Lightweight Detection Model for Animal Wet-Blue Hide Surface Defects Based on Yolov5s

by
Qixin Han,[1] Yushan Wan,[1] Luwen Cao,[1] Rong Luo,[2*] Yafei Sun[2] and Weikuan Jia[1*]
[1]*School of Information Science and Engineering, Shandong Normal University*, Jinan 250358, China
[2]*State Key Laboratory of Biobased Materials and Green Papermaking, Qilu University of Technology (Shandong Academy of Science),* Jinan 25035, China

Abstract

In the process of animal leather processing, the surface damage of wet-blue hides restricts the quality of leather products. To ensure the efficiency and quality of animal leather processing, a lightweight model for detecting surface defects on wet-blue hides based on optimized YOLOv5s is proposed. The new model adopts the lightweight EfficientNetV2 network to extract surface defect features and incorporates a spatial pyramid pooling–fast (SPPF) structure at the end of the network to obtain features at different scales. Efficient multi-scale attention (EMA) was embedded in the bottom-up structure of the Neck section to achieve comprehensive feature extraction and retention, ensuring that spatial semantic features are adequately distributed in each feature. A dataset of wet-blue hide defects was constructed and used to verify the performance of the new model. the experimental results show that, the new model is superior to the commonly used classical detection models. The precision rates for detecting three types of leather surface defects, namely imprint, puncture, and breakage, are 86.5%, 95.3%, and 87.9%, respectively. These results can provide technical support for research of surface damage detection in other leather processing applications.

1 Introduction

Leather and its products are among the most traded products in the world, with an annual international trade volume of more than $80 billion.[1] These products encompass a variety of items such as clothing, footwear, and decorative accessories.[2] People's demand for leather is not only practical but also related to the pursuit of aesthetics, thus posing a more stringent challenge to the production of leather products.[3-5] However, surface defects not only cause damage to the appearance of leather but also may have an impact on quality management.[6,7] Therefore, research on the detection of leather surface defects during animal leather processing can help to improve the safety, efficiency and quality of leather production. In the process of animal leather manufacturing, the grading of leather quality currently relies predominantly on manual inspection of the surface defects of wet-blue hides.[8-10] However, this approach faces challenges such as time-consuming processing and significant errors. There is an urgent need for efficient and precise research on the detection of surface defects in wet-blue hides.

Traditional machine learning based methods play an important role in the detection of leather surface defects. Georgieva et al.[11] used rotational invariance and scale invariance of grayscale histograms of leather images for feature analysis processing. On the other hand, Kwak et al.[12] used texture features to identify defects on leather, surfaces and were able to accurately detect a variety of defects, such as pinholes, scratches, and wrinkles, while missing few small defects. Jawahar et al.[13] used an innovative multilevel holding algorithm to segment the leather surface defect region to objectively quantify the leather surface defects and demonstrated the potential of classifying leather defects. Gan et al.[14] proposed an automatic leather defect localization and detection system using AlexNet as a feature descriptor and support vector machine (SVM) as a classifier to determine the presence or absence of defects on leather patches. Jawahar et al.[15] proposed an SVM classifier image processing technique based on wavelet features that can automatically identify leather blemish defects. Although the above methods have achieved a series of results in the process of defect detection, some of the defect features are difficult to extract due to the influence of noisy environmental factors in the processing workshop, which further restricts the detection accuracy of surface defects.

Deep learning methods, which can learn features directly from raw data, have been widely used in the fields of surface defect detection,[16,17] equipment fault diagnosis[18] and medical imaging[19] and have achieved remarkable results; inspired by these methods, deep learning theory has been introduced into the detection of surface defects in animal leather. For example, Liong et al.[20] developed an automatic defect detection technique based on AlexNet and U-Net to classify leather images into three categories (normal, black lines, and wrinkles) to achieve accurate pixel-level defect location. Liong et al.[21] reduced the number of defect boundary points while maintaining the shape of the defects for tick bite damage, and ultimately used a robotic arm to automatically outline the boundaries of the defects on the leather surface. Luo et al.[22] proposed a robust breakage detection network (RBD-Net) model for leather breakage detection, which can effectively address various interfering factors and ensure reliable detection of leather in

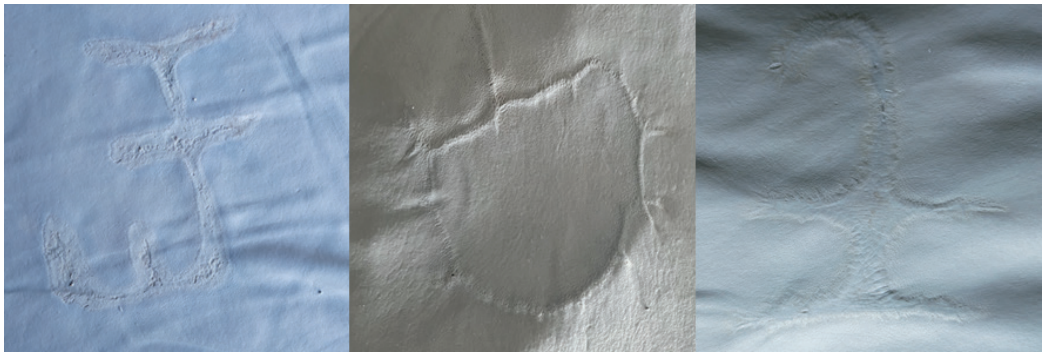
*Corresponding authors email: R Luo, lrcity@qlu.edu.cn ; WK Jia, jwk_1982@163.com
Manuscript received January 30, 2024, February 26, 2024.

practical applications. Iqbal et al.[23] developed an automated system for detecting defects in leather images based on visual surface analysis and introduced a multi-layer residual convolutional neural network (MLR-Net), which effectively and accurately identifies and segments defects in leather images. In comparison to traditional machine learning algorithms, the aforementioned defect detection algorithm exhibits significantly improved accuracy and robustness.

Current algorithms are mostly designed to operate under ideal conditions. Faced with noisy processing environments, factors such as lighting, defect types, and shapes can significantly impact the efficient and accurate detection of surface defects on wet-blue hides. In the large-scale production process of animal leather, achieving real-time and efficient detection is crucial. Employing lightweight detection models can meet this requirement while providing effective solutions and reducing hardware costs. Furthermore, with the proliferation of industrial smart devices such as robots and intelligent cameras, deploying lightweight detection models on these embedded devices is more feasible, thereby enhancing system flexibility and scalability. The mentioned lightweight model is characterized by reduced parameter count and computational complexity, while also demonstrating faster detection and inference speeds. Therefore, this study proposed a lightweight detection model based on optimized YOLOv5s. The primary contributions of this study include the following:

(1) The introduction of the lightweight EfficientNetV2 network is employed to extract surface defect features from wet-blue hides, thereby reducing the model's parameter count. This approach aims to address the challenges associated with inadequate feature extraction and low training efficiency.

(2) Introducing the spatial pyramid pooling–fast (SPPF) structure at the end of the backbone network facilitates the fusion of features across multiple scales. Additionally, embedding efficient multi-Scale attention (EMA) in the bottom-up structure of the Neck section ensures the effective distribution of spatial semantic features, achieving cross-space multi-scale aggregation.



a) Brand

b) Broken hole

c) Broken surface

Figure 1. Three Different Types of Defects

(3) Considering issues such as overlapping areas and geometric features of detection boxes, the complete IoU (CIoU) loss is employed as the regression loss function, to enhance the convergence speed and performance of the network.

The rest of the paper is organized as follows. Section 2 introduces the dataset. Section 3 describes the proposed method. Section 4 conducts experiments and analyzes the results. Finally, Section 5 summarizes the work.

2 The wet-blue hide defect dataset

2.1 Image acquisition

The leather images selected for this study were obtained from preliminarily processed cowhides, resulting in wet-blue hides. The details of the acquisition are outlined below:

Defect types: Images were collected for three distinct surface defect types on wet-blue hides: punctures, breaks, and imprints. Figure 1 provides examples of wet-blue hide images depicting these three types of defects.

Collection sites: The data collection was conducted at Shandong Dexin Leather Co., Ltd. and Zibo Dahuan Jiubao'en Leather Group Co., Ltd., both of which are located in Zibo city, Shandong Province, China.

Collection environment: The processing environment of wet-blue hides is inherently complex. To simulate the actual factory conditions as closely as possible, a diverse set of leather images was collected. These images encompassed various time periods, different shooting angles and distances, diverse lighting conditions, and different types of interference and damage. Samples closer to the camera were treated as large targets, while those farther away served as small targets, thereby enriching the dataset comprehensively.

Equipment for data acquisition: The wet-blue hide images were captured using both smartphones and a Sony Alpha 7 II camera, resulting in a total of 2162 images. To enhance the network's detection quality for low-resolution images under real-time detection requirements, the images were compressed and uniformly scaled to 800 pixels × 600 pixels.

2.2 Data augmentation

Training a detection model often relies on a sufficient amount of sample data. Therefore, to enhance the model's generalizability and stability, and to further improve its detection accuracy while avoiding overfitting due to insufficient data collection, this study employs Mosaic data augmentation[24] to process the input images. In each iteration cycle (Epoch), this method randomly replaces images. The replacement involves random scaling, random cropping, and random arrangement to generate new images. During the training process, images with different combinations are reconfigured in the subsequent iteration cycle, thereby enhancing the network's generalization capability. Figure 2 shows the annotated images after surgery with Mosaic data augmentation.

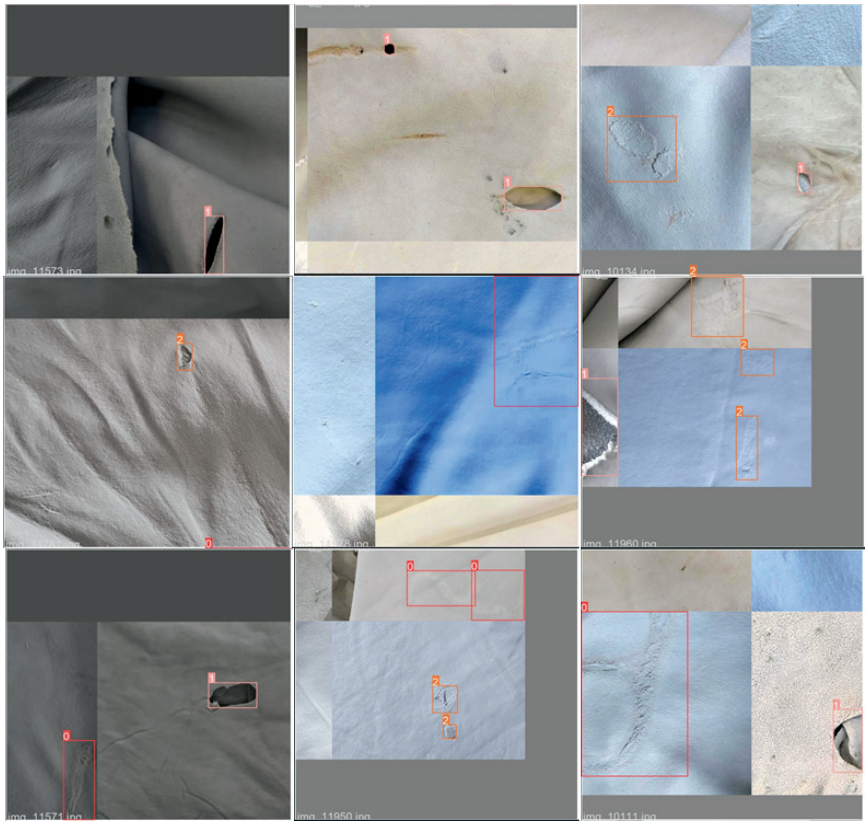


Figure 2. Mosaic Data Augmentation

Table I

Statistical Summary of Defect Quantities by Different Sizes.

	small target	medium target	large target	total
Training Set	260	236	1776	2272
Validation Set	150	129	750	1029
Total	410 (12.42%)	365 (11.06%)	2526 (76.52%)	3301

2.3 Dataset creation

First, the leather defect dataset images were effectively annotated using LabelMe software, with labels categorized as Brand, Broken hole, and Broken surface. The annotated points denote the damaged regions on the wet-blue hides, while the remaining areas serve as the background. All the annotation information is stored in JSON files corresponding to the original images, and these JSON files are converted into a leather defect dataset in Microsoft COCO format.[25] Following the annotation of target areas, defects are further categorized into small, medium, and large sizes based on the following criteria:

$$\text{small target: area} \le 32^2.$$
$$\text{medium target: } 32^2 < \text{area} \le 96^2.$$
$$\text{large target: area} > 96^2$$

Finally, based on the generated dataset, the data were partitioned into training and validation sets at a 7:3 ratio. The annotation information for each dataset was consolidated, forming the ultimate wet-blue hide surface defect dataset. The quantity of defects of different sizes in the dataset is presented in Table I.

3 Lightweight detection model for surface defects on wet-blue hides

Object detection networks are generally complex and involve large model sizes. To improve the efficiency and quality of animal leather processing, this study proposes a lightweight detection model tailored for surface defects on wet-blue hides based on the YOLOv5s model. This model not only allows for rapid and efficient identification of surface defects but also facilitates deployment on embedded devices. As illustrated in Figure 3, the overall network is primarily divided into four parts: Input, Backbone, Neck, and Output. The Backbone stage utilizes the EfficientNetV2[26] +SPPF structure to reduce model complexity and expedite training. The Neck stage adopts the feature pyramid network (FPN)+bottom-up[27] structure and embeds the EMA attention module[28] into the bottom-up structure, enabling better retention of spatial semantic features and improved fusion of multiscale features. The output layer of the network employs the CIoU loss[29] as the loss function.

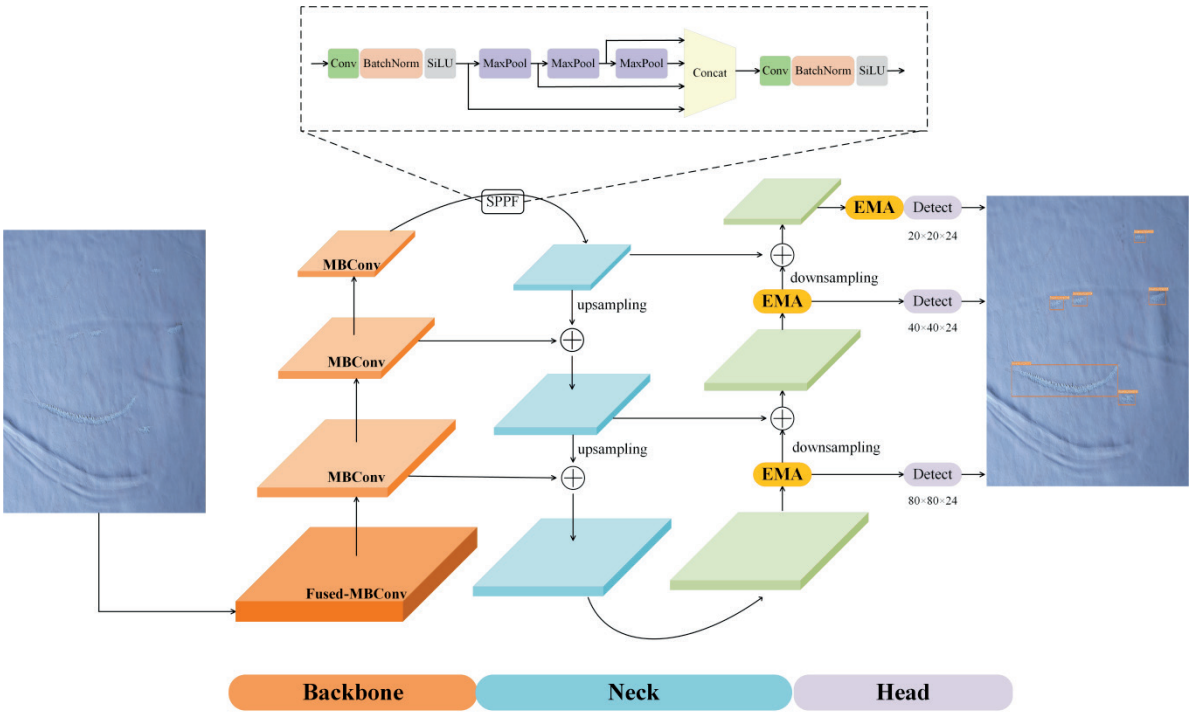


Figure 3. Lightweight Detection Model for Surface Defects on Wet-Blue Hides

3.1 Defect feature extraction architecture

In the intricate production processes of animal leather processing, ensuring that the detection equipment can accurately discern surface defects on leather is of paramount importance. To achieve this goal, embedding a fast and straightforward network model into the detection equipment becomes imperative. This allows for efficient detection of surface defects on animal leather in complex environments while ensuring precise classification and processing of the animal leather. In this study, the backbone network of the new model utilizes the compact yet powerful EfficientNetV2 network and the SPPF structure.

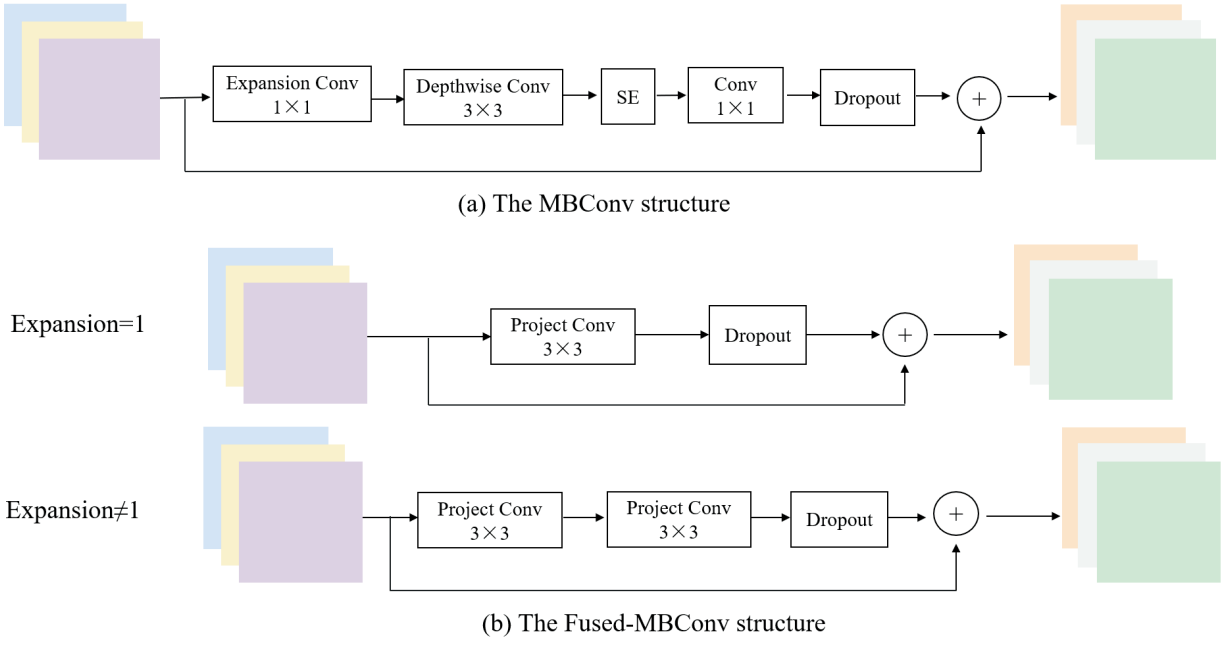
The EfficientNetV2 model[30] is characterized by its compact size, fast training speed, and high parameter efficiency, and is achieved through the utilization of multiple convolutions with a stride of 2 to reduce the data size. The network architecture of EfficientNetV2 is presented in Table II. This architecture employs fewer parameters to attain higher accuracy and efficiency, thereby striking a balance between model precision and operational speed.

The EfficientNetV2 network is primarily composed of shallow Fused-MBConv and deep MBConv layers, as illustrated in Figure 4.

Table II

The architecture of the EfficientNetV2 network

Stage	Operation	Stride	Channels	Layers
0	Conv3×3	2	24	1
1	Fused-MBConv1 k3×3	1	24	2
2	Fused-MBConv4 k3×3	2	48	4
3	Fused-MBConv4 k3×3	2	64	4
4	MBConv4, SE0.25 k3×3	2	128	6
5	MBConv6 ,SE0.25 k3×3	1	160	9
6	MBConv6,SE0.25 k3×3	2	256	15
7	Conv1×1 & Pooling & FC	-	1280	1



(a) The MBConv structure

(b) The Fused-MBConv structure

Figure 4. The Fused-MBConv and MBConv structures

In the MBConv module, depth wise separable convolution is employed, where each convolution kernel is responsible for one input channel, meaning that each channel is convolved by only one kernel. The number of channels obtained on the feature map is exactly consistent with the number of input channels. This convolution comprises depth wise convolution and pointwise convolution. The Fused-MBConv module replaces Depth wise Conv3×3 and Expansion Conv1×1 in MBConv with the standard Conv3×3. Through the combination of training-aware Neural Architecture Search (NAS) and scaling, both modules work collaboratively to optimize the model's training speed and parameter efficiency.

While the rational combination of the aforementioned two modules significantly reduces the number of parameters and computations, there is a trade-off in terms of feature extraction efficacy. This is because the modified feature extraction network, as part of the streamlined model, loses a portion of the feature information. To mitigate the risk of excessive information loss, an SPPF structure is added after the EfficientNetV2 network architecture. This structure utilizes pooling kernels of different sizes to extract features at various scales, capture objects at different scales, and better understand the contextual information in the input image. This addition aims to alleviate the problem of information loss.

3.2 EMA attention module

To better integrate the defect features extracted by the backbone network, EMA attention is embedded in the bottom-up structure of the Neck section. This approach facilitates the spatial aggregation of multiscale features, ensuring an improved distribution of spatial semantic features within each feature. The structure of EMA attention comprises three main components: Feature Grouping, Parallel Subnetworks, and Cross-spatial learning, as illustrated in Figure 5.

First, the EMA divides any given feature map into G sub-features along the channel dimension to learn different semantics. Second, within the parallel subnets, the EMA employs three parallel routes to extract attention weight descriptors for grouped feature maps. Two parallel routes employ two one-dimensional global average pooling operations encoding channels along two spatial directions, and the third route merely stacks a single 3×3 kernel to capture multiscale feature representations. Finally, in cross-space learning, two tensors are introduced to encode the global spatial information via two-dimensional global average pooling for the outputs of the two branches in the parallel subnet. The corresponding outputs are subsequently transformed into the respective dimensional shapes, denoted as $R_1^{1\times C//G} \times R_3^{C//G\times H*W}$ and $R_3^{1\times C//G} \times R_1^{C//G\times H*W}$, before being subjected to the joint activation mechanism of channel features. The employed two-dimensional global average pooling is expressed as follows:

$$Z_c = \frac{1}{H\times W}\sum_j^H \sum_i^W x_c(i,j) \tag{1}$$

which is designed to encode global information and model long-range dependencies, where, x_c represents the input feature at the *c-th* channel, C represents the number of input channels, and H and W represent the spatial dimensions of the input feature.

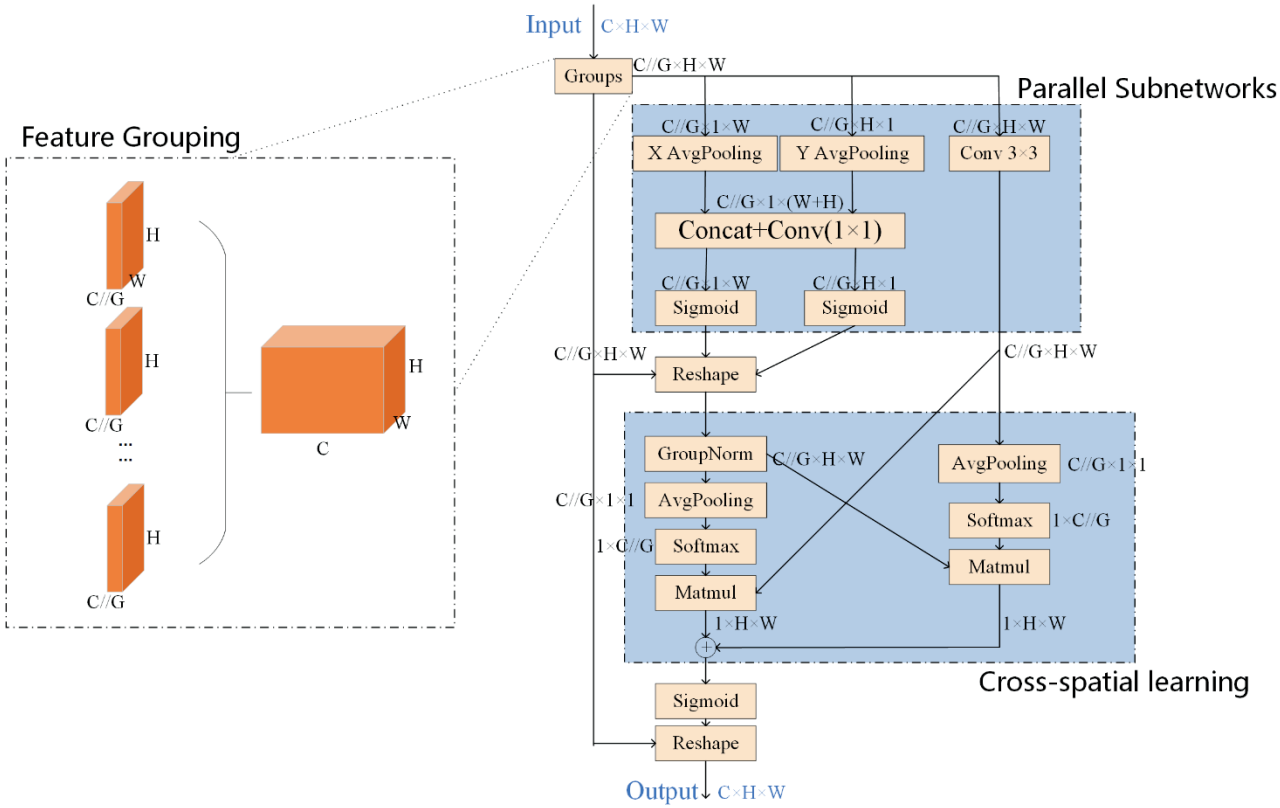


Figure 5. Diagram of the Efficient Multi-Scale Attention Structure

Introducing EMA attention allows focused retention of information in each channel while reducing computational overhead. It redefines a portion of channels on a batch level and divides the channel dimension into multiple sub-features, ensuring that spatial semantic features are well-distributed within each feature. This approach addresses issues such as unclear edges in defect images, where regions formed by the leather edge and the input image boundary are prone to being mistakenly identified as punctures. Consequently, this approach enhances the ability of the model to detect surface defects on leather surfaces.

3.3 The CIoU loss function

The choice of a loss function plays a crucial role in model training and is vital for optimizing the iterative process and achieving optimal training results through gradient backpropagation. This study enhances the loss function of the YOLOv5s model to improve its ability to detect surface defects on leather surfaces. The loss function for YOLOv5s is presented in Formula (2), which consists of three parts, with λ_1, λ_2, λ_3, as the balance coefficient. The first part, loss_cls, employs the binary cross entropy (BCE) loss to calculate the classification loss, focusing solely on positive samples. In the second part, loss_loc, the generalized IoU (GIoU) loss is utilized to compute the regression loss for only positive samples. Finally, loss_conf employs the BCE loss to calculate the target confidence loss for all samples.

$$\text{Loss} = \lambda_1 L_{\text{cls}} + \lambda_2 L_{\text{loc}} + \lambda_3 L_{\text{conf}} \tag{2}$$

However, when two prediction boxes overlap completely, the GIoU metric fails to accurately reflect the distance and positional relationship between the prediction box and the ground truth box. When the GIoU values are equal, the overlap effect of the two boxes may differ, making it challenging to quickly and accurately determine the optimization direction for the localization box. The CIoU loss function addresses this limitation by incorporating the Euclidean distance between the prediction box and the ground truth box, as well as the loss of the detection box scale, into the penalty term. This approach enhances the aspect ratio of the target box, progressively bringing the prediction box closer to the ground truth box during continuous training, thereby improving the convergence speed and regression accuracy of the network. Consequently, the model proposed in this study replaces the regression loss function with the CIoU loss, and the penalty term of the CIoU loss is shown in Formula (3).

$$\mathfrak{R}_{CIoU} = \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \tag{3}$$

$$\alpha = \frac{v}{(1\text{-}IoU)+v} \tag{4}$$

$$v = \frac{4}{\pi^2}\left(arctan\ \frac{w^{gt}}{h^{gt}} - arctan\ \frac{w}{h}\right)^2 \tag{5}$$

Where α is a positive balancing parameter, v measures the consistency of aspect ratios, b and b^{gt} represent the center coordinates of the predicted and ground truth boxes, respectively; $\rho(b, b^{gt})$ is the Euclidean distance between the center coordinates of the predicted and ground truth boxes; c is the Euclidean distance between the two diagonal vertices of the minimum bounding rectangle for the predicted and ground truth boxes; w, h, w^{gt} and h^{gt} are the widths and heights of the predicted and ground truth boxes, respectively; and intersection over union (IoU) is the intersection over union between the predicted and ground truth boxes. Thus, the CIoU loss is defined as follows:

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b,b^{gt})}{c^2} + \alpha v \tag{6}$$

The introduction of the CIoU loss expedites model convergence, addressing the drawback where gradients were ineffective at backpropagation when the predicted box overlaps with the ground truth box. This enhancement results in more stable regression for the target box.

4. Experiments

To better validate the effectiveness of the model in detecting surface defects on animal leather, this study conducted the following experiments, providing detailed descriptions of the experimental setup and performing comparative analyses of the results. First, a detailed description of the experimental platform was provided. Second, experimental details for both the training and testing phases were outlined. For the training process, the optimal model was selected and applied to the validation set, allowing for a comparison and evaluation of the experimental results. Finally, comparative experiments were conducted using existing classical object detection algorithms under the same experimental configuration to assess the performance of the proposed model in this study.

4.1 Experimental operating platform

All the experiments conducted in this study were completed on the same server system, which is primarily configured with the Ubuntu 16.04 LTS operating system. The processor utilized is an Intel® Xeon(R) Silver 4214R CPU @ 2.40 GHz × 45, complemented by a 10 GB NVIDIA GeForce RTX 3080 GPU and V11.4 CUDA environment. All the models are executed using the Python language and the PyTorch 1.7 deep learning library.

4.2 Details of the experimental implementation

To conduct more effective experiments, enhance the model's adaptability to the dataset, and improve the detection accuracy of surface defects on leather, the image size is standardized to (800, 600) before feeding it into the training network. During formal training, a small-batch method is employed for 300 epochs of iteration. After each epoch, the model's training status is validated using the validation set. This approach facilitates better adjustment of hyperparameters to obtain optimal values and enhances the overall efficiency of the model.

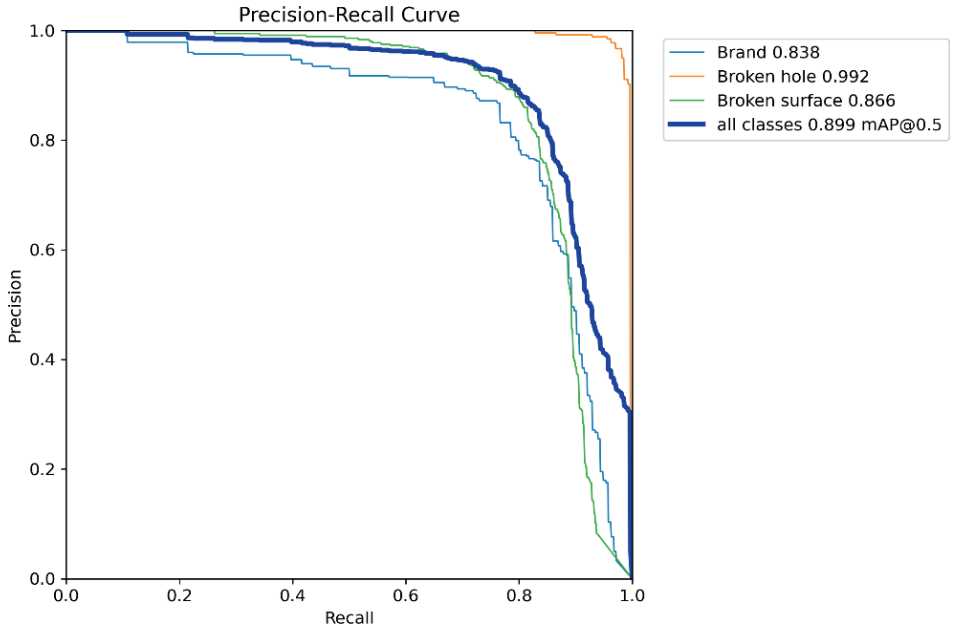


Figure 6. The Precision-Recall (PR) curve for model training

Weights are normalized using batch normalization at each refreshment, and model parameters are updated using stochastic gradient descent. During the training process, the learning rate, weight decay, and momentum are set to 0.01, 0.0005, and 0.900, respectively. The training progress of the model is illustrated in Figure 6.

4.3 Evaluation criteria

As the model needs to accurately predict the results of surface defect detection on leather, this experiment employs precision and recall as evaluation metrics for the model's effectiveness. The precision is calculated using Formula (7), and the recall is calculated using Formula (8).

$$precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

Where *TP, FP, FN* and *TN* represent the sample quantities of true positive, false positive, false negative, and true negative instances of defects, respectively. To comprehensively evaluate the model, further calculations are performed using Formula (9) to compute the Average Precision (AP) metric at specified *IoU* thresholds and *mAP*, as per Formula (10), to assess the mean average precision.

$$AP = 1/101 \sum_{r \in R} P(r) \tag{9}$$

$$mAP = 1/n \sum_{i=\mathrm{I}}^{n} AP_i \tag{10}$$

Where the letter *r* represents the recall, and *R* is the set of recall rates, which consists of 101 values: [0.0, 0.01, 0.02, ..., 0.98, 0.99, 1.0]. P(r) denotes the precision related to the recall rate. The letter *i* represents the *IoU* threshold, and *I* is the set of *IoU* thresholds, comprising 10 values: [0.5, 0.55, 0.6, 0.65, ..., 0.9, 0.95]. n is the number of categories, and in this study, n is set to 3, corresponding to three defect categories.

When the predicted category of a defect on the leather surface is correct and the *IoU* is greater than a certain threshold (the *IoU* threshold in the experiment is set to 0.5), the detection result is considered correct. The model's evaluation results for the detection of three different defect types in the leather dataset are presented in Table III, and the actual detection performance of the three defects is shown in Figure 7.

Table III

Results of Comparative Evaluations for Defect Detection in Three Leather Categories

Type	Precision/%	Recall/%	mAP@.5/%	mAP@.5:.95/%
Brand	86.5	76.6	83.8	52.5
Broken hole	95.3	98.1	99.2	80.9
Broken surface	87.9	79.9	86.6	55.8

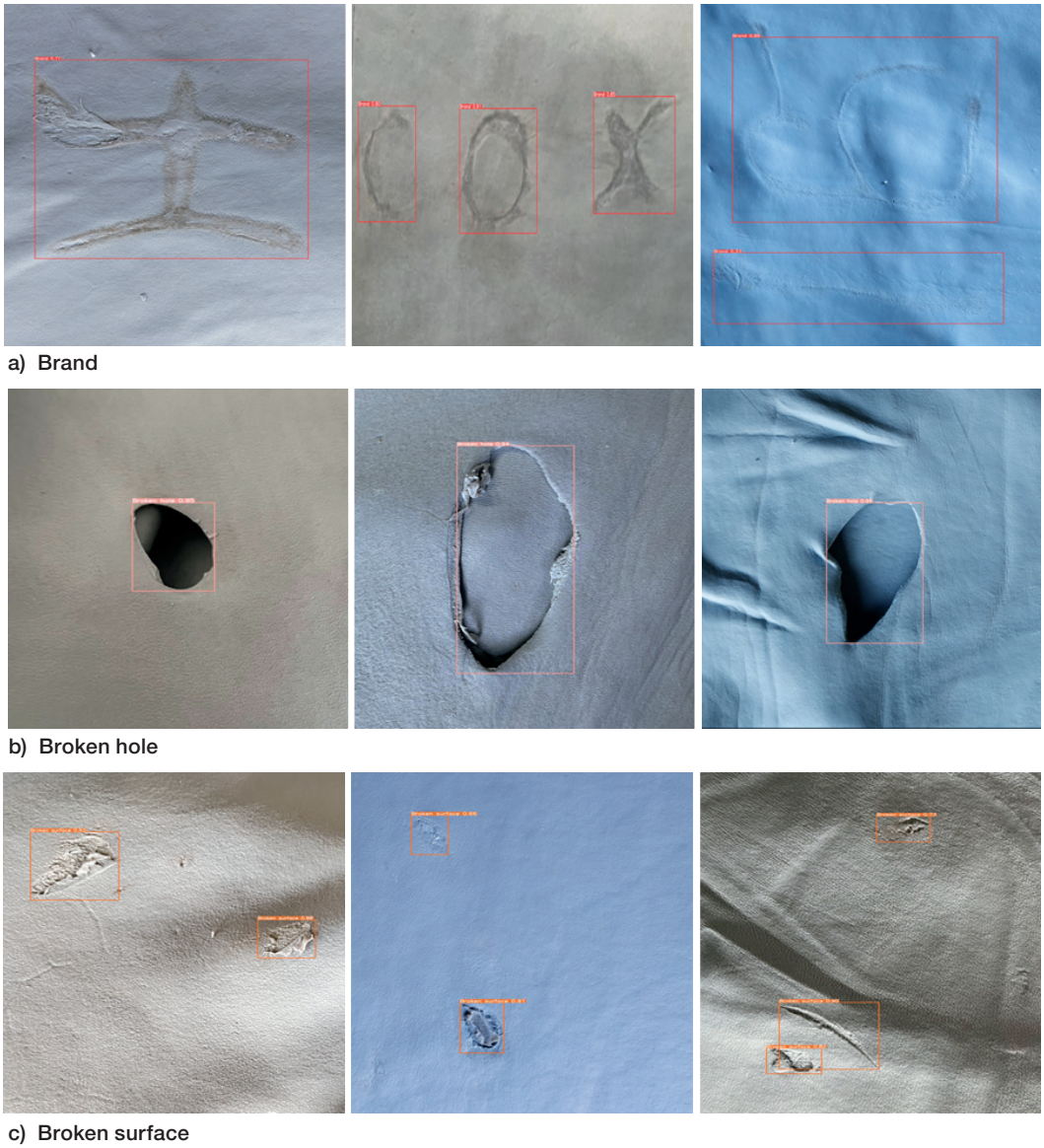


Figure 7. Results of Leather Surface Defect Detection

4.4 Ablation studies

In this section, the effectiveness of the model's backbone network and attention module was validated through ablation studies. To comprehend the contributions of the EfficientNetV2 network, EMA attention module, and CIoU loss function to detection efficiency and accuracy, the original backbone network is used as a baseline for comparison against the impacts of EfficientNetV2, EMA, and CIoU loss.

Table IV

The Impact of the EfficientNetV2 and EMA Modules on the Experimental Results

Baseline Model	Backbone Network	EMA Attention	Loss Function	precision/%	recall/%	mAP@.5/%	Params/M
YOLOv5s	CSPDarknet	×	GIoU loss	87.5	84.8	88.3	7.03
	EfficientNetV2	√	GIoU loss	88.7	84.8	89.2	5.60
	EfficientNetV2	×	CIoU loss	91.5	82.8	89.0	5.60
	CSPDarknet	√	CIoU loss	87.4	86.3	88.4	7.03
YOLO v5s	EfficientNetV2	√	CIoU loss	89.6	84.9	89.9	5.60

As shown in Table IV, when the backbone network is replaced with EfficientNetV2 and the CIoU loss is utilized as the loss function, the improvement in accuracy is most pronounced, with an increase of 4%. When using the original backbone network and incorporating EMA attention with CIoU loss as the loss function, the *recall* rate shows the most significant improvement, increasing by 1.5%. Furthermore, when all three factors work together, all three metrics improve, with the *mAP* experiencing the most notable increase, increasing by 1.6%. When EfficientNetV2 is employed as the backbone network for all the cases, the model's parameter count decreases by 1.43M. These results indicate that the improved YOLOv5s model not only effectively reduces the model's parameter count but also enhances the detection accuracy, validating the effectiveness of the various modules in the new model.

4.5 Comparison experiments

To validate the detection performance of the improved YOLOv5s model, comparisons were made with classical and state-of-the-art detection algorithms, including Faster R-CNN,[31] FSAF,[32] Mask R-CNN,[33] Grid RCNN,[34] Yolact,[35] YOLOv3,[36] and YOLOv5s. The experiments were conducted in the same environment using identical equipment and image preprocessing procedures. Following model detection, a uniform application of Non-Maximum Suppression (NMS) and the same *IoU* threshold was employed for filtering. The specific experimental results are presented in the following table. Notably, the value of the *mAP* is particularly emphasized because it accurately reflects the detection accuracy of leather defect types, facilitating subsequent determination of leather grades and corresponding processing. Thus, the *IoU* threshold of 0.5 and the *mAP* corresponding to small, medium, and large targets were selected as the comparative evaluation metrics.

According to the results presented in Table V, the improved model achieved a *mAP* of 89.9%. For large targets with an *IoU* threshold of 0.5, the AP values were 67.1% and 74.4%, generating an $AR_{maxDet=100}$ of 76.2%. According to the results presented in Table VI, the model's parameter count is 5.60M, surpassing the metrics of other comparative models. However, to ensure the precision of detection, the model's frames per second (FPS) and inference time do not reach the optimal level. Therefore, the improved model exhibits superior detection capabilities and lower computational complexity. Moreover, in comparison with the actual detection results of YOLOv5s shown in Figure 8, the improved YOLOv5s demonstrate finer handling of regions formed by the leather edge and image boundary, avoiding misidentifying edge regions as puncture damage.

Table V

The detection results of both classical and state-of-the-art detection models for leather defects.

Model	Backbone Network	mAP/%	AP_s/%	AP_m/%	AP_l/%	$AR_{maxDet=100}$/%
Faster RCNN	ResNet50	83.9	32.3	60.3	63.7	70.0
Mask RCNN	ResNet50	71.7	26.6	45.6	50.7	57.4
Grid RCNN	ResNet50	84.6	40.9	63.1	67.6	74.5
FSAF	ResNet50	81.5	52.7	55.1	57.1	67.7
Retinanet	ResNet50	82.7	54.2	60.3	65.1	71.4
Yolact	ResNet50	73.0	18.9	44.8	52.6	55.1
YOLO v3	Darknet-53	85.8	35.3	63.4	65.6	69.4
YOLO v5s	CSPDarknet	88.3	41.0	63.9	73.6	75.5
Ours	EfficientNetV2	89.9	38.9	67.1	74.4	76.2

Table VI

The parameter count, computational complexity, and detection performance outcomes of the detection model.

Model	Backbone Network	Params/M	FPS/s	Inference time/ms	Flops/GFLOPs
Faster RCNN	ResNet50	41.3	16.0	62.1	71.7
Mask RCNN	ResNet50	44.4	12.7	78.7	125.0
Grid RCNN	ResNet50	64.5	9.8	102.0	291.0
FSAF	ResNet50	36.4	12.9	77.5	60.7
Retinanet	ResNet50	38.0	14.1	70.9	61.5
Yolact	ResNet50	34.8	36.2	27.6	61.3
YOLO v3	Darknet-53	62.0	18.3	54.6	58.6
YOLO v5s	CSPDarknet	7.0	45.2	20.9	16.0
Ours	EfficientNetV2	5.6	25.9	37.5	6.9

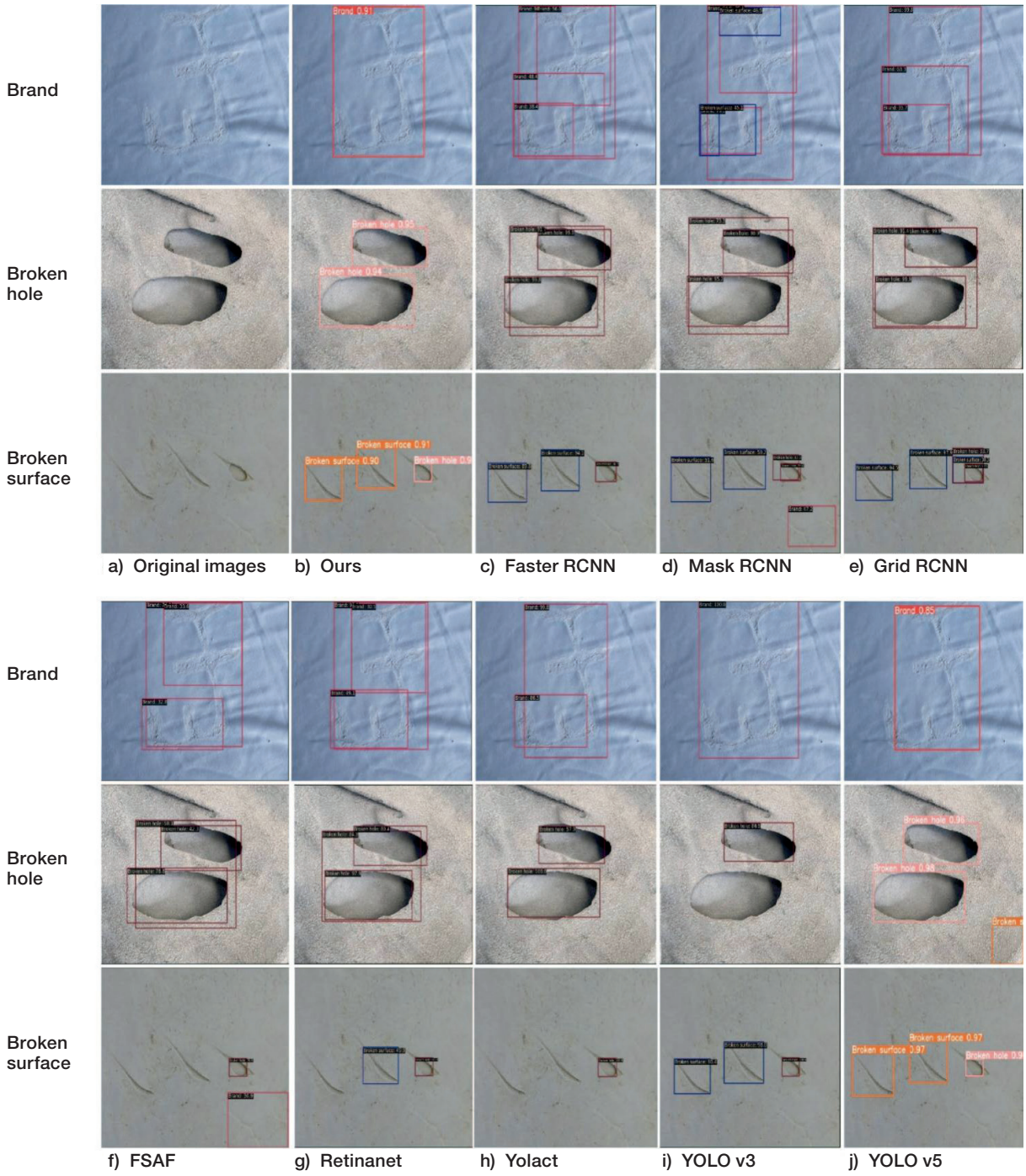


Figure 9. Comparison of Detection Images in the Wet-Blue Hide Dataset

The above analysis indicates that the improved detection model, which has the characteristics of a simple network structure, low computational complexity, and fast model speed, achieves high detection accuracy. This ensures a balance between accuracy and speed, making it suitable for application in animal leather production inspection. The improved model can be deployed in the leather surface defect detection process on production lines. Figure 9 shows a comparison of the images detected in the wet-blue hide dataset.

5. Conclusion

Addressing the issue of surface defect detection in animal leather, this research optimizes the YOLOv5s model tailored to different defect types, thereby achieving robust results in leather surface defect detection. In the proposed model, the EfficientNetV2 network and SPPF structure replace the original backbone network, effectively reducing the parameter count and computational complexity while maintaining high accuracy. Additionally, the EMA attention

module is embedded in the Neck section, preserving semantic features extracted from space more effectively and ensuring their distribution across each feature. Finally, in the Head section of the model, a more precise CIoU loss is employed as the regression loss function, which not only accelerates the model convergence but also ensures more stable regression.

This study focused on surface defects in wet-blue hides in animals and presents an improved lightweight detection model based on YOLOv5s. The model exhibits a 23% reduction in parameters, achieving the *mAP* of 89.9%. In comparison to other lightweight networks, the proposed algorithm not only reduces the number of parameters but also maintains high accuracy. This makes it more suitable for deployment on mobile devices with limited computational resources, facilitating efficient detection and identification of surface defects in animal leather.

By simulating scenarios in animal leather production, this model demonstrates the ability to overcome various interferences, enabling cost-effective detection and exhibiting strong stability. Given the model's efficiency in detecting surface defects on animal leather, its detection algorithm can be extended to other production inspection contexts. While the model achieves efficient detection results, the complexity of production work needs to be taken into consideration. Future research may focus on further enhancing the model's efficiency and practical operability.

Acknowledgments

This work is supported by National Nature Science Foundation of China (No.: 21978139); Natural Science Foundation of Shandong Province in China (ZR2019MB030); New Twentieth Items of Universities in Jinan (2021GXRC049).

References

1. Omoloso O, Mortimer K, Wise W R, et al. Sustainability research in the leather industry: A critical review of progress and opportunities for future research. *Journal of Cleaner Production*, 2021, 285: 125441.
2. Cadirci B H, Ozgunay H, Vural C, et al. A new defect on leather: Microbial bio-film. *JALCA* 2010, **105**(04): 129-134.
3. Aslam M, Naqvi S S, Khan T M, et al. Trainable guided attention based robust leather defect detection. *Engineering Applications of Artificial Intelligence*, 2023, **124**: 106438.
4. Jawahar M, Babu N K C, Vani K, et al. Vision based inspection system for leather surface defect detection using fast convergence particle swarm optimization ensemble classifier approach. *Multimedia Tools and Applications*, 2021, **80**: 4203-4235.
5. Mohammed K M C, Kumar S S, Prasad G. Optimized fuzzy c-means clustering methods for defect detection on leather surface. *Journal of Scientific and Industrial Research*, 2020, **79**(9): 833-836.
6. Liu Y, Yuan Y, Liu J. Deep learning model for imbalanced multi-label surface defect classification. *Measurement Science and Technology*, 2021, **33**(3): 035601.
7. Liu X, He W, Zhang Y, et al. Effect of dual-convolutional neural network model fusion for Aluminum profile surface defects classification and recognition. *Mathematical Biosciences and Engineering*, 2022, **19**(1): 997-1025.
8. Aslam M, Khan T M, Naqvi S S, et al. An ensemble of fine-tuned deep learning networks for wet-blue leather segmentation. *JALCA* 2022, **117**(4).
9. Aslam M, Khan T M, Naqvi S S, et al. Ensemble convolutional neural networks with knowledge transfer for leather defect classification in industrial settings. IEEE Access, 2020, 8: 198600-198614.
10. Aslam M, Khan T M, Naqvi S S, et al. Learning to recognize irregular features on leather surfaces. *JALCA* 2021, **116**(5).
11. Georgieva L, Krastev K, Angelov N. Identification of surface leather defects. *CompSysTech*, 2003, **3**, 303–307.
12. Kwak C, Ventura JA, Tofang-Sazi K. Automated defect inspection and classification of leather fabric. *Intelligent Data Analysis*, 2001, **5**(4):355–370.
13. Jawahar, M., & Vani, K. Machine vision inspection system for detection of leather surface defects. *JALCA* 2019, **114**(1), 10–19.
14. Gan Y S, Liong S T, Zheng D, et al. Detection and localization of defects on natural leather surfaces. *Journal of Ambient Intelligence and Humanized Computing*, 2021: 1-15.
15. Jawahar M, Babu N K C, Vani K. Leather texture classification using wavelet feature extraction technique. 2014 IEEE International Conference on Computational Intelligence and Computing Research. IEEE, 2014: 1-4.
16. Saberironaghi A, Ren J, El-Gindy M. Defect detection methods for industrial products using deep learning techniques: A review. *Algorithms*, 2023, **16**(2): 95.
17. Aslam M, Khan T M, Naqvi S S, et al. Putting current state of the art object detectors to the test: towards industry applicable leather surface defect detection. 2021 *Digital Image Computing: Techniques and Applications* (DICTA). IEEE, 2021: 01-08.
18. Kaya Y, Kuncan F, ERTUNÇ H M. A new automatic bearing fault size diagnosis using time-frequency images of CWT and deep transfer learning methods. *Turkish Journal of Electrical Engineering and Computer Sciences*, 2022, **30**(5): 1851-1867.
19. Kaya Y, Yiner Z, Kaya M, et al. A new approach to COVID-19 detection from X-ray images using angle transformation with GoogleNet and LSTM. *Measurement Science and Technology*, 2022, **33**(12): 124011.
20. Liong S T, Zheng D, Huang Y C, et al. Leather defect classification and segmentation using deep learning architecture. *International Journal of Computer Integrated Manufacturing*, 2020, **33**(10-11): 1105-1117.
21. Liong S T, Gan Y S, Huang Y C, Yuan C A, Chang H C. Automatic defect segmentation on leather with deep learning. arXiv preprint arXiv: 1903.12139, 2019.
22. Luo R, Chen R, Jia F, et al. RBD-Net: robust breakage detection algorithm for industrial leather. *Journal of Intelligent Manufacturing*, 2023, **34**(6): 2783-2796.
23. Iqbal S, Khan T M, Naqvi S S, et al. MLR-Net: A multi-layer residual convolutional neural network for leather defect segmentation.

Engineering Applications of Artificial Intelligence, 2023, 126: 107007.
24. Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934, 2020.
25. Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13. Springer International Publishing, 2014: 740-755.
26. Tan M, Le Q. Efficientnetv2: Smaller models and faster training. International conference on machine learning. *PMLR*, 2021: 10096-10106.
27. Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
28. Ouyang D, He S, Zhang G, et al. Efficient Multi-Scale Attention Module with Cross-Spatial Learning. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2023: 1-5.
29. Zheng Z, Wang P, Liu W, et al. Distance-IoU loss: Faster and better learning for bounding box regression. Proceedings of the AAAI conference on artificial intelligence. 2020, **34**(07): 12993-13000.
30. Tan M, Le Q. Efficientnet: Rethinking model scaling for convolutional neural networks. International conference on machine learning. *PMLR*, 2019: 6105-6114.
31. Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks. International Conference on Neural Information Processing Systems. MIT Press, 2015:91-99.
32. Zhu C, He Y, Savvides M. Feature selective anchor-free module for single-shot object detection. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019: 840-849.
33. He K, Gkioxari G, Dollár P, et al. Mask r-cnn. Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
34. Lu X, Li B, Yue Y, et al. Grid r-cnn. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 7363-7372.
35. Bolya D, Zhou C, Xiao F, et al. Yolact: Real-time instance segmentation. Proceedings of the IEEE/CVF international conference on computer vision. 2019: 9157-9166.
36. Redmon J, Farhadi A. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.