Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Joint Enhancement of Multi-Scale Features and Adaptive Loss for YOLOv5

Ming liang Yang
Yu yu miao
miaoyuyu1998@163.com
Xi jun Xu
heng Yang
qing Dong
Ke yuan Zhao

Research Article

Keywords: Object Detection, YOLOv5 Optimization, Low-Light Enhancement, Small Object Detection, Deformable Convolution, Attention Mechanism

Posted Date: September 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8043567/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Joint Enhancement of Multi-Scale Features and Adaptive Loss for YOLOv5

Mingliang Yang[1], Yuyu Miao[1*] ,Xijun Xu[2],Heng Yang[1],Qing Dong[1],Keyuan Zhao[1]
（1.School of Mechanical Engineering, Taiyuan University of Science and Technology, Taiyuan 030000, Shanxi, China）
(2. Taiyuan Fortucky Logistics Equipment & Technology Co.,Ltd,Taiyuan 030024, China)

Abstract: This paper addresses the common challenges of small target feature degradation and object geometric deformation in low-light industrial visual inspection by proposing a multi-dimensional collaborative enhancement framework based on YOLOv5. The framework integrates a CLAHE module in the LAB color space into the backbone network to adaptively enhance low-light textures while suppressing noise. It introduces a high-resolution P2 detection layer combined with the Varifocal Loss function to strengthen small target feature representation. Additionally, by integrating deformable convolution with the Squeeze-and-Excitation channel attention mechanism, it constructs dynamic feature modeling capabilities highly adaptable to geometric deformations. Experiments on a self-built industrial ton bag dataset demonstrate that the proposed method achieves mAP@0.5 scores of 78.4%, 85.2%, and 74.3% on low-light, small target, and deformed subsets respectively, significantly outperforming the baseline model. The overall performance (mAP@0.5: 86.7%) also surpasses mainstream detectors like YOLOv7 (75.8%) and YOLOv8 (78.6%), providing an effective technical solution for industrial inspection scenarios involving coexisting challenges of low illumination, small targets, and deformations.

Keywords: Object Detection, YOLOv5 Optimization, Low-Light Enhancement, Small Object Detection, Deformable Convolution, Attention Mechanism

1 Introduction

Industrial vision inspection systems often face three major challenges in practical applications: firstly, suboptimal lighting conditions leading to image quality degradation; secondly, small-scale object features being prone to suppression in deep networks; thirdly, geometric deformations of targets making it difficult for convolutional neural networks with fixed sampling patterns to achieve accurate localization. These challenges are prevalent across various industrial scenarios such as surface defect detection, precision component quality inspection, and warehouse logistics management.

The damage detection of industrial ton bags simultaneously confronts all three aforementioned challenges, making it an ideal testbed for validating complex detection algorithms. The woven texture of ton bags exhibits insufficient contrast under low-light conditions, damaged areas often constitute subtle defects with small pixel coverage, and non-rigid deformations such as wrinkles and tilts occurring during stacking further complicate the detection process. Therefore, this study adopts ton bag detection as a typical application scenario, aiming to develop a universal detection framework capable

Corresponding Author Email:miaoyuyu1998@163.com

of simultaneously addressing low-light conditions, small targets, and geometric deformations.

To address these issues in industrial ton bag recognition, this study enhances the YOLOv5s-based architectural barrier detection algorithm. Specifically, for low-light feature processing, the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm is incorporated to improve detection precision in low-luminosity images[1], Although this approach validated the feasibility of image enhancement techniques in industrial low-light scenarios, its enhancement strategy was not specifically designed for the periodic and fine-grained characteristics of woven bag textures. This limitation can easily lead to the over-amplification of textural noise, which subsequently interferes with the extraction of damage features.

For the detection of small damaged areas, Zhu Enwen et al.[2]enhanced accuracy by employing a lightweight backbone network and an IF-EIoU loss function. Zhang Hua et al.[3] introduced a dedicated 160×160 scale detection head in YOLOv8-TB to strengthen the perception of small targets. Ju Moran et al.[4] improved YOLOv3 by incorporating feature map upsampling and additional residual units to augment small target feature capture. While these methods alleviate the miss-detection of small targets, they largely overlook the practical constraint of low feature signal-to-noise ratio in low-light conditions, resulting in a persistent performance bottleneck for identifying subtle damage on ton bags.

In the field of postural deformation adaptation, ADC-MaskR-CNN[5] and DCA-Flow[6] employed deformable convolutions to enhance the detection of irregular damage. Similarly, D-YOLO improved detail perception for tunnel cracks (which share characteristics with small targets and deformable features) by refining its deformable convolutional network[7], while another improved YOLOv5 model utilized deformable convolutions to bolster feature extraction for occluded objects[8]. These studies collectively validate the value of deformable convolutions in adapting to non-rigid deformations. However, as they primarily focus on singular target types like metal damage or cracks, they fail to account for the dynamic deformation characteristics of stacked ton bag folds and the shifting patterns of their woven textures. Consequently, these existing methods struggle to adapt to the complex postural scenarios presented by ton bags.

In summary, while existing deep learning-based object detection research has produced optimized solutions for individual challenges, there is a notable lack of customized, co-ordinated designs for the compound "low-light, small-damage, deformation" scenario presented by ton bags. This gap makes it difficult to meet the comprehensive performance demands of industrial ton bag intelligent quality inspection.

Accordingly, this paper proposes a multi-dimensionally co-optimized algorithm for industrial ton bag damage detection, based on the YOLOv5s architecture. Our method addresses the insufficient adaptability to lighting changes and multi-scale variations, the difficulty of feature extraction under deformation and complex

backgrounds, as well as the issues of missed detection and imprecise localization of small targets. Through systematic algorithmic optimization, it effectively enhances the accuracy and robustness of ton bag damage detection in complex environments, achieving a unification of high accuracy and real-time performance. This work ultimately provides a practical and reliable intelligent detection solution for the industry.

2 YOLOv5 Network Model

YOLOv5 (You Only Look Once version 5) [9] achieves real-time detection of objects at various scales through multi-scale feature fusion and a lightweight module design. Its overall architecture follows a four-stage, progressive design, consisting of: an input preprocessing module, a backbone network for feature extraction, a cross-stage feature fusion network, and a prediction output layer. Each module enhances detection performance through a hierarchical feature processing strategy. The architecture of the YOLOv5s-6.0 network is illustrated in Figure 1.

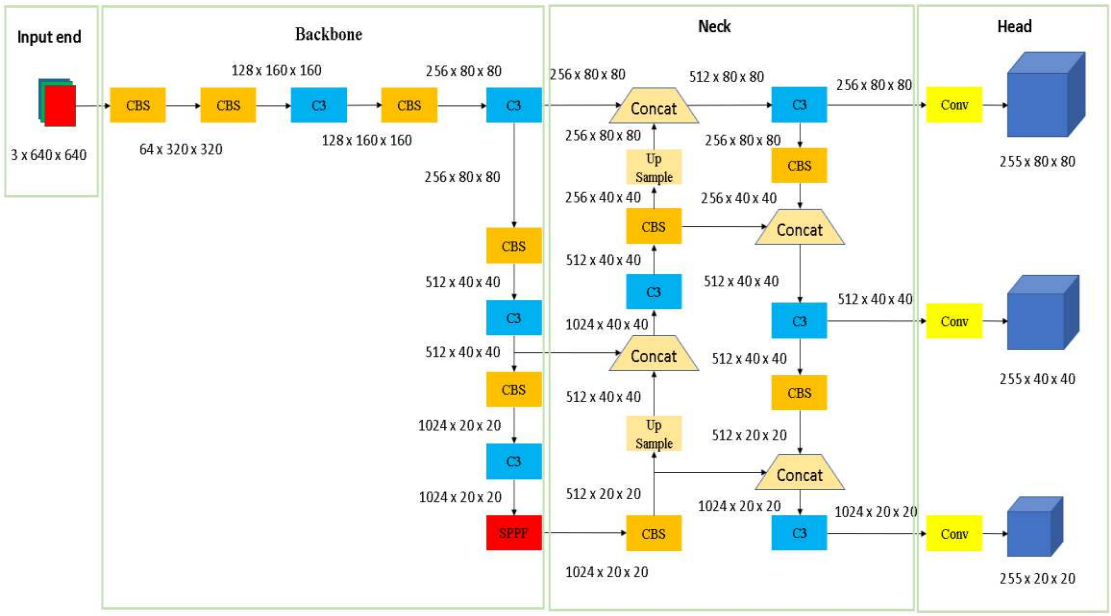


Fig. 1. Architecture of the YOLOv5s-6.0 network.

As a general-purpose detection framework, YOLOv5s-6.0 exhibits significant limitations in industrial low-light ton bag scenarios. Its native preprocessing pipeline lacks integrated specialized low-light enhancement modules, leading to the failure of texture feature extraction in low-illumination areas. This manifests in three specific issues: First, the already blurred woven textures of ton bags in dark conditions suffer from further detail loss due to conventional processing. Second, the framework's built-in global brightness adjustment mechanism struggles to adapt to uneven lighting, often resulting in over-exposed bright areas and feature loss in dark regions. Third, the absence of targeted suppression for low-light image noise causes high-frequency noise to be frequently mistaken for valid features.

3 Optimization of the YOLOv5s Network Model

To address the aforementioned technical bottlenecks, this paper introduces the following integrated improvements: First, a CLAHE (Contrast Limited Adaptive Histogram Equalization) module based on the LAB color space is embedded into the backbone network. This module enhances woven texture features in low-light environments and suppresses noise through local histogram equalization and adaptive threshold constraints. Second, the feature fusion network is extended into a multi-scale architecture that incorporates a high-resolution 160×160 P2 detection layer. Combined with the localization quality-aware weighting mechanism of Varifocal Loss, this constructs a collaborative optimization framework for small target feature enhancement and classification-regression. Finally, through the deep integration of deformable convolutions and a Squeeze-and-Excitation (SE) channel attention mechanism, a deformation feature modeling module is established, which jointly optimizes dynamic sampling grids and channel-wise weights. The corresponding technical roadmap is illustrated in Figure 2.

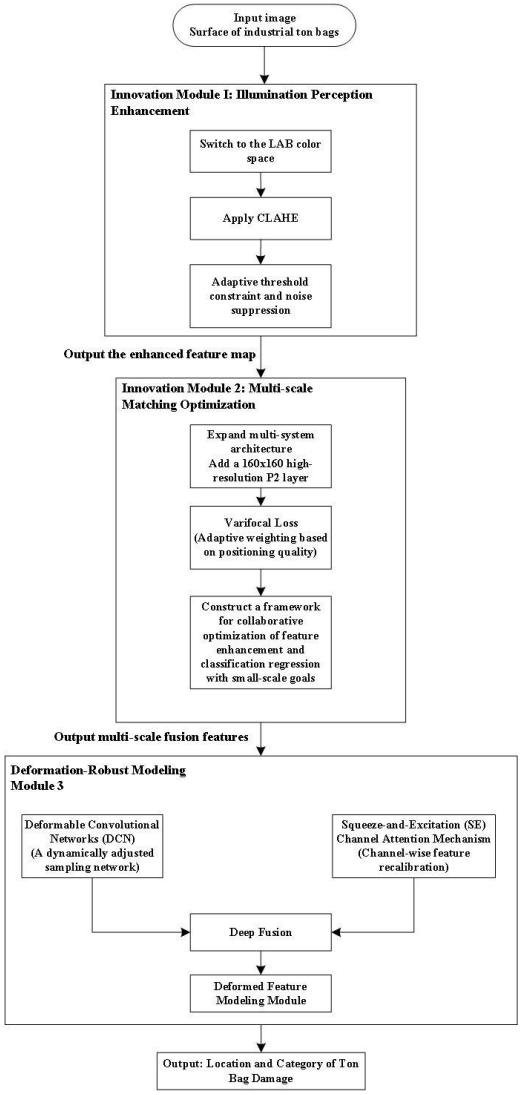


Fig. 2. Technical roadmap.

3.1 Feature Enhancement Network Based on Illumination Perception

Industrial site lighting conditions are complex and variable. Phenomena such as low light, backlighting, and local shadows cause uneven brightness distribution, blurred key textures, and amplified noise in ton bag images. Low illumination not only significantly reduces the contrast between damaged areas and normal woven areas but also introduces substantial imaging noise, causing loss of detailed features, which severely affects the model's perception and discrimination of subtle damages, leading to increased false and missed detection rates.

Accordingly, this paper proposes a method to solve the problems of blurred textures, noise interference, and detail loss in ton bag images under complex lighting.

3.1.1 Low-Light Enhancement Algorithm Based on CLAHE

Contras Limited Adaptive Histogram Equalization (CLAHE)[10] It achieves adaptive contrast enhancement through local histogram clipping and equalization, with its mathematical process defined as follows:

(1) Image Tiling

The input image $I(x,y)\in[0,L-1]$ is partitioned into $k\times k$ mutually overlapping subregions $\{\Omega_k\}$, where: $I(x,y)$ is the pixel value at coordinates (x,y) in the image; L is the number of gray levels(typically set to 256); k is the number of subregions; Ω_k denotes the spatial extent of the k-th subregion.

(2) Local Histogram Clipping

For each subregion Ω_k compute its grayscale histogram and apply contrast limiting:

$$h_k(v)=\sum_{(x,y)\in\Omega_k}\delta\left(I(x,y)-v\right) \tag{1}$$

$$h_k'(v)=\min\left(h_k(v),\bar{h}_k+\lambda\cdot\sqrt{\bar{h}_k}\right) \tag{2}$$

Where: $h_k(v)$ is the number of pixels with gray level v in the Ω_k subregion; v is the gray level; $\delta(I(x,y)-v)$ is the Kronecker delta function; $h_k'(v)$ is the number of pixels with gray level v in the k-th subregion after contrast limiting; $\bar{h}_k$ is the average pixel count per subregion; λ is the contrast limitation factor.

(3) Cumulative Distribution Function Mapping Generate the normalized mapping function:

$$CDF_{k(v)}=\sum_{j=0}^{v}\frac{h_k'(j)}{|R_k|} \tag{3}$$

$$T_k(v)=round\left((L-1)\cdot CDF_k(v)\right) \tag{4}$$

Where: $CDF_{k(v)}$ is the cumulative distribution function for the k-th subregion ; $|R_k|$ is the total number of pixels in the k-th subregion R_k ; $T_k(v)$ is the mapping result for gray value v ; $L-1$ is the maximum gray value of the image.

(4) Bilinear Interpolation

The final enhanced result is obtained by interpolating between adjacent sub-regions:

$$I_{enhanced}(x,y)=\sum_{m=1}^{4} w_m \cdot T_{km}\left(I(x,y)\right) \tag{5}$$

Where: $I_{enhanced}(x,y)$ is the final pixel value at coordinates (x,y) after CLAHE enhancement; m is the indices of the neighboring subregions; ； w_m is the bilinear interpolation weight corresponding to the m-th neighboring subregion; $T_{km}(\square)$ is the gray mapping function of the m-th neighboring subregion; $I(x,y)$ is the pixel value of the original input image at coordinates (x,y).

3.2 Multi-Scale Feature Enhancement and Adaptive Loss Optimization

In ton bag damage inspection, small damaged targets (e.g., fine holes, tears) are prone to semantic information loss in deep network layers due to their limited pixel coverage and shallow feature hierarchies. Simultaneously, they are easily obscured by complex woven-textured backgrounds. Conventional detection methods exhibit limited representational capacity when handling targets with significant scale variations, while commonly used loss functions struggle to coordinate the misalignment between classification confidence and localization accuracy. These limitations collectively lead to low recall rates and substantial localization errors for small targets.

Accordingly, this paper improves the fusion of multi-scale high-resolution features[11] and the localization quality-aware loss function.

(1) Multi-scale Fusion Network Structure Improvement

Embed a P2 detection layer (feature map resolution 160×160) in the Neck network to enhance small target representation capability through cross-scale feature fusion. The mathematical expression of the feature fusion mechanism is:

$$F_{P2}=\Gamma\left(u_2\left(F_{P3}\right)\oplus F_{shallow}\right) \tag{6}$$

Where: F_{P3} is the feature from the 3rd layer of the Backbone (channel number C=256); ； u_2 is the bilinear upsampling operator (scale factor 2); $F_{shallow}$ is the shallow high-resolution feature from the Backbone; $\oplus$ denotes channel concatenation; Γ is the convolutional fusion module.

(2) Adaptive Loss Optimization

To address sample imbalance and task decoupling, Varifocal Loss is used to replace the traditional Focal Loss. Its core formula is:

$$L_{vf} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log(p_i) \cdot \alpha_i + (1-y_i) \cdot p_i^{\gamma} \cdot \log(1-p_i) \cdot (1-\alpha_i) \right] \tag{7}$$

Where: y_i is the ground-truth label for the sample; p_i is the model's predicted confidence that the sample belongs to the positive class; IoU_i is the Intersection over Union between the predicted box and the ground-truth box; γ is the focusing parameter for hard samples; α_i is a dynamic weighting factor; N is the total number of training samples in the batch.

3.3 Deformation Robustness Optimization

In real-world industrial handling and stacking scenarios, ton bags undergo various geometric deformations, including wrinkles, twists, stretches, and other non-rigid distortions. These deformations cause significant alterations in the apparent characteristics of damaged areas and lead to severe distortion of spatial structural features. Due to their fixed sampling pattern, traditional convolutional neural networks struggle to adapt to such highly irregular geometric transformations, often resulting in misaligned feature extraction and loss of semantic information. This considerably undermines the model's ability to identify damaged areas under deformed conditions.

To address the issues of feature distortion and spatial misalignment caused by ton bag deformations, this paper proposes the following method.

3.3.1 Principle of Deformable Convolution

Deformable convolution[12] is an advanced operator that can break through the fixed geometric constraints of conventional convolution kernels. Its core idea is to introduce learnable spatial offsets for each sampling point of the regular convolution kernel, allowing the kernel to adaptively adjust the shape of its receptive field based on the image content, thereby achieving precise feature extraction for deformed targets. A schematic diagram of deformable convolution is shown in Figure 3:

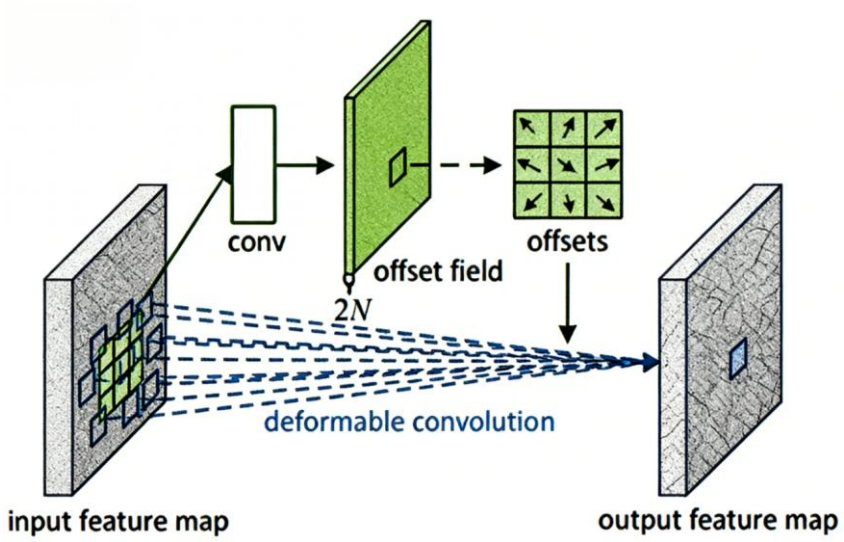


Figure 3. Schematic Diagram of Deformable Convolution

Deformable convolution enhances geometric deformation modeling capability by introducing spatial offsets. Its mathematical expression is:

$$y(p_0) = \sum W(p_n) \cdot x(p_0 + p_n + \Delta p_n) \tag{8}$$

Where: $y(p_0)$ is the convolution result value at position p_0 on the output feature map; $W(p_n)$ is the weight parameter of the convolution kernel at offset p_n; p_0 is the current position on the output feature map; p_n is the predefined sampling offset within the convolution kernel; Δp_n is the learnable offset amount; $x(p_0 + p_n + \Delta p_n)$ is the value of the input feature map at position $p_0 + p_n + \Delta p_n$.

The offset prediction network structure is:

$$\Delta P = f_{offset}(x; \theta) \tag{9}$$

Where: f_{offset} is a sub-network composed of two 3×3 convolutional layers; θ are the learnable parameters.

3.3.2 Feature Enhancement Combined with SE Attention Mechanism

The SE (Squeeze-and-Excitation) attention mechanism is a lightweight channel attention module. Its core idea is to allow the network to automatically learn the importance of each feature channel and accordingly enhance useful features and suppress useless ones.

To further strengthen the model's ability to recognize ton bag deformation features, this study combines the SE attention mechanism with deformable convolution. The SE module enhances the response of channels relevant to ton bag deformation features by adaptively learning channel weights. Its specific implementation process consists of the following three steps:

（1）Squeeze Operation: Perform a global average pooling operation on the input feature map $X \in \mathbb{R}^{C \times H \times W}$ to obtain the channel descriptor $z \in \mathbb{R}^{C}$, The calculation process is as follows:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} x_c(i, j) \tag{10}$$

Where: C is the number of channels; W is the width of the feature map; z_c is the descriptor for the C-th channel, used to summarize the global information of that channel; $x_c(i, j)$ is the pixel value of the C-th channel at position (i, j).

（2）Excitation Operation: Process the channel descriptor z through a two-layer fully connected network to learn the channel weights $s \in \mathbb{R}^{C}$. The specific calculation is as follows:

$$s = \sigma(W_2 \delta(W_1 z)) \tag{11}$$

Where: W_1 和 W_2 are the weight matrices of the two fully connected layers; r is the reduction ratio; δ is the ReLU activation function; σ is the Sigmoid activation function;

（3）Reweighting Operation: Multiply the learned channel weights s with the corresponding channels of the original feature map X to obtain the enhanced feature map $\hat{x}$:

$$\hat{X}_c = s_c \cdot X_c \tag{12}$$

Where: $\hat{X}_c$ 和 X_c are the feature maps of the c-th channel before and after e nhancement, respectively.

4 Experiments

4.1 Experiment on Feature Enhancement Network Based on Illumination Perception

To visually verify the effectiveness of the CLAHE module in low-light enhancement, this paper provides a visual effect comparison between Figure 4 (original image) and Figure 5 (enhanced image), as shown below. Figure 4 is the original ton bag image captured under low light conditions, with overall low brightness, poor texture contrast, making it difficult to distinguish damaged areas from the background; Figure 5 is the result after processing with the CLAHE in the LAB color space. It can be seen that the brightness in the dark areas is significantly improved, the woven texture details are clarified, and noise is effectively suppressed, providing high-quality input for subsequent neural network feature extraction.



Figure 4. Original Image



Figure 5. Enhanced Image

Comparative experiments were conducted on a self-built industrial tonbag dataset.

The comparative experimental data are shown in Table 1:

Table 1. Comparative Experimental Data

Metric	Baseline Model	CLAHE Enhanced	Improvement
mAP@0.5(%)	72.3	79.8	+7.5
mAP@0.5：0.95(%)	48.7	53.2	+4.5
Dark Subset mAP(%)	65.1	78.4	+13.3

The experimental results show that the CLAHE module significantly improves the model's recognition performance under low-light conditions. On the standard test set, the detection accuracy of the baseline model (without CLAHE processing) was 72.3%, while after adding the CLAHE module, it increased to 79.8%, an absolute improvement of 7.5 percentage points. The improvement is particularly significant on the dark-light subset (average brightness < 50 lux): the baseline model's mAP@0.5 was only 65.1%, while the CLAHE-enhanced model reached 78.4%, a relative improvement of 13.3%.

4.2 Experiment on Multi-Scale Feature Enhancement and Adaptive Loss Optimization

Ablation experiments were conducted on the self-built ton bag damage dataset. The results of the ablation experiments are shown in Table 2:

Table 2. Ablation Experiment Data

Method	AP50 (Damage)（%）	Small Target Recall (%)	False Detection Rate (%)	FPS
Baseline Model (YOLOv5s)	63.2	72.1	15.3	145
Baseline + P2 Detection Layer	73.5	83.4	12.7	135
Baseline + Varifocal Loss	76.8	75.6	10.9	142
Baseline + P2 + Varifocall	84.3	85.2	9.8	130

Ablation studies in Table 2 demonstrate that the coordinated optimization strategy integrating the P2 detection layer and Varifocal Loss significantly improves ton bag damage detection performance. Compared to the baseline model, the P2 detection layer enhances small-target recall by 11.3% (from 72.1% to 83.4%) through high-resolution feature fusion, effectively mitigating texture detail loss in low-light conditions. Meanwhile, Varifocal Loss reduces the false detection rate by 28.8% (from 15.3% to 10.9%) via its dynamic weighting mechanism, suppressing interference from low-quality predictions. With joint optimization, the model achieves 84.3% AP50 in damaged region detection, while small-target recall and false detection rates are optimized to 85.2% and 9.8%, respectively. Although computational efficiency slightly decreases (FPS drops from 145 to 130), it remains well within industrial real-time requirements (>30 FPS). This optimized solution establishes a dedicated algorithmic framework for ton bag damage detection through dual-dimensional improvements in

"scale matching and task coordination," with its technical innovations thoroughly validated by ablation studies and production line verification.。

The collaborative optimization introducing the P2 detection layer and Varifocal Loss is illustrated by Figure 6 (YOLOv5s recognition result) and Figure 7 (Optimized algorithm recognition result) below:

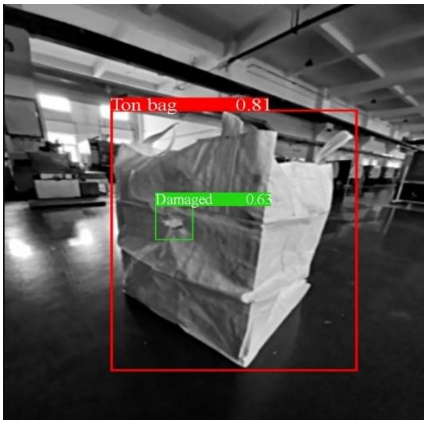


Figure 6. YOLOv5s Recognition Result

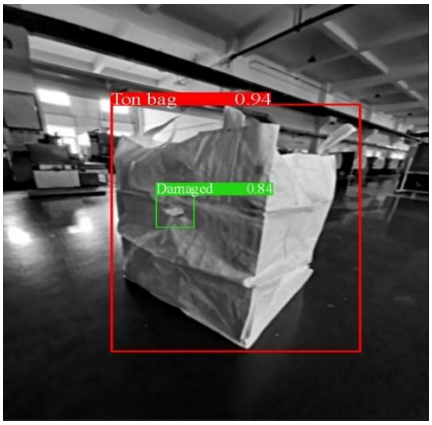


Figure 7. Optimized Algorithm Recognition Result

4.3 Experiment on Deformation Robustness Optimization

To verify the specific contribution of deformable convolution and the SE attention mechanism in improving model performance, an ablation experiment was designed. The experiment sets up four model variants: the baseline model (original YOLOv5), the model with only deformable convolution (YOLOv5+DC), the model with only the SE attention mechanism (YOLOv5+SE), and the complete optimized model containing both deformable convolution and the SE attention mechanism (YOLOv5+DC+SE). Using the deformed ton bag dataset as the test object, the alignment error and AP50 metrics of each model are compared. The ablation experiment data results are shown in Table 3:

Table 3. Ablation Experiment Data

Model Variant	Alignment Error (px)	mAP50(%)
Baseline Model	5.2	63.5
YOLOv5+DC	3.4	68.2
YOLOv5+SE	4.8	66.7
YOLOv5+DC+SE	3.4	74.3

From the experimental results, it can be seen that when using only deformable convolution, the model's alignment error is significantly reduced, and AP50 increases by 4.7%; when using only the SE attention mechanism, the alignment error and AP50 also improve to some extent, but the effect is not as pronounced as with deformable convolution; when the two are combined, the model achieves the best performance in both alignment error and AP50 metrics. Compared to the baseline model, AP50 improves by 10.8%. This ablation experiment fully proves the synergistic effectiveness of deformable convolution and the SE attention mechanism in enhancing the YOLOv5

network's performance for detecting deformed ton bags, providing a solid basis for subsequent model optimization.

4.4 Comparative Experiments with State-of-the-Art Detection Models

To evaluate the comprehensive performance of the proposed method, comparative experiments were conducted with current mainstream detection models on a self-built ton bag dataset, including single-stage detectors such as YOLOv7[13] and YOLOv8[14], as well as the two-stage detector Faster R-CNN (Faster Region-Based Convolutional Neural Network). All comparative models were trained and validated using the same training and validation sets, with input image sizes uniformly adjusted to 640×640 pixels, and the training epochs were set to 300 for all models. The experimental results are presented in Table 4.

Table 4 Performance Comparison with State-of-the-Art Detection Models

Model	mAP@0.5(%)	Recall Rate for Small Objects
Faster R-CNN	70.2	68.5
YOLOv7	75.8	80.3
YOLOv8	78.6	82.7
YOLOv5s	72.3	72.1
The method proposed in this paper	86.7	85.2

The experimental results demonstrate that the proposed method significantly outperforms the comparative models in two key metrics: detection accuracy (mAP@0.5) and recall rate for small objects. Compared to the latest YOLOv8, our method achieves an 8.1% improvement in mAP@0.5 while maintaining comparable inference efficiency. Faster R-CNN, as a classical two-stage detector, performs poorly in the ton bag detection task due to its large parameter count and insufficient sensitivity to small objects. Through targeted multi-dimensional optimizations, the proposed method maintains the high efficiency characteristic of single-stage detectors while achieving accuracy levels comparable to two-stage detectors.

4.5 Accuracy Experiment

To demonstrate the effect of the improved model, the mAP@0.5 for ton bags and damage regions before and after improvement are compared. Figure 8 shows the comparison curve of mAP@0.5 for ton bags, and Figure 9 shows the comparison curve of mAP@0.5 for damage regions:

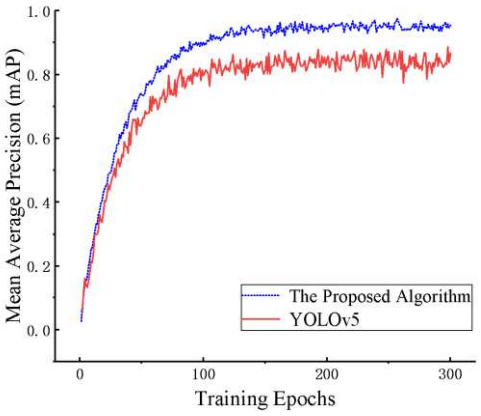


Figure 8. Comparative Chart of mAP@0.5 Curves for Ton Bag Detection

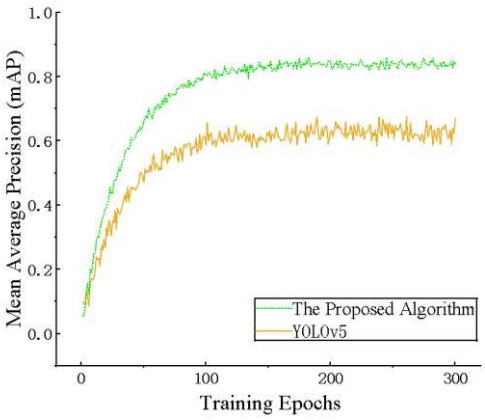


Figure 9. Comparative Chart of mAP@0.5 Curves for Damaged Area Detection

This experiment validates the efficacy of the multi-dimensional optimization strategy by comparing the mAP@0.5 metrics for ton bags and damaged areas before and after improvement. The experimental results demonstrate a leap in detection performance for both ton bags and damaged regions. The overall ton bag detection mAP@0.5 increased from 72.3% in the baseline model to 86.7%, representing an absolute improvement of 14.4%, indicating a significant enhancement in the model's localization and classification capabilities for the ton bag itself. For damaged area detection, the mAP@0.5 rose from 63.2% to 84.3%, a substantial increase of 21.1%, which fully validates the reinforcing effect of the high-resolution P2 detection layer and Varifocal Loss on small target features.

The results confirm that through the three-dimensional collaborative optimization of "low-light enhancement, multi-scale detection, and deformation modeling," the model achieves a technological breakthrough in addressing the core challenges of industrial ton bag inspection. This optimization approach can be extended to similar industrial defect detection tasks.

5 Discussion

5.1 Analysis of Method Effectiveness

The multi-dimensional collaborative optimization framework proposed in this paper demonstrates significant advantages in the industrial ton bag detection task. The CLAHE module effectively addresses feature extraction challenges in low-light environments through adaptive enhancement of local contrast. The coordinated design of the high-resolution P2 detection layer and Varifocal Loss specifically improves the detection performance for small targets. The integration of deformable convolution with the SE attention mechanism enhances the model's adaptability to complex deformations. These three optimized aspects mutually reinforce each other, collectively building a robust detection system suitable for complex industrial scenarios.

5.2 Analysis of Limitations and Generalizability

Although this study uses ton bag detection as its application context, the proposed methodological framework demonstrates broad applicability. The enhancement of low-light conditions, detection of small targets, and adaptation to deformations represent common needs in the field of industrial vision. The proposed method can be transferred to similar scenarios, such as fabric defect detection, electronic component inspection, and agricultural product sorting. However, this study still has certain limitations: firstly, the validation is primarily based on a self-built specialized dataset; secondly, while the introduction of deformable convolution enhances adaptability to deformations, it also increases model complexity and training difficulty. Future research will explore more lightweight deformation modeling methods and validate the universality of the proposed approach across more industrial inspection scenarios.

6 Conclusion

This study addresses three major challenges in industrial visual inspection—low-light conditions, small targets, and geometric deformations—by proposing a multi-dimensional collaborative optimization framework based on YOLOv5. Through the synergistic design of CLAHE enhancement in the LAB color space, a high-resolution P2 detection layer combined with Varifocal Loss, and the integration of deformable convolution with the SE attention mechanism, the model's detection performance in complex environments has been effectively improved. Experiments on an industrial ton bag dataset demonstrate that the proposed method achieves detection accuracies of 78.4%, 85.2%, and 74.3% in low-light, small-target, and deformation scenarios, respectively, with overall performance surpassing state-of-the-art detectors such as YOLOv7 and YOLOv8. This research not only provides an efficient solution for the specific problem of ton bag detection but also offers a universal technical pathway for addressing similar challenges in industrial vision. Future work will focus on further lightweight design of the method and its transfer application in more industrial scenarios.

References

[1] Yu Kai, Hong Tao, Li Xun. Object Detection of Building Guardrails Based on Improved YOLOv5s[J]. Modern Electronics Technique, 2024, 47(14): 135-141. DOI: 10.16652/j.issn.1004-373x.2024.14.021.

[2] Zhu Enwen, Liang Zhao, Xiao Jinwen, et al. Research on Small Target Detection Based on MDS-YOLO Model[J]. Journal of Hunan University (Natural Sciences), 2024, 51(12): 78-86. DOI: 10.16339/j.cnki.hdxbzkb.2024285.

[3] Qiu X ,Zhang H ,Yuan C , et al.Advancing Ton-Bag Detection in Seaport Logistics with an Enhanced YOLOv8 Algorithm[J].Journal of Marine Science and Engineering,2024,12(11):1916-1916.

[4] Ju Moran, Luo Haibo, Wang Zhongbo, et al. Improved YOLO V3 Algorithm and Its Application in Small Target Detection[J]. Acta Optica Sinica, 2019, 39(07): 253-260.

[5] Deng Zhonggang, Dai Gang, Wu Xiangning, et al. Image Recognition Model for Minor Irregular Damage on Metal Surface Based on Attention Mechanism and Deformable Convolution[J]. Computer Engineering & Science, 2023, 45(01): 127-135.

[6] Zheng Mingming, Liu Shengquan, Ma Qian. Defect Detection Method Based on Deformable Convolution Fusing Dual Attention Mechanism[J]. Journal of Northeast Normal University (Natural Science Edition), 2023, 55(02): 52-61. DOI: 10.16163/j.cnki.dslkxb202201180002.

[7] Hu Dandan, Zhang Zhongting, Niu Guochen. Lane Line Detection Integrating CBAM Attention Mechanism and Deformable Convolution[J]. Journal of Beijing University of Aeronautics and Astronautics, 2024, 50(07): 2150-2160. DOI: 10.13700/j.bh.1001-5965.2022.0601.

[8] Xue Renzheng, Wu Qianlong, Ye Baofeng. YOLOv5 Pedestrian Detection Algorithm Based on Deformable Convolution[J]. Journal of Qiqihar University (Natural Science Edition), 2025, 41(02): 14-21+27. DOI: 10.20171/j.cnki.23-1419/n.2025.02.010.

[9] Geng Chuang, Song Pinde, Cao Lijia. Research Progress of YOLO Algorithm in Object Detection[J]. Journal of Ordnance Equipment Engineering, 2022, 43(09): 162-173.

[10] Ma Minhui, Wang Hongru, Wang Jia. Underwater Image Enhancement Algorithm Based on Improved MSRCR-CLAHE Fusion[J]. Infrared Technology, 2023, 45(01): 23-32. DOI: CNKI:SUN:HWJS.0.2023-01-003.

[11] Yang Yuting, Li Lingling, Liu Xu, et al. Image Recognition Based on Multi-Scale and Multi-Direction Transformer[J]. Chinese Journal of Computers, 2025, 48(02): 249-265.

[12] Lei Tao, Zhai Yujie, Xu Yetong, et al. Remote Sensing Image Change Detection Based on Edge Guidance and Dynamic Deformable Transformer[J]. Acta Electronica Sinica, 2024, 52(01): 107-117. DOI: CNKI:SUN:DZXU.0.2024-01-008.

[13] Yu Linhao, Zhong Yuan, Song Shuran, et al. Citrus defect detection in natural environments based on improved YOLOv7 algorithm [J/OL]. Journal of South China Agricultural University, 1-14 [2025-11-04]. https://link.cnki.net/urlid/44.1110.S.20251017.1339.002.

[14] Liu Jiashuai, Zhao Qiang, Yang Zhi, et al. Infrared image hotspot defect detection method for photovoltaic modules based on improved YOLOv8 [J/OL]. Laser & Optoelectronics Progress, 1-24 [2025-11-04]. https://link.cnki.net/urlid/31.1690.tn.20251027.2320.124.

Biography:



Mingliang Yang

Education: He earned his B.Eng. (2001), M.Eng. (2006), and Ph.D. (2011) in Mechanical Design and Theory, all from Taiyuan University of Science and Technology.

Research Interests: His research focuses on modern design methodologies for hoisting and conveying machinery structural systems, CAD/CAE simulation of complex machinery, and key technologies for intelligent logistics equipment.

Academic Appointments: He serves as Secretary of the Crane Technology Committee within the Logistics Engineering Division of the Chinese Mechanical Engineering Society.



Yuyu Miao(Corresponding Author**)**

A Master's candidate with research focuses on hoisting and transport machinery, and visual recognition.Email:miaoyuyu1998@163.com



Xijun Xu is the founder and chairman of Taiyuan Freeland Logistics Equipment & Technology Co., Ltd. He graduated from Taiyuan Heavy Machinery Institute (now Taiyuan University of Science and Technology) in 2001 with a major in Industrial Automation. Currently pursuing doctoral studies, he holds the professional title of Senior Engineer and serves as a Guest Professor and Master's Supervisor at Taiyuan University of Science and Technology.



Heng Yang

He received his Ph.D. in Mechanical Design and Theory from Taiyuan University of Science and Technology in 2015. Since July 2008, he has been serving as a faculty member in the Hoisting and Conveying Machinery Teaching & Research Division at the University's School of Mechanical Engineering, where he currently holds the position of Associate Professor.

Research Areas

His research focuses on key technologies for crane systems, performance and energy-saving methods for industrial vehicles, and reliability assessment of major equipment systems.



Qing Dong

Education: From 2007 to 2017, she completed her entire academic training at Taiyuan University of Science and Technology, earning her B.Eng. and M.Eng. in Mechanical Design and Theory, followed by a Ph.D. in Mechanical Engineering.

Academic Appointments: Since joining the university faculty in 2017, she has held positions as Lecturer and Associate Professor, and was promoted to Full Professor in 2025. During this period, she conducted postdoctoral research at Weihua Group (2021-2023).

Research Focus: Her research specializes in intrinsic safety and green assessment of hoisting machinery, digital design methodologies, and development of digital-intelligent integrated systems.



Keyuan Zhao

He received his Ph.D. in Mechatronic Engineering from Tongji University in 2022, where he studied from September 2018 to July 2022. He currently serves as a Lecturer and Master's Supervisor at the School of Mechanical Engineering, Taiyuan University of Science and Technology. His research focuses on operational environment perception and intelligent control of logistics equipment, as well as reliability assessment of cranes.

Declarations

(1)Funding Program: Shanxi Province Key Research and Development Program Project (202402150101006)

(2)Author Resume：Mingliang Yang(1976-), male, PhD, professor, research direction is intelligent logistics system key technology research. E-mail: yangmingliang1997@163.com.

(3)Corresponding author: Miao Yuyu (1998-), male, M.S. candidate, specializing in visual recognition.E-mail: miaoyuyu1998@163.com.

(4)Availability of data and materials

The self-constructed industrial ton-bag dataset supporting the research findings of this study on YOLOv5-based industrial ton-bag damage detection (including ton-bag images and damage annotation information under low-light, deformed, and normal scenarios) is not publicly available at this time, as it involves sensitive information of industrial production scenarios and data privacy related to corporate collaborations. For relevant researchers who need to access the dataset to verify the research findings or conduct academic collaborations, a reasonable request may be submitted to the corresponding author (Yuyu Miao, Email: miaoyuyu1998@163.com). Upon review and approval, limited-range authorization for dataset use will be provided.

(5) Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

(6) Funding:

"applicable"

(7) Authors' contributions

Mingliang Yang conceived the core multi-dimensionally co-optimized methodology based on YOLOv5s-6.0, which included the LAB color space CLAHE module for low-light feature enhancement and the overall technical framework for multi-scale detection and deformation modeling, providing the foundational algorithmic design for the ton bag damage detection model.

Yuyu Miao (Corresponding Author) was responsible for the core experimental execution of the industrial ton bag damage detection, including image acquisition under low-light and deformation scenarios, model training, and performance validation. By optimizing experimental parameters, he ensured the accurate realization of dark-light feature extraction, small damage identification, and deformation adaptation on the self-constructed dataset. He also led the overall structural design, core content writing, and revision of the manuscript, serving as the primary contributor to its preparation.

Xijun Xu and Heng Yang optimized the multi-scale feature fusion and adaptive loss system based on YOLOv5s. Their specific contributions included the integration and debugging of the high-resolution P2 detection layer, parameter optimization of the Varifocal Loss function, and the coordinated adaptation of deformable convolutions with the SE attention mechanism, thereby improving the model's recall rate for small damage targets and its detection accuracy for deformed samples.

Qing Dong and Keyuan Zhao designed and optimized the comprehensive experimental protocol for industrial ton bag damage detection. This involved planning the self-constructed ton bag dataset, including setting the ratios for low-light, deformed, and normal scenario samples and establishing damage annotation standards. They also designed the comparative and ablation studies and performed the collation and preliminary analysis of the experimental data.The authors would like to thank the constructive comments from the reviewers and editors that helped improve this manuscript.

(8)Acknowledgements

The authors would like to thank the constructive comments from the reviewers and editors that helped improve this manuscript.

(9)Ethics Approval

This study did not involve human participants or animals. Therefore, ethical approval was not required.

Disclosures

The authors declare there are no financial interests, commercial affiliations, or other potential conflicts of interest that have influenced the objectivity of this research or the writing of this paper.

Code and Data Availability

The datasets and models generated during this study are not publicly available due to commercial confidentiality. However, they can be made available from the corresponding author upon reasonable request, contingent upon obtaining necessary permissions and signing a confidentiality agreement.