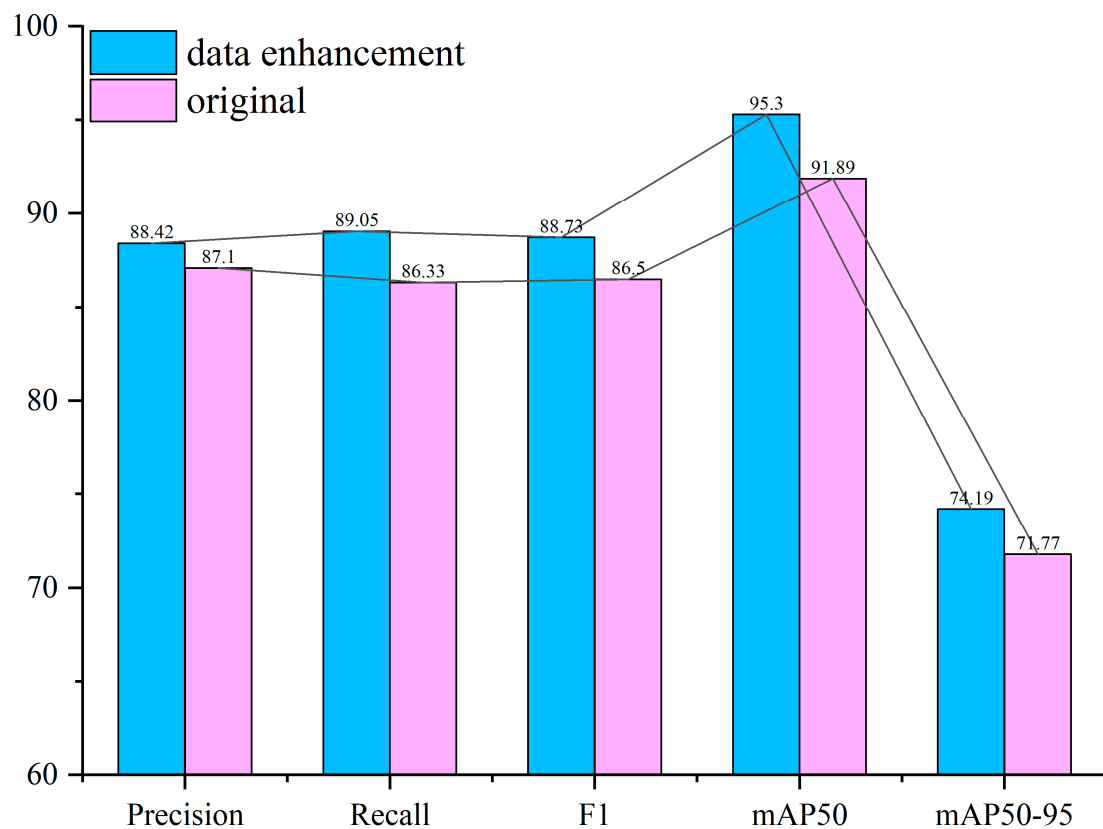


Figure 13. Comparison of data-enhanced metrics.

3.6. Comparison Experiment

To fully evaluate the performance of the DGS-YOLOv8 model, we systematically compared it with a range of state-of-the-art target detection models, including SSD, EfficientDet, and several versions of the YOLO family (YOLOv3, YOLOv3-Tiny, YOLOv5n, YOLOv5s, YOLOv7, YOLOv7-Tiny, YOLOv8n, and YOLOv10). Table 8 and Figure 14 of the experimental results show that DGS-YOLOv8 performs well on several key performance indicators and is highly feasible in practical applications. The performance of each model on five key performance indicators—precision, recall, F1 score, mAP50, and mAP50-95—were evaluated, and the number of parameters and file size of each model were analyzed. A detailed comparison reveals that DGS-YOLOv8 outperforms the other models in most performance metrics. Specifically, SSD, EfficientDet, and YOLOv3-Tiny significantly underperform compared to DGS-YOLOv8.

Table 8. Comparison experiments with other models.

Model	Precision (%)	Recall (%)	F1 (%)	mAP50 (%)	mAP50-95 (%)	Parameters (%)	Weight (MB)
SSD	58.9	73.61	65.44	87.3	58.7	14.34	48.1
EfficientDet	41.1	62.41	49.56	71.51	40.44	3.87	15.3
YOLOv3	73.91	81.09	77.34	85.40	55.69	61.5	123.6
YOLOv3-Tiny	36.91	55.37	44.28	37.62	11.59	8.67	17.5
YOLOv5n	80.09	85.53	82.74	89.1	55.5	1.78	3.9
YOLOv5s	84.61	84.89	84.66	91.2	59.2	7.03	14.5
YOLOv7	69.31	72.97	71.10	80.31	54.3	37.21	74.8
YOLOv7-Tiny	50.8	76.98	61.24	62	33.31	6.02	12.3
YOLOv8n	81.56	88.52	84.88	92.57	70.37	3.0	6.3
YOLOv10n	72.3	80.11	75.96	83.24	56.17	2.7	5.7
YOLOv10s	77.02	84.19	80.36	88.21	66.2	8.07	16.6
DGS-YOLOv8	88.42	89.05	88.73	95.3	74.19	2.86	6.0

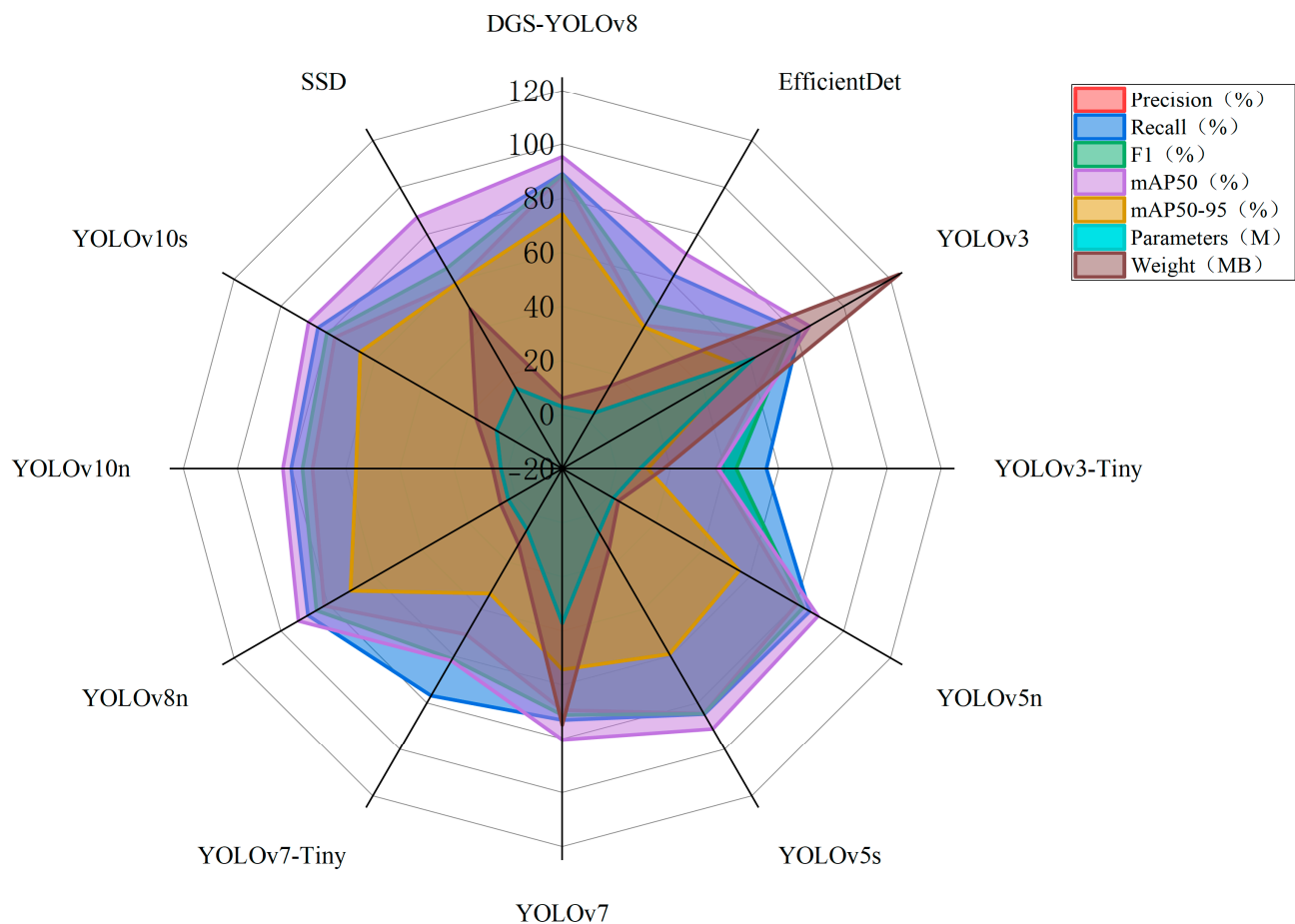


Figure 14. Comparison experiments with other models.

In contrast, YOLOv3 and YOLOv10s have vast numbers of parameters and file sizes, although they perform better in some metrics. For example, YOLOv3 has a precision of 73.91%, mAP50 of 85.40%, and mAP50-95 of 55.69%, while YOLOv10s has a recall of 84.19%, mAP50 of 88.21%, and mAP50-95 of 54.3%. These models' large number of parameters and file sizes limit their usefulness. Although YOLOv5n and YOLOv10n have fewer parameters, they still do not perform as well as DGS-YOLOv8 in some performance metrics. While YOLOv5s, YOLOv7-Tiny, and YOLOv8n perform better in some metrics, they still do not surpass DGS-YOLOv8 overall. The superior performance of DGS-YOLOv8 is mainly attributed to its adoption of specific features and improvements such as C2f-DCNv2, Slim-Neck, and SimAM. C2f-DCNv2, an improved convolutional structure, can better capture the detailed features of ginseng's appearance and quality during feature extraction, enhancing the detection accuracy of the model. Slim-Neck, a lightweight neck network structure, not only reduces the number of parameters and file size of the model but also maintains efficient feature fusion capability, thus improving the model's overall performance. By introducing the SimAM attention mechanism, the model can focus more on essential features when dealing with complex scenes, improving the accuracy and stability of detection.

Figure 15 shows the detection results of DGS-YOLOv8, SSD, EfficientDet, YOLOv3, YOLOv3-Tiny, YOLOv5n, YOLOv5s, YOLOv7, YOLOv7-Tiny, YOLOv8n, and YOLOv10n on the test set dataset. The comparative analysis demonstrates the superior performance of DGS-YOLOv8 in terms of accuracy and stability. In contrast, SSD and EfficientDet show false positives. DGS-YOLOv8 provides higher confidence in detecting targets and significantly increases bounding box accuracy. Its bounding boxes are more compact and closely aligned with the target than other models.

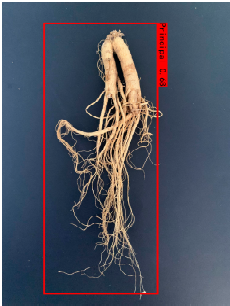
	Principal	First-class	Second-class
SSD			
EfficientDet			
YOLOv3			
YOLOv5n			
YOLOv5s			

Figure 15. Cont.



Figure 15. Pictures of detection results of different models.

4. Discussion

With the rapid development of machine learning and deep learning technologies, research on ginseng appearance and quality recognition has made remarkable progress in recent years. Traditional methods of ginseng classification, which rely heavily on manual inspection, are time-consuming, labor-intensive, and prone to human error. While existing CNN-based models have improved the situation, they often face challenges regarding

efficiency and deployment on edge devices. In this study, data augmentation techniques, such as simulated red rust, different lighting conditions, and complex backgrounds, were used to enhance the dataset and improve the robustness of the DGS-YOLOv8 model. This approach achieves reliable and precise results in ginseng testing, ensuring good model performance under various conditions. The proposed DGS-YOLOv8 model has significant advantages over manual classification and existing CNN-based models: (1) The DGS-YOLOv8 model demonstrates superior detection and classification performance, ensuring more consistent and reliable results when evaluating ginseng appearance and quality. This improvement highlights its effectiveness in practical applications compared to previous studies and methods. (2) The DGS-YOLOv8 model is optimized for computational efficiency, with smaller model sizes, fewer parameters, and faster processing times than traditional CNN models, making it suitable for real-time applications. Its compact size and efficiency make it ideal for deployment on edge devices with limited processing power, critical for agricultural applications requiring real-time decision making.

Despite these achievements, enhancing the model's generalization capabilities remains a crucial area for further research. The current validation of the model primarily involves a specific set of conditions and ginseng types, which may limit its broader applicability. To address this, future research should focus on the following: (1) Expanding the Dataset: To improve the model's adaptability, it is essential to expand the dataset to include a wider variety of ginseng types, such as American and red ginseng, which have distinct appearance characteristics. This expansion should also encompass samples from different geographic regions, cultivation techniques, and processing methods (e.g., steaming and drying). This diversity will help the model generalize better to various ginseng forms and environmental conditions. (2) Testing Under Diverse Conditions: Future studies should assess the model's performance under varying environmental conditions, including different lighting scenarios, seasonal changes, and background complexities. This will provide a more comprehensive understanding of how well the model adapts to real-world variations in ginseng appearance. (3) Incorporating Processing Variations: Including samples of ginseng processed through various methods will further enhance the model's robustness. Evaluating ginseng in different states, such as raw, steamed, and dried, will ensure that the model can accurately classify ginseng regardless of its preparation state. (4) Cross-Domain Applications: Exploring the model's applicability in related domains, such as other medicinal herbs with similar morphological characteristics, could provide insights into its versatility and effectiveness across different types of plants. This approach will contribute to developing a more generalizable model for broader agricultural applications. By addressing these aspects, the DGS-YOLOv8 model's generalization capabilities can be significantly improved, making it more effective across various conditions and types of ginseng.

5. Conclusions

This study adopted the YOLOv8 architecture as the foundational framework and implemented a series of modifications to enhance the performance of the DGS-YOLOv8 model. Using data augmentation techniques such as simulating red rust, varying light conditions, and complex backgrounds, the standard appearance features of ginseng and its surrounding environments were emulated. Through rigorous experimental validations, substantial improvements were observed in the DGS-YOLOv8 model. Key performance indicators of the DGS-YOLOv8 model are as follows: precision is 88.42%, recall is 89.05%, F1 score is 88.73%, mAP50 is 95.3%, mAP50-95 is 74.19%, the model size is 2.86 million parameters, and the model weight is 6.0 MB. These results demonstrate the model's robust detection and classification capabilities. Compared to conventional detection models such as SSD, EfficientDet, and various iterations of the YOLO series (including YOLOv3, YOLOv3-Tiny, YOLOv5n, YOLOv5s, YOLOv7, YOLOv7-Tiny, YOLOv8n, and YOLOv10n), the optimized DGS-YOLOv8 model exhibited superior performance. This enhanced model is especially beneficial for agricultural applications due to its high computational efficiency

and fast processing times, critical for real-time decision making. The significance of this work lies in its ability to provide a practical solution for the accurate and automatic detection of ginseng appearance and quality grades, addressing the limitations of traditional manual methods and existing CNN-based models. The innovation of DGS-YOLOv8 lies in its ability to operate efficiently on edge devices, making it suitable for deployment in field conditions where real-time analysis is crucial. This capability enhances the accuracy and reliability of crop monitoring and management, contributing significantly to intelligent agriculture's advancement. The objective is to continue refining and optimizing the DGS-YOLOv8 model. Expanding the variety and quantity of samples aims to bolster the model's robustness and further promote the development of precision agriculture technologies. Future research will focus on integrating additional environmental factors and diverse ginseng types to enhance the model's adaptability and effectiveness. This ongoing research aims to provide robust technical support for advancing intelligent agriculture, contributing to more efficient and effective farm management practices.

Author Contributions: Conceptualization, L.Z. and Z.L.; Data curation, H.Y. and D.L.; Formal analysis, L.Z. and Y.L.; Funding acquisition, Y.L.; Investigation, Z.W.; Methodology, H.Y. and S.Y.; Project administration, Y.L.; Resources, Z.W. and H.J.; Software, Z.L.; Validation, L.Z. and H.J.; Visualization, H.Y.; Writing—original draft, H.Y.; Writing—review and editing, D.L. and C.Z. All authors will be informed about each step of manuscript processing, including submission, revision, revision reminder, etc., via emails from our system or assigned Assistant Editor. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Natural Science Foundation of China (61806024, 62206257); Jilin Province Science and Technology Development Plan Key Research and Development Project (20210204050YY); Wuxi University Research Start-up Fund for Introduced Talents (2023r004, 2023r006). We thank the anonymous reviewers for their helpful and constructive comments.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The datasets generated for this study are available upon request from the corresponding author.

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

References

1. Liu, L.; Hu, J.; Mao, Q.; Liu, C.; He, H.; Hui, X.; Yang, G.; Qu, P.; Lian, W.; Duan, L. Functional compounds of ginseng and ginseng-containing medicine for treating cardiovascular diseases. *Front. Pharmacol.* **2022**, *13*, 1034870. [[CrossRef](#)] [[PubMed](#)]
2. Loo, S.; Kam, A.; Dutta, B.; Zhang, X.; Feng, N.; Sze, S.K.; Liu, C.-F.; Wang, X.; Tam, J.P. Broad-spectrum ginsentides are principal bioactives in unraveling the cure-all effects of ginseng. *Acta Pharm. Sin. B* **2024**, *14*, 653–666. [[CrossRef](#)] [[PubMed](#)]
3. Pang, Y.; Tian, X.; Wang, D.; Wang, H. Species authentication of *Panax ginseng* CA Mey. and ginseng extracts using mitochondrial nad2 intron 4 region. *J. Appl. Res. Med. Aromat. Plants* **2024**, *41*, 100554.
4. Lee, K.-Y.; Shim, S.-L.; Jang, E.-S.; Choi, S.-G. Ginsenoside stability and antioxidant activity of Korean red ginseng (*Panax ginseng* CA meyer) extract as affected by temperature and time. *LWT* **2024**, *200*, 116205. [[CrossRef](#)]
5. Fan, W.; Fan, L.; Wang, Z.; Mei, Y.; Liu, L.; Li, L.; Yang, L.; Wang, Z. Rare ginsenosides: A unique perspective of ginseng research. *J. Adv. Res.* **2024**; *in press*.
6. Zhang, Z.; Chen, X.; Zhang, K.; Zhang, R.; Wang, Y. Research on the current situation of ginseng industry and development counter-measures in Jilin Province. *J. Jilin Agric. Univ.* **2023**, *45*, 649–655.
7. Fang, J.; Xu, Z.-F.; Zhang, T.; Chen, C.-B.; Liu, C.-S.; Liu, R.; Chen, Y.-Q. Effects of soil microbial ecology on ginsenoside accumulation in *Panax ginseng* across different cultivation years. *Ind. Crops Prod.* **2024**, *215*, 118637. [[CrossRef](#)]
8. Ye, X.-W.; Li, C.-S.; Zhang, H.-X.; Li, Q.; Cheng, S.-Q.; Wen, J.; Wang, X.; Ren, H.-M.; Xia, L.-J.; Wang, X.-X.; et al. Saponins of ginseng products: A review of their transformation in processing. *Front. Pharmacol.* **2023**, *14*, 1177819. [[CrossRef](#)] [[PubMed](#)]
9. Ryu, J.Y.; Kim, H.U.; Lee, S.Y. Deep learning improves prediction of drug–drug and drug–food interactions. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, E4304–E4311. [[CrossRef](#)]
10. Li, D.; Yang, C.; Yao, R.; Ma, L. Origin Identification of *Saposhnikovia divaricata* by CNN Embedded with the Hierarchical Residual Connection Block. *Agronomy* **2023**, *13*, 1199. [[CrossRef](#)]
11. Kim, M.; Kim, J.; Kim, J.S.; Lim, J.-H.; Moon, K.-D. Automated Grading of Red Ginseng Using DenseNet121 and Image Preprocessing Techniques. *Agronomy* **2023**, *13*, 2943. [[CrossRef](#)]

12. Li, D.; Piao, X.; Lei, Y.; Li, W.; Zhang, L.; Ma, L. A Grading Method of Ginseng (*Panax ginseng* C. A. Meyer) Appearance Quality Based on an Improved ResNet50 Model. *Agronomy* **2022**, *12*, 2925. [\[CrossRef\]](#)
13. Li, D.; Zhai, M.; Piao, X.; Li, W.; Zhang, L. A Ginseng Appearance Quality Grading Method Based on an Improved ConvNeXt Model. *Agronomy* **2023**, *13*, 1770. [\[CrossRef\]](#)
14. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
15. Chen, X.; Liu, T.; Han, K.; Jin, X.; Wang, J.; Kong, X.; Yu, J. TSP-yolo-based deep learning method for monitoring cabbage seedling emergence. *Eur. J. Agron.* **2024**, *157*, 127191. [\[CrossRef\]](#)
16. Yang, S.; Wang, W.; Gao, S.; Deng, Z. Strawberry ripeness detection based on YOLOv8 algorithm fused with LW-Swin Transformer. *Comput. Electron. Agric.* **2023**, *215*, 108360. [\[CrossRef\]](#)
17. Liu, Z.; Rasika, D.; Abeyrathna, R.M.; Mulya Sampurno, R.; Massaki Nakaguchi, V.; Ahamed, T. Faster-YOLO-AP: A lightweight apple detection algorithm based on improved YOLOv8 with a new efficient PDWConv in orchard. *Comput. Electron. Agric.* **2024**, *223*, 109118. [\[CrossRef\]](#)
18. Ma, L.; Yu, Q.; Yu, H.; Zhang, J. Maize Leaf Disease Identification Based on YOLOv5n Algorithm Incorporating Attention Mechanism. *Agronomy* **2023**, *13*, 521. [\[CrossRef\]](#)
19. Jiang, M.; Liang, Y.; Pei, Z.; Wang, X.; Zhou, F.; Wei, C.; Feng, X. Diagnosis of breast hyperplasia and evaluation of RuXian-I based on metabolomics deep belief networks. *Int. J. Mol. Sci.* **2019**, *20*, 2620. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Liu, Y.; Li, Y.; Zhao, Y.; Na, X. Image Classification and Recognition of Medicinal Plants Based on Convolutional Neural Network. In Proceedings of the 2021 IEEE 21st International Conference on Communication Technology (ICCT), Tianjin, China, 13–16 October 2021; pp. 1128–1133.
21. Lu, J.; Wu, W. Fine-grained image classification based on attention-guided image enhancement. *Proc. J. Phys. Conf. Ser.* **2021**, *1754*, 012189. [\[CrossRef\]](#)
22. Terven, J.; Córdova-Esparza, D.-M.; Romero-González, J.-A. A comprehensive review of yolo architectures in computer vision: From yolov1 to yolov8 and yolo-nas. *Mach. Learn. Knowl. Extr.* **2023**, *5*, 1680–1716. [\[CrossRef\]](#)
23. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Michael, K.; Fang, J.; Yifu, Z.; Wong, C.; Montes, D. ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation. *Zenodo* **2022**. Available online: <https://ui.adsabs.harvard.edu/abs/2022zndo...7347926j/abstract> (accessed on 11 August 2024).
24. Ge, Z.; Liu, S.; Wang, F.; Li, Z.; Sun, J. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
25. Zhu, X.; Hu, H.; Lin, S.; Dai, J. Deformable convnets v2: More deformable, better results. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9308–9316.
26. Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhan, Z.; Ren, Q. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles. *arXiv* **2022**, arXiv:2206.02424.
27. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
28. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 11863–11874.
29. Woo, S.; Park, J.; Lee, J.-Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
30. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13713–13722.
31. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
32. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
33. Liu, Y.; Shao, Z.; Hoffmann, N. Global attention mechanism: Retain information to enhance channel-spatial interactions. *arXiv* **2021**, arXiv:2112.05561.
34. Liu, Y.; Shao, Z.; Teng, Y.; Hoffmann, N. NAM: Normalization-based attention module. *arXiv* **2021**, arXiv:2111.12419.
35. Selvaraju, R.R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; Batra, D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 618–626.

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.