IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 January 2024, accepted 18 February 2024, date of publication 26 February 2024, date of current version 6 March 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3369705

RESEARCH ARTICLE

Experimental Study on YOLO-Based Leather Surface Defect Detection

ZHIQIANG CHEN[1], QIRUI ZHU[1,2], XIAOFAN ZHOU[3], JIEHANG DENG[4], AND WEI SONG[5]

[1]College of Electrical and Information Engineering, Quzhou University, Quzhou, Zhejiang 324000, China
[2]School of Computing, Hangzhou Dianzi University, Hangzhou, Zhejiang 310000, China
[3]Zhejiang Qianjiang Robot Company Ltd., Wenling, Zhejiang 317500, China
[4]School of Computers, Guangdong University of Technology, Guangzhou 510006, China
[5]School of Mechanic Engineering and Automation, Shanghai University, Shanghai 200444, China

Corresponding author: Zhiqiang Chen (czq@qzc.edu.cn)

This work was supported in part by the Basic Public Welfare Research Plan Project of Zhejiang Province under Grant LGG22F010011; in part by the Science and Technology Plan Project of Quzhou City, Zhejiang Province, under Grant 2023K228; in part by the Key Research and Development Project of Wenling City, Zhejiang Province, under Grant 2023G00020; and in part by the Startup Research Found Plan Project funded by Quzhou University, China, under Grant BSYJ202107.

ABSTRACT Accurate, reliable, and fast intelligent detection of leather surface defects has become an important subject in industrial inspection, which aims at improving production efficiency and increasing automation levels. This work focuses on the rapid defect recognition and localization of leather surface defects for industrious applications, which is based on the state-of-the-art real-time detection model YOLO. Three experimental Schemes with different challenges were designed to find the optimal YOLO-based leather surface defect detection scheme. Typical tanned leather surface defect images from the factory were collected, which are comprised of eight types of defects, namely rotten surface, hole, scratch, crease, healed injury, bacterial injury, growth line, and pinhole, which exhibit variations in shapes, sizes, and colors, reflecting the various characteristics found in tanned leather defects. A comprehensive and in-depth review of the YOLO series of models is presented, including YOLOv1 to YOLOv8. The systematic and extensive experiments were conducted,which indicate that the YOLO models can simultaneously detect multiple types of defects present in each leather image. The multi-defect detection task achieved a maximum of 52.3% mean average precision (mAP), 58.2% precision, and 68.7% recall. For single-class detection tasks, the highest performance reached 85.1% mAP, 90.9% precision, and 81.8% recall. These works provide feasible intelligent solutions for surface defects in the leather industry, laying a solid foundation for the design and development of new solutions for leather defect detection.

INDEX TERMS Leather, defect detection, YOLO, deep learning.

I. INTRODUCTION

Under the driving of intelligent manufacturing, automation, and intelligent technologies are rapidly being applied in industrial product inspection. In the leather and leather product manufacturing process, traditional surface defect inspection and grading still heavily rely on manual inspection by personnel, requiring extensive judgmental experience and capabilities. Manual inspection struggles to maintain the stability and consistency of leather surface defect detection [1], and is highly inefficient. This has become a potential bottleneck in leather product manufacturing. Global economic pressures and social development are compelling the reform of these inefficient production processes. Enterprises urgently need to enhance automation and intelligence levels in leather surface quality inspection to address challenges from all directions. Therefore, intelligent inspection of leather surface defects has become a significant topic in industrial inspection.

Leather surface defect detection is a typical object detection task, involving the localization of defects and identification of their respective categories. In the past two decades, intelligent machine vision systems have been the core of industrial inspection and monitoring, and have been widely used in product surface defect detection. Many machine

The associate editor coordinating the review of this manuscript and approving it for publication was Jolanta Mizera-Pietraszko.

vision-based technologies have been developed and applied to leather surface defect detection. These methods can be mainly divided into three categories: methods based on classical image processing techniques, machine learning methods using handcrafted features or shallow learning techniques, and methods based on deep learning [2]. Although since the 1990s, both domestic and international scholars, as well as automatic detection equipment suppliers, have been paying attention to the automatic detection of leather surface defects based on machine vision, many enterprises still rely on traditional manual defect inspection and grading. Some companies have achieved semi-automation with partial human involvement, but a truly fully automated defect detection system for leather surface defects has not been realized in practical applications. The presented achievements are often confined to self-defined and collected leather defect data, leading to insufficient generalization. There have been relatively few studies that have deeply researched this field. Very few studies consider the real-time requirements of leather surface defect detection. To address the challenges in the field of leather surface defect detection, it is necessary to conduct further in-depth research.

In practical applications, leather surface defect detection often requires simultaneous localization and recognition tasks. Most existing work either focuses on identifying the types of surface defects in leather or only segmenting leather defects.This work aims to particularly explore defect detection technology that combines defect recognition and localization on leather surfaces. The YOLO series models, as the state-of-the-art one-stage end-to-end object detection algorithms, have been widely used in the field of real-time object detection. However only a few literature reported they are applied to leather surface defect detection, and there is a lack of in-depth systematic comparison and evaluation for this field. Considering the outstanding performance of YOLO models in real-time object detection, this work focuses on evaluating the performance of the YOLO series model [3], [4], [5], [6], [7], [8], [9], [10], including YOLOv5-YOLOv8, in leather surface defect detection, aiming to answer the following several questions:

(1) How is the performance of the YOLO series models in the field of leather surface defect detection from the first version V1 to the latest version V8 currently released? Do they meet the real-time requirements for leather surface defect detection?

(2) What are the challenges in the field of leather surface defect detection?

II. RELATED WORK

Since the 1990s, traditional image processing methods such as edge detection [11], threshold segmentation [12], [13], leather texture analysis techniques [14], wavelet transform [15], and image saliency analysis [16] have been applied to leather surface defect detection and have shown some effectiveness in reported datasets. However, these algorithms often require multiple thresholds for various defects and are highly sensitive to lighting conditions and background colors. When faced with a new problem, these thresholds need to be adjusted, or these algorithms may need to be redesigned. Furthermore, the testing datasets are relatively small and lack diversity in defects, and they do not consider the dynamic changes in leather defects. As a result, it is challenging to ensure the generalization performance of these algorithms [2].

A large number of machine learning methods have also been used for leather surface defect type recognition, with a focus on identifying the types of defects [15], [17], [18], [19], [20], [21]. However, most of these approaches were tested on custom small local datasets, lacking comparable benchmarks and comprehensive evaluations. Their work mainly focused on defect presence or recognition of a few defect types, without considering the other crucial task of defect localization in object detection. They also did not address the issue of dynamic changes in leather defects, leading to limited generalization capabilities of their models. The stability and effectiveness of their methods need to be evaluated on various types of defects in real-world leather samples in an industrial environment, indicating limitations in practical applications. In our early research [2], [40], various image processing methods based on edge detection and threshold segmentation, as well as traditional machine learning approaches, were developed for leather surface defect detection. However, the overall performance of these methods was not satisfactory. While the traditional machine learning methods achieved approximately 80% defect recognition accuracy in simple application scenarios, their accuracy rapidly declined when dealing with more complex scenarios.

Aslam et al. [22] believes that deep learning holds great promise in developing new solutions for leather surface defect inspection. Some researchers have also developed corresponding solutions [23], [24], [25], [26]. Although Liong's team conducted an in-depth exploration, their work was also limited to a small local dataset. Despite the use of GAN(Generative Adversarial Network)-generated data, the overall dataset size was still relatively small, and the variety of defects was quite limited. In our early research, 26 deep learning models were evaluated for leather surface defect type recognition, including ResNet [27], GoogleNet [28], DenseNet [29], AlexNet [30], VGG [31], SqueezeNet [32], and Shufflenet [33], for leather surface defect recognition. The results demonstrated that deep learning models have promising potential in the field of leather surface defect detection and can replace some manual labor in industrial applications. The above work demonstrates that deep learning models have great potential in the field of leather surface defect detection. However, several challenges still exist:

(1) Limited data: Leather defect datasets are relatively small and may not cover all types of defects with varying morphologies. It is difficult to obtain large-scale datasets, similar to ImageNet, for training.

(2) Research on the dual priority of localization and recognition tasks is still rare: Leather defect detection is a target

detection task that involves both localization and recognition. Efficiently achieving good performance in both sub-tasks is essential. Current achievements focus on different applications, often emphasizing either localization or recognition, but not both with equal priority.

(3) Leather defects exhibit multiple spatial scales and aspect ratios: Leather defects occur at various spatial scales, and even defects of the same type can exhibit significant variations in size and shape. The aspect ratios of bounding boxes used for localization can also differ substantially.

(4) Class imbalance: The number of samples for different defect classes may vary significantly, leading to class imbalance issues.

(5) Real-time detection requirements: Practical applications often require real-time detection of leather surface defects to respond promptly to production needs.

(6) This work is committed to addressing the aforementioned challenges of dual priority of localization and recognition tasks, as well as the requirement for real-time detection by researching on the application of the YOLO model for leather surface defect detection.

III. YOLO MODELS

A. OVERVIEW OF MODELS

Real-time object detection has become a crucial component in various applications, playing a significant role in fields such as autonomous driving vehicles, robotics, video surveillance, and augmented reality. Among numerous object detection algorithms, the YOLO [3] framework stands out for its exceptional balance between detection speed and accuracy, enabling fast and reliable identification of objects in images. Despite the continuous emergence of approaches like transform [34] and its variants that have made waves in the object detection task, the industrial sector still widely supports the application of the YOLO model, as it considers the trade-off between model training cost, detection speed, and accuracy.

In terms of detection performance, the most significant feature of the YOLO series models and their subsequent versions is that they have fewer false positive predictions in the background, leading to higher recall rates. This feature makes the YOLO model particularly suitable for industrial use. Especially for leather production companies, accurate defect detection is crucial, and the impact of false negatives is greater than that of false positives. In addition, the YOLO model can better learn the abstract representation of the target. They utilize information from the entire image for prediction, rather than region based methods. Therefore, compared with the R-CNN model [35], this model implicitly encodes contextual information of different categories and shapes, thereby improving the detection performance of abstract objects such as artworks. Since its inception, the YOLO family has undergone several iterations, the release timeline for YOLO series models is shown in Figure 1. The following sections provide a detailed description of the YOLO model by different teams.

The detailed evolution process is shown in Table 1. The backbone network has evolved from darknet19 to the improved CSPDarkNet53. The neck network has evolved from none to incorporating the FNP (Feature Pyramid Network) [36], and then to the SPP (Spatial Pyramid Pooling) +PAN (Pyramid Attention Network). The detection head has evolved from the initial fully connected to fully convolution coupled detection head, and then to the advanced decoupled detection head. The activation function has evolved from Leaky ReLU to SiLU. Many of the ideas in these network structures represent high-value models at the time in the field of artificial intelligence.

Besides network architecture, the loss function is also a significant area of research in the field of object detection. The loss function in the YOLO series models consists of three components: confidence error loss, bounding box regression loss, and classification error loss. Each component contributes to the specific task's loss, and the ideas behind these loss functions were also at the forefront of technology at the time [37]. The specific evolution process of these loss functions (Binary Cross Entropy (BCE), Mean Squared Error (MSE), Virtual Adversarial Training Loss (VFL), and Distribution Focal Loss (DFL) is illustrated in Table 2. Indeed, the innovations in the subsequent versions of YOLO extend beyond network architecture and loss functions.

Since the publication of YOLOv1 in 2016, its unique and effective innovations have attracted numerous researchers to continuously optimize and improve its foundational structure, giving rise to the YOLO series of models. Moreover, this trend of development is still ongoing. Many outstanding optimization ideas within the YOLO family are still considered state-of-the-art and widely applicable. These include various data augmentation methods, training techniques, and more. A variety of innovative approaches in the YOLO series complement each other, continuously improving detection performance, reducing training and inference costs, and accelerating inference speed.

B. YOLOv5∼v8

In the evolution process of the YOLO model, YOLOv5 model is a milestone achievement. Its high-quality and efficient code implementation standardizes the development of YOLO-based model. Subsequent versions of the YOLO model have made performance improvements from different perspectives. Therefore, in this work, we mainly focus on the application development of the YOLOv5∼v8 family in the field of leather surface defect detection. Figure 2∼5 present their network architecture diagram.

YOLOv5: The most significant highlight of YOLOv5 model is the implementation of model pruning. There are five different scaled versions of YOLOv5: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The implementation of the YOLOv5 largely continues the YOLOv4. However, some optimizations have been made in the details. Substantial feature extraction on the image in the first convolutional layer

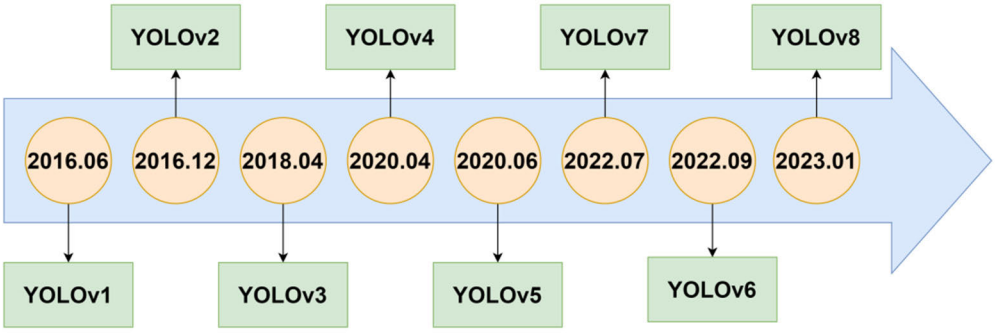


FIGURE 1. Release timeline for YOLO series models.

TABLE 1. The evolution process of the network structure of YOLO series models.

Model	network structure			Label Allocation
	backbone	Nick	Head	
YOLOv1	Improved GoogleNet	-[1]	Fully Connected	IOU threshold matching
YOLOv2	darknet-19	-	Conv, anchor boxes	IOU threshold matching
YOLOv3	darknet-53	FPN	Conv, anchor boxes, scale detection logic, Multi-label classification	IOU threshold matching
YOLOv4	CSPDarkNet53	SPP, PAN	Conv, anchor boxes, scale detection logic, Multi-label classification	Class-agnostic Regression
YOLOv5	CSPDarkNet53	SPP, PAN	Conv, anchor boxes, scale detection logic, Multi-label classification, Auto Learning Bounding Box	GIoU threshold matching
YOLOv7	ELAN, MPConv,	SPPSCP improved PAN	Conv, anchor boxes, scale detection logic, Multi-label classification, Auto Learning Bounding Box, RepVGG style, auxiliary Head	coarse-to-fine lead guided
YOLOv6	EfficientRep	Rep-PAN	decoupled detection head, anchor free	Task Alignment Learning
YOLOv8	improved CSPDarkNet53	PAN-FPN	decoupled detection head, anchor free	Task Aligned Assigner

The symbol "-" indicates that this structure is not present in this version.

is performed. The original SPP layer has been replaced with the SPPF (spatial pyramid pooling fast) layer, which serves the same purpose of handling features at different scales but with reduced computational cost. Moreover, the model incorporates the adaptive anchor box method. It analyzes the target boxes in the training set using clustering algorithms to determine the statistical characteristics of target boxes with different scales and ratios. Based on the clustering results, anchor boxes that adapt to the target features are automatically generated. These prior anchor boxes are better suited

TABLE 2. Evolution process of loss function in models of various versions in YOLO series.

Model	loss function		
	Confidence error loss	box regression loss	classification error loss
YOLOv1	BCE	MSE	BCE
YOLOv2	BCE	MSE	BCE
YOLOv3	BCE	MSE	BCE
YOLOv4	BCE	CIoU	BCE
YOLOv5	BCE	CIoU	BCE
YOLOv7	BCE	CIoU	BCE
YOLOv6	-	SIoU/GIoU	VFL
YOLOv8	-	DFL + CIOU	BCE

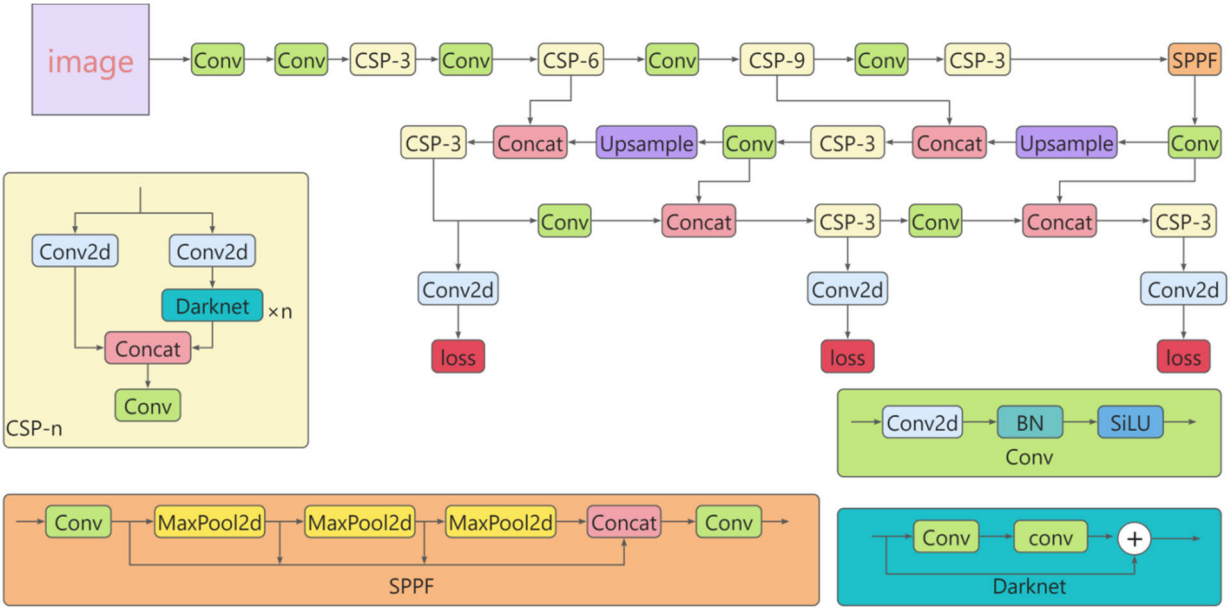


FIGURE 2. The network architecture of YOLOv5.

for the network's learning process. Furthermore, the model adopts a domain-positive sample allocation strategy to balance the class imbalance by adjusting the weights of positive samples.

YOLOv6: In September 2022, the team at Meituan Inc., led by Li et al., introduced the YOLOv6 model. This new model not only includes improvements in the network architecture but also incorporates many practical industrial enhancements. The authors incorporated state-of-the-art network designs, training strategies, testing techniques, quantization, and optimization methods available at that time. Additionally, they integrated their team's practical experience and unique insights from working with YOLO models.

The YOLOv6 model incorporates an efficient network called EfficientRep as its backbone, which is designed based on the RepVGG style structure. The RepVGG style structure outperforms in hardware computational capability, memory bandwidth, compilation optimization features, and network representation capability. The backbone network of the model is not a fixed structure but depends on the model's complexity. For small models, RepBlock is used to construct computational blocks, while for larger models, a more efficient CSPStackRep block is used. These blocks heavily utilize stacks of 3×3 convolutional layers, making full use of hardware computational capacity. The neck network of the model utilizes an improved PAN structure called RepPAN, which replaces the CSP modules in the PAN structure of the YOLOv5 model with Rep modules. The detection head of YOLOv6 is an efficient decoupled head with a mixed-channel strategy, inspired by the decoupled head in the YOLOX model. The author named it the Efficient Decoupled Head. Unlike the coupled detection head of YOLOv5, the decoupled head of YOLOX separates the classification and regression branches and adds two additional 3×3 convolutional layers to improve accuracy. However, this brought additional computational costs. To address this issue, the authors designed

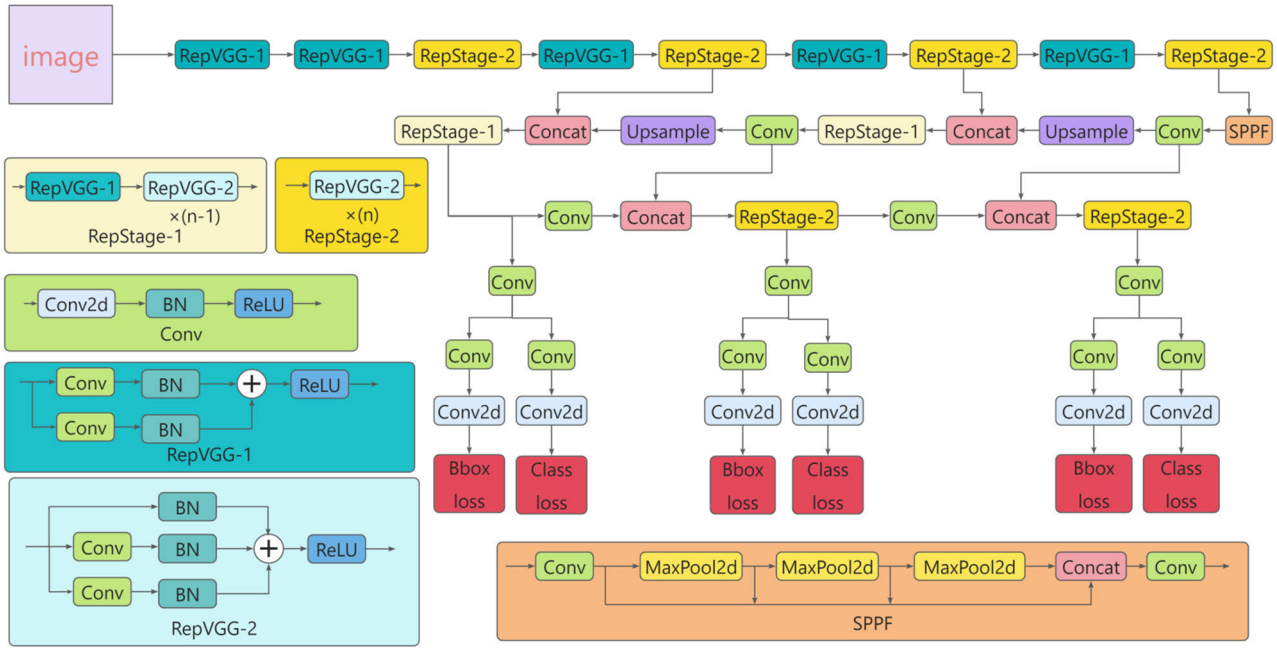


FIGURE 3. The network architecture of YOLOv6.

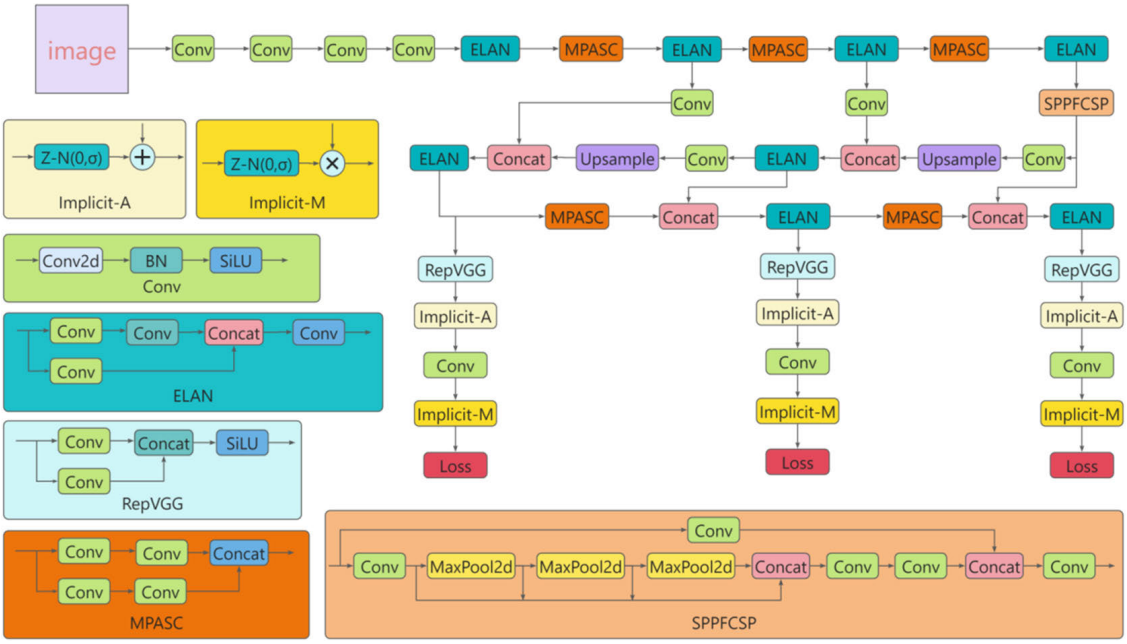


FIGURE 4. The network architecture of YOLOv7.

a more efficient decoupled head structure using the Hybrid Channels strategy, which maintains accuracy while reducing latency.

In the YOLOv6 model, there is an improvement in the label assignment strategy, where they adopt Task Alignment Learning (TAL) [38] as the default label assignment approach. TAL introduces a unified metric for both classification score and prediction box quality, replacing IoU (Intersection over Union), to assign object labels. This new metric helps address the problem of misalignment between the tasks of classification and bounding box regression to some extent. By using TAL, the model can better align the two tasks and improve the overall performance of object detection.

YOLOv7: In July 2022, Wang et al. released YOLOv7 model. Similar to YOLOv4, YOLOv7 introduced architectural changes and a series of ''bag of freebies'' techniques,

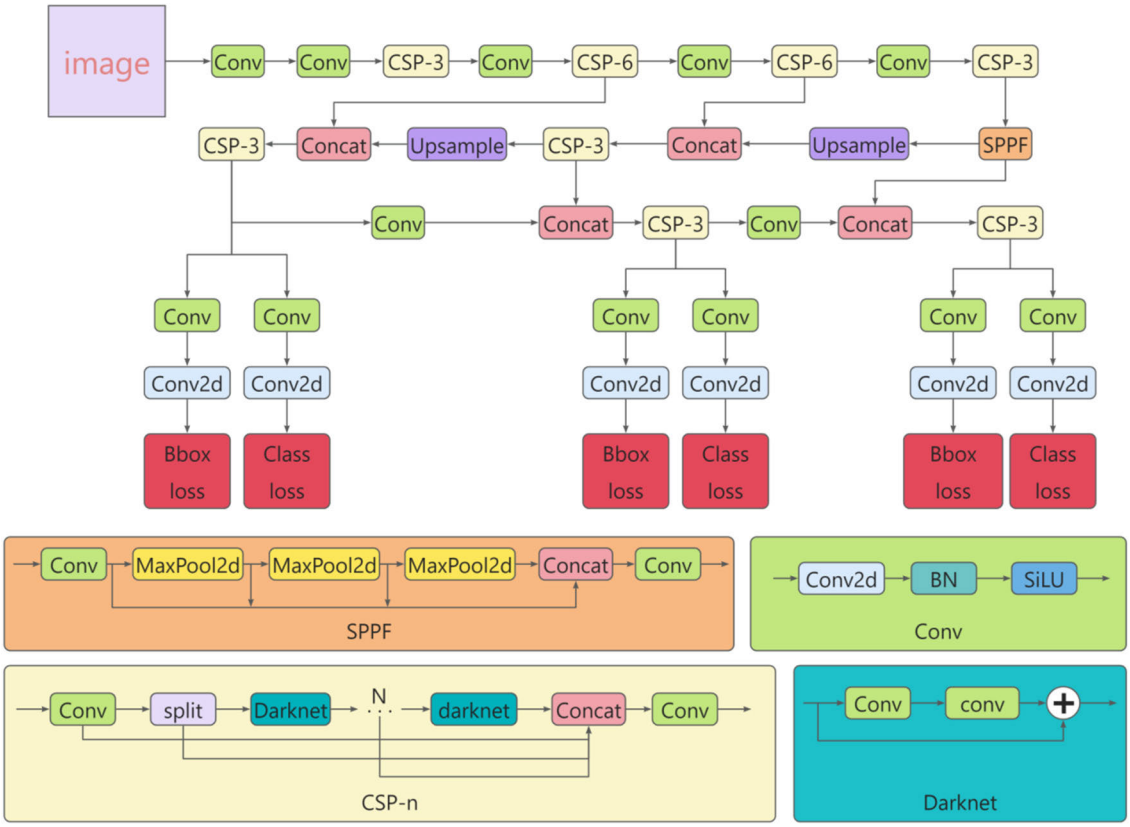


FIGURE 5. The network architecture of YOLOv8.

enabling the model to achieve detection speeds and accuracy within the specific range of 5FPS to 160FPS, surpassing all known detectors at that time.

YOLOv7 model introduced a novel network architecture called Extended Efficient Layer Aggregation Network (E-ELAN) as its backbone. E-ELAN is based on ELAN [39], which is a strategy that controls the shortest and longest gradient paths to enable more effective learning and convergence in deep models. E-ELAN is specifically designed for models with infinitely stacked computational blocks, significantly improving the learning and convergence performance of deep models while preserving the original gradient paths to avoid model degradation. E-ELAN achieves this by employing the ''expand, shuffle, and merge cardinality'' methods to combine features from different groups without disrupting the original gradient paths, thereby continuously enhancing the network's learning capabilities.

In the neck network, the YOLOv7 model incorporates the SPPCSP module and optimized PAN module. The SPPCSP module is an extension of the SPP module, where a concatenation operation is added to fuse the output of the SPP module with the feature map before the SPP operation. This further enriches the feature information. On the other hand, the optimized PAN module builds upon the PAN module and adopts the E-ELAN idea, leveraging strategies like expand, shuffle, and merge cardinality to enhance the network's learning capabilities without disrupting the original gradient paths. Significant improvements to the detection head of YOLOv7, the RepVGG style network structure was introduced. It's worth noting that the training model and inference model do not share the same network architecture. During the training process, multiple branches are used to enhance the network's learning capabilities. In contrast, during the inference process, structural reparameterization is employed to accelerate the inference speed without sacrificing performance.

YOLOv8: In January 2023, YOLOv8 model code was released. The overall code style and network structure of the model also have some similarities with YOLOv5. The backbone network has been modified from CSPDarkNet53 by replacing YOLOv5's C3 structure with a more gradient-rich C2f structure. Additionally, different scale models have been adjusted with varying channel numbers, showing a carefully tuned model architecture to enhance detection performance. The neck network in YOLOv8 utilizes the PAN-FPN structure, which is inspired by the PANet's backbone network. The detection head of the model adopts the decoupling head structure of the YOLOv6 model, as well as the anchor-free idea. Additionally, the model adopts the Task Aligned Assigner as the label assignment strategy. For training, the model refers to YOLOX by disabling Mosiac augmentation in the last 10 epochs.

(a) Cow skin

(b) Sheep skin

FIGURE 6. Whole-skin images.

By integrating these cutting-edge ideas and modules, YOLOv8 further improves the detection performance of the YOLO series. On the publicly available COCO dataset, the model achieves unprecedented heights in both detection accuracy and speed, making YOLOv8 become a state-of-the-art model.

IV. EXPERIMENTAL DESIGN

A. DATA COLLECTION AND ANALYSIS

1) ULTRA-HIGH-DEFINITION WHOLE-SKIN IMAGING

There are three main characteristics of leather surface defect imaging [40]: (1) large imaging area: a whole skin area of up to 2m×3m; (2) small defect size: the defect area can be as small as 150um×150um, the maximum average diameter of the thin spot is approximately 0.98 mm, and the minimum average circular spot diameter is approximately 1.20 mm, and (3) the leather surface is a textured surface and the defects are usually hidden behind the irregular textured background of the leather surface. Therefore, the image acquisition of leather defects requires a large camera view and high resolution. A high-resolution imaging system of the same literature [64] is used to capture full-skin images and collect a large amount of leather surface defect images from factories. The imaging system consisted of a leather fixed platform, an ultra-high-definition CCD camera (resolution of 8688 × 5792 pixels), a light source, and an image processing workstation. To distinguish leather from the background, the color of the leather fixing platform was fixed to blue after counting the color of the leather. Figure 6 shows the original whole-skin image samples.

2) DATA ANNOTATION AND OBSERVATION

All data came from a leather production enterprise in Guangdong Province, China. The collected leather defect images are common in actual production. Combining Aslam et al.'s [22] definition of leather surface defects, after annotation by experienced engineers in the factory, 2855 images of leather surface defects with a size of 2268 × 4432 pixels were obtained, including the following eight defects: cavity, pinhole, scratch, rotten surface, growth line, healing wound, crease, and bacterial wound. The open-source annotation tool called ''labelme'' was used for annotating defects.

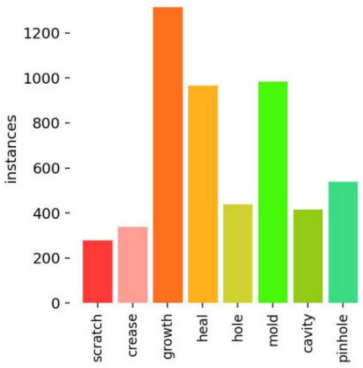
To observe the defect color, shape, and size of the collected datasets and more intuitively express the diversity of datasets, qualitative and quantitative analyses were conducted on the datasets. The visualized statistics of the dataset are shown in Figure 7, depicting the distribution of data in the training set. Growth lines, healed injuries, and bacterial injuries are the most common types of surface defects in leather, as they have a higher number of instances in the dataset. The size of the bounding boxes is primarily concentrated in medium-sized ones around the image center, with larger boxes around the edges. Various sizes of boxes coexist, indicating that the dataset possesses a certain degree of sample balance, effectively capturing the position uncertainty and size variability of surface defects in leather as semantic features. As shown in Figure 7, the defect data in datasets remain diverse in size, position, shape, texture, and color. Compared to previous research, the collected data is not limited to a single type of animal skin but includes both sheepskin and cowhide, and the whole-hide imaging without any stitching operation ensures stability and consistency.

B. EXPERIMENTAL SCHEMES

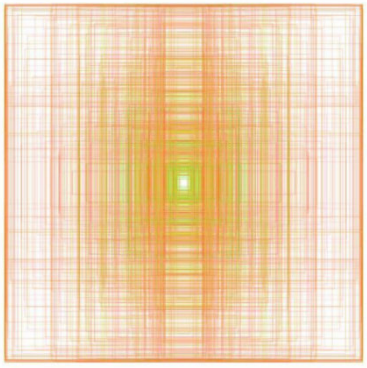
This work aims at explore a suitable leather surface defect detection scheme based on the YOLO model. For this purpose, three experimental schemes were designed. Corresponding to these schemes, three datasets were constructed, whose details are shown in Table 3.

Scheme I: Multi-Class Defect Detection Experiment With Input of Large-Size Images: Here data set is made up of 1250 leather surface images with distinct defects like Figure 8, each sized 4032 × 2268 pixels, which is named Dataset I. The number and categories of leather surface defects in each image are random but ensured to have at least one defect object per image. This dataset includes eight types of defects, namely cavity, pinhole, scratch, rotten surface, growth line, healing wound, crease, and bacterial wound. In the top left corner of the figure, it shows a bar chart representing the number of instances for each defect in the data set. Some examples and annotations are presented in Figure 9.

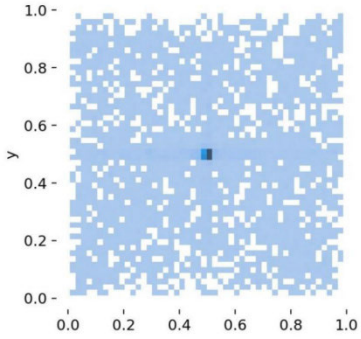
Scheme II: Multi-Class Defect Detection Experiment With Input of Medium-Size Images: Considering the high resolution of the image data in Dataset I, which brings a significant computational burden and limits the model's performance, we performed another round of cropping and obtained 10,000 leather surface defect images with a



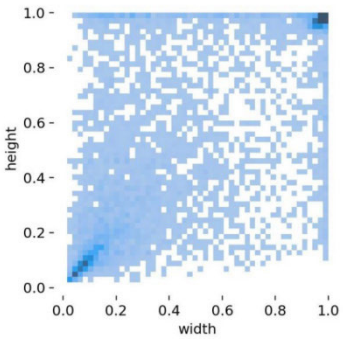
(a) The number of instances or each defect in data set



(b) Overall distribution of all annotated bounding boxes



(c) Distribution of the center coordinates of all annotated bounding boxes



(d) Distribution of the sizes of all annotated bounding boxes.

FIGURE 7. The data set statistics visualization.

TABLE 3. Details of datasets I ∼ III.

Dataset	Scratch	Crease	Growth line	Healed injury	Hole	Bacterial injury	Rotten surface	Pinhole	Background image size		Total
I	217	111	126	218	134	109	102	149	84	4032×2268	1250
II	353	457	1753	1113	603	1037	530	740	3414	1008×1134	10000
III	The images used in each subset are the same as in Dataset II, with the only difference being that each subset focuses exclusively on a single type of defect, while other types of defects are set as "background".										

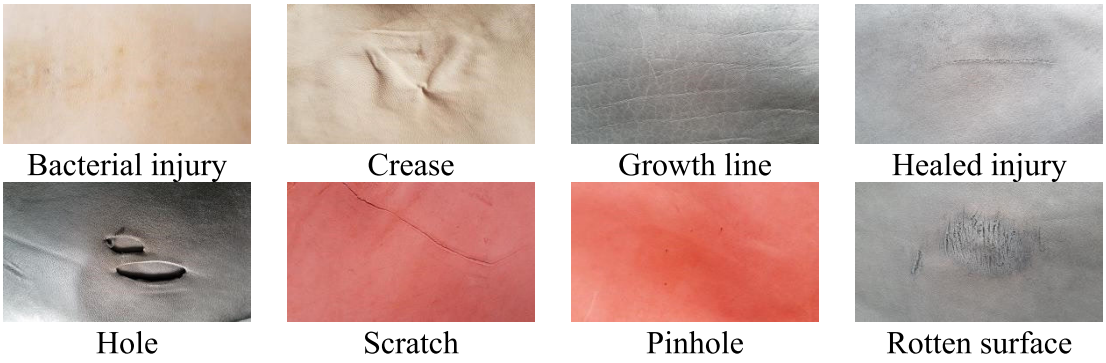


FIGURE 8. Examples of eight types of leather surface defects.

resolution of 1008 × 1134, forming Dataset II. The distribution of instances in the dataset is shown in Table 3. Similar to Dataset I, the defective objects are randomly distributed in the images, but this dataset also includes images without defects, referred to as ‘‘background’’ images. Some examples and annotations are presented in Figure 10.

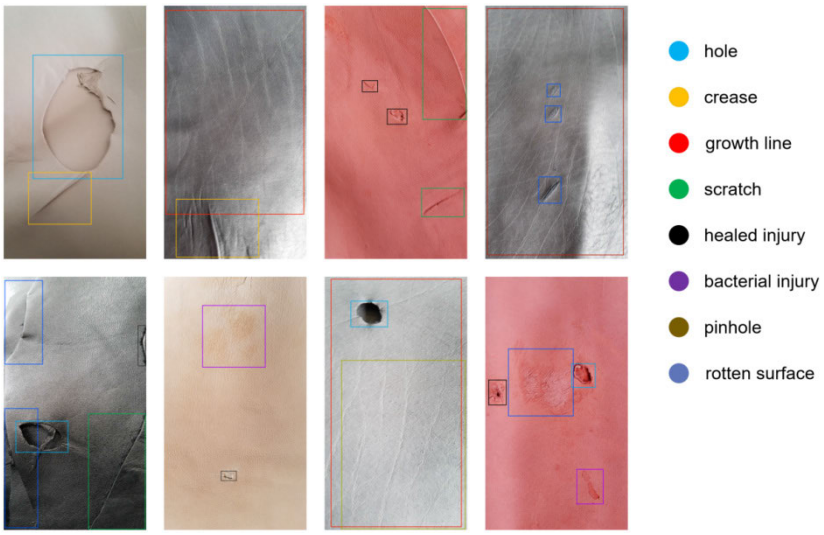


FIGURE 9. The figures display some visualizations of annotated samples from Dataset I.

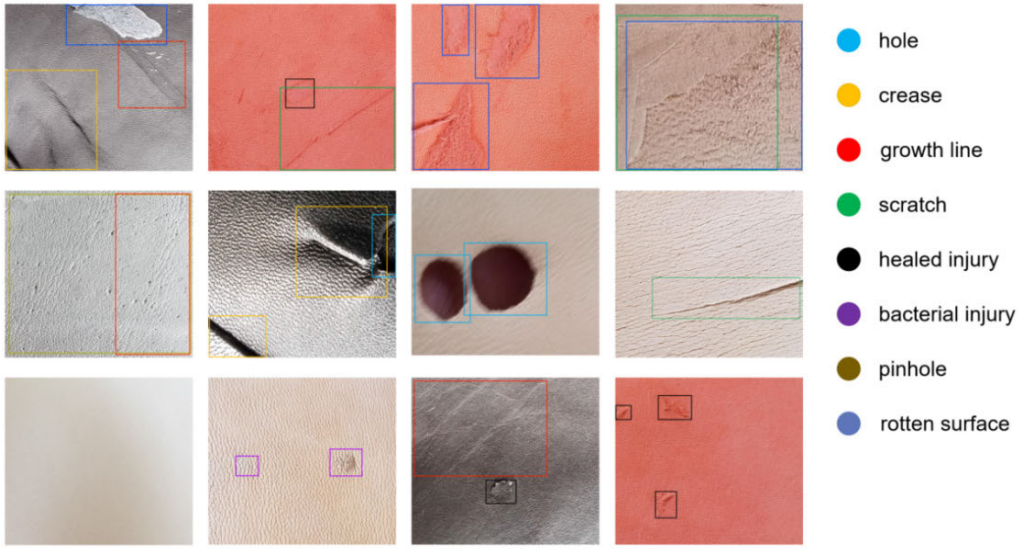


FIGURE 10. The figures display some representative visualizations of annotated samples from Dataset II. The characteristic of this dataset is that the appearance of defects in each image is flexible, which includes images without defects (e.g., the sample in the bottom left corner), images with multiple defects, and images with various types of defects.

Scheme III: Single-Class Defect Detection Experiment With The Input of Medium-Size Images: Both Scheme I and Scheme II detect multiple defects simultaneously, which poses great challenges. In practical applications, only one type of defect is often detected at a time. Considering the practical needs for defect detection of individual types in industrial applications, Dataset II was further divided into separate subsets, forming a total of 8 sub-datasets: Rotten surface defect, Crease defect, Growth line defect, Scratch defect, Hole defect, Healing injury defect, Pinhole defect, and Bacterial injury defect subset. We collectively refer to these subsets as Dataset III. The images used in each subset are the same as in Dataset II, with the only difference being that each subset focuses exclusively on a single type of defect, while other types of defects are set as ''background''.

C. EXPERIMENTAL CONFIGURATION

The experimental hardware adopted is the Inspur Yingxin server NF5280M6 with the GPU graphics card NVIDIA A40 having 46GB of memory. The software environment was Ubuntu 18.04 LTS operating system. All the evaluated deep learning models were derived from PyTorch (version 1.13.0). A total of 25 YOLO models were evaluated experimentally. Furthermore, for each version of the YOLO model, specific training parameters as Table 4 were configured for the following experiments. Considering the significant impact of

missing defects compared to false alarms in industrial production, the following detection metrics were selectively adopted for evaluation:

Precision: It refers to the proportion of true positive samples among the detected positive samples. Precision = TP / (TP + FP), where TP represents the number of true positive samples and FP represents the number of false positive samples.

Recall: It represents the proportion of correctly detected positive samples among all the actual positive samples. In other words, it measures how many of the existing targets are detected. Recall = TP / (TP + FN), where TP represents the number of true positive samples and FN represents the number of false negative samples.

mAP: It is the average of the Average Precision (AP) for all classes, calculated at an IoU threshold of 0.5. mAP is commonly used to assess the performance of object detection algorithms. AP is computed as the area under the Precision-Recall curve.

FPS (frames per second): This value indicates how many images the model can process per second with a batch size of 1. A higher value indicates a faster detection speed of the model.

V. EXPERIMENTAL EVALUATION

A. PERFORMANCE EVALUATION FOR EXPERIMENTAL SCHEME I

This scheme mainly examines the performance of detecting multiple defects simultaneously under high-resolution input. Dataset I is characterized by a large resolution (4032 × 2268) and a relatively small original dataset size (1250 images). Based on Dataset I, the defect recognition performance of a total of 10 models, including all versions of YOLOv5 n/s/m/l/x and YOLOv8n/s/m/l/x were evaluated on high-resolution leather surface defect images. Each version of the model varies in terms of the number of layers and parameters.

The results of multi-defect detection are shown in Figure 11, with the highest recall rate achieved by the YOLOv5l model at 49.6%, and the highest precision and mAP achieved by the YOLOv8 model in its x and s versions, 55.2% and 47.8%, respectively. In Table 5, the defect detection performance of all versions of the YOLOv5 and YOLOv8 models was presented, revealing a significant discrepancy in how these models detect defects. While some defects such as hole can achieve mAP of over 80%, others pose more challenges, with mAP below 20%. Figure 12 displays some detection results, indicating that the models are capable of simultaneously detecting multiple defects on leather surface images with accurate localization and classification. Visualizations of the Confusion Matrix for YOLOv5l and YOLOv8x were selected and displayed in Figure 13.

From the detection result, it can be observed that the ''hole'' defect is the easiest to detect, with a maximum accuracy of 80%. The next most manageable defects are ''crease,''

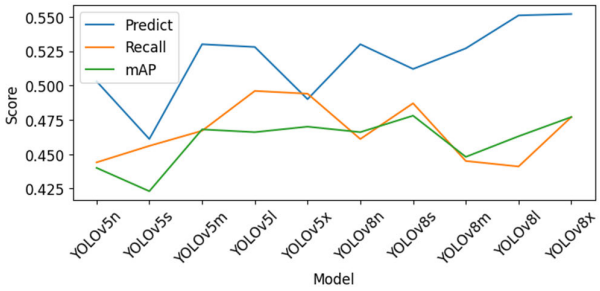


FIGURE 11. A comparison for multi-defect detection tasks on Scheme I for YOLOv5n to YOLOv8x.

''growth line,'' and ''bacterial injury,'' all with accuracy above 50%. On the other hand, the most challenging defects are ''scratch'' and ''pinhole.'' The former only achieves 16% detection accuracy, with 80% of instances misclassified as background, and the latter achieves only 20% detection accuracy, with 79% of instances being missed. Similarly difficult to detect are the ''rotten surface'' and ''healed injury'' defects, with misclassification rates above 50%. The experimental results indicate that the constructed Scheme I poses certain challenges, with higher rates of missed detections observed for small-sized defects. The overall convergence is smooth, and there is no issue of underfitting due to too few samples.

B. PERFORMANCE EVALUATION FOR EXPERIMENTAL SCHEME II

Can the lower detection accuracy of Scheme I be attributed to the large image resolution in Dataset I? Is it possible that this will limit the performance of the YOLO models? To verify this hypothesis, Experimental Scheme II is constructed and further experimental verification is conducted. A more detailed evaluation of the YOLO v5-v8 models on Scheme II, which has a smaller resolution, was conducted based on Dataset II. In this data set, defects are relatively larger in size, making them more prominent and easier for the model to detect. This smaller resolution reduces the computational burden on the model, allowing for a larger batch size during training, which enhances the model's batch-processing capabilities. Additionally, compared to Dataset I, Dataset II has a significantly increased overall sample size, providing the model with a larger and more diverse set of training data.

The specific experimental results are shown in Figure 14, Table 6. Figure 14 presents Precision, Recall and mAP of all models in this evaluation. Table 6 provides a detailed listing of all experimental data in SchemeII. Among these YOLOv5 models, YOLOv5l achieved the highest precision at 57.2%, while YOLOv5x obtained the highest recall and mAP50 at 50.6% and 51.3%, respectively. Among the YOLOv6 series of models, the YOLOv6l version performed the best on Dataset II, with precision, recall, and mAP50 reaching 53.7%, 68.7%, and 49.9%, respectively. YOLOv7 series models did not perform well in this round of evaluation,

TABLE 4. Some of the training parameters that were uniformly set during model training.

Model Parameter	Value	Parameter Explanation
epoch	300	number of training epochs
lr0	0.01	Initial learning rate
lrf	0.2	Final learning rate ratio
momentum	0.937	Momentum for optimization
weight_decay	0.0005	Weight decay for regularization
warmup_epochs	3	Number of warm-up epochs
warmup_momentum	0.8	Momentum value during warm-up
warmup_bias_lr	0.1	Bias learning rate during warm-up
hsv_h	0.015	Hue parameter for HSV color augmentation
hsv_s	0.7	Saturation parameter for HSV color augmentation
hsv_v	0.4	Value parameter for HSV color augmentation
translate	0.2	Range of translation for affine transformation augmentation
scale	0.9	Range of scaling for affine transformation augmentation
fliplr	0.5	Probability of left-right flipping augmentation
mosaic	1	Probability of Mosaic data augmentation
mixup	0.15	Probability of MixUp data augmentation

TABLE 5. The performance for all versions of YOLOv5 and YOLOv8 models.

Model	Metrics	all	scratch	fold	growth line	healed injury	hole	bacterial injury	rotten surface	pinhole
YOLOv5n	map	0.44	0.37	0.222	0.731	0.567	0.77	0.433	0.187	0.241
	recall	0.444	0.466	0.222	0.721	0.476	0.76	0.339	0.247	0.323
YOLOv5s	map	0.423	0.265	0.147	0.737	0.532	0.795	0.487	0.189	0.232
	recall	0.456	0.448	0.209	0.744	0.405	0.801	0.475	0.233	0.333
YOLOv5m	map	0.468	0.405	0.17	0.733	0.628	0.838	0.458	0.234	0.277
	recall	0.467	0.495	0.174	0.721	0.524	0.82	0.407	0.205	0.392
YOLOv5l	map	0.466	0.443	0.129	0.727	0.693	0.864	0.359	0.228	0.288
	recall	0.496	0.586	0.198	0.744	0.662	0.86	0.322	0.219	0.376
YOLOv5x	map	0.47	0.481	0.169	0.718	0.59	0.841	0.411	0.224	0.327
	recall	0.494	0.586	0.224	0.791	0.476	0.88	0.362	0.233	0.376
YOLOv8n	map	0.466	0.455	0.174	0.773	0.683	0.809	0.418	0.193	0.211
	recall	0.461	0.466	0.198	0.726	0.645	0.82	0.475	0.137	0.226
YOLOv8s	map	0.478	0.438	0.158	0.75	0.638	0.816	0.541	0.216	0.233
	recall	0.487	0.517	0.279	0.744	0.381	0.84	0.576	0.205	0.306
YOLOv8m	map	0.448	0.433	0.112	0.707	0.576	0.833	0.402	0.266	0.252
	recall	0.445	0.501	0.264	0.674	0.5	0.779	0.22	0.26	0.364
YOLOv8l	map	0.463	0.48	0.119	0.736	0.587	0.839	0.395	0.221	0.326
	recall	0.441	0.414	0.186	0.674	0.643	0.84	0.276	0.178	0.316
YOLOv8x	map	0.477	0.463	0.203	0.684	0.687	0.823	0.458	0.198	0.297
	recall	0.477	0.638	0.186	0.648	0.619	0.8	0.407	0.151	0.366

with the lowest detection precision of only 50.2%. Other metrics, including recall and mAP50, reach 54.6% and 50.7% respectively. This discrepancy is related to the different versions of YOLOv7. The final evaluation included the YOLOv8 series, which represents the most recent version of YOLO. In this evaluation, the YOLOv8 series models demonstrated the best detection precision and mAP50 among the entire series. Specifically, the YOLOv8m model achieved a precision of 58.2%, while the YOLOv8x model reached a mAP50 of 52.3%.

Compared to Scheme 1, all evaluation indicators have improved. It is evident that the improvement in the dataset has led to a notable performance boost for the YOLO models. For the same model, the mAP has increased by more than 4.5%, and the detection precision has improved by over 3%. Certain challenging defect types in Dataset I, such as scratches and pinhole defects, achieved satisfactory detection results in this evaluation experiment. By comparing the confusion matrices in Figure 15 and Figure 13, an overall improvement in the detection performance on Dataset II was

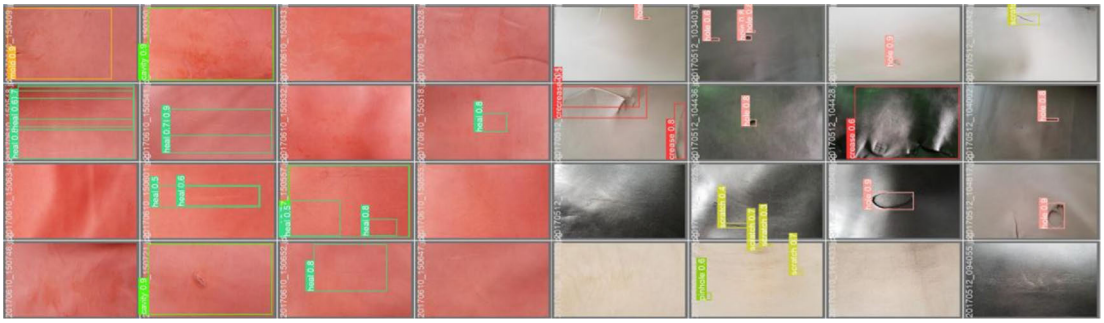


FIGURE 12. Visualization of the YOLOv8x model for detection on Scheme I.

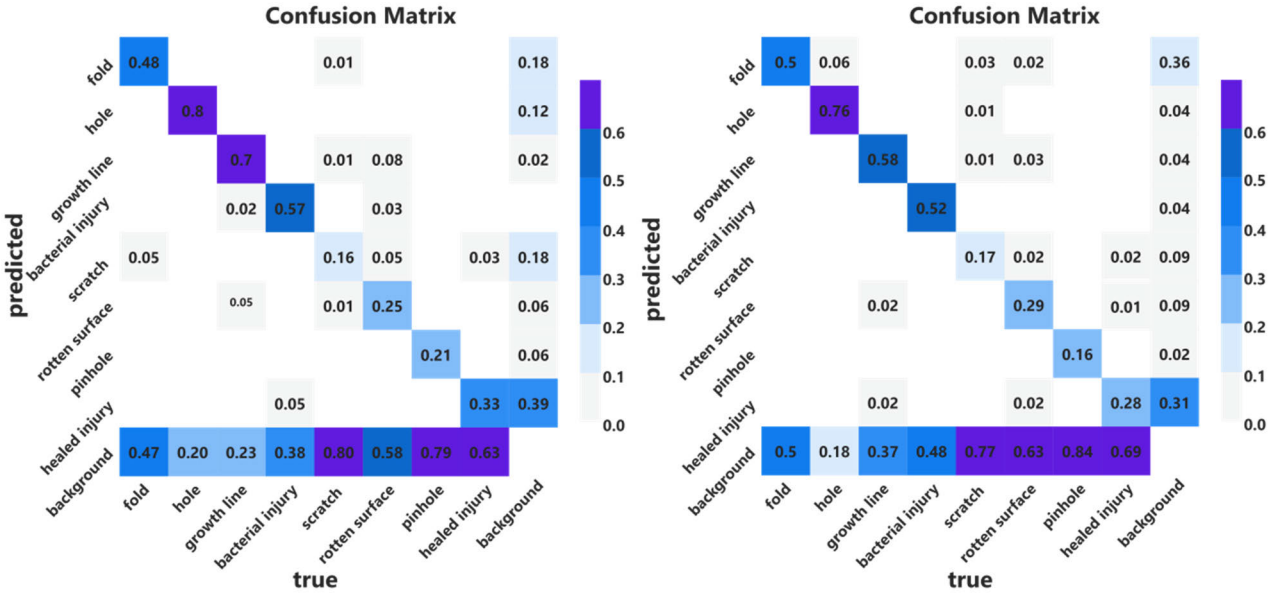


FIGURE 13. Confusion Matrix for Multi-Category Detection in Scheme I for YOLOv5l and YOLOv8x.

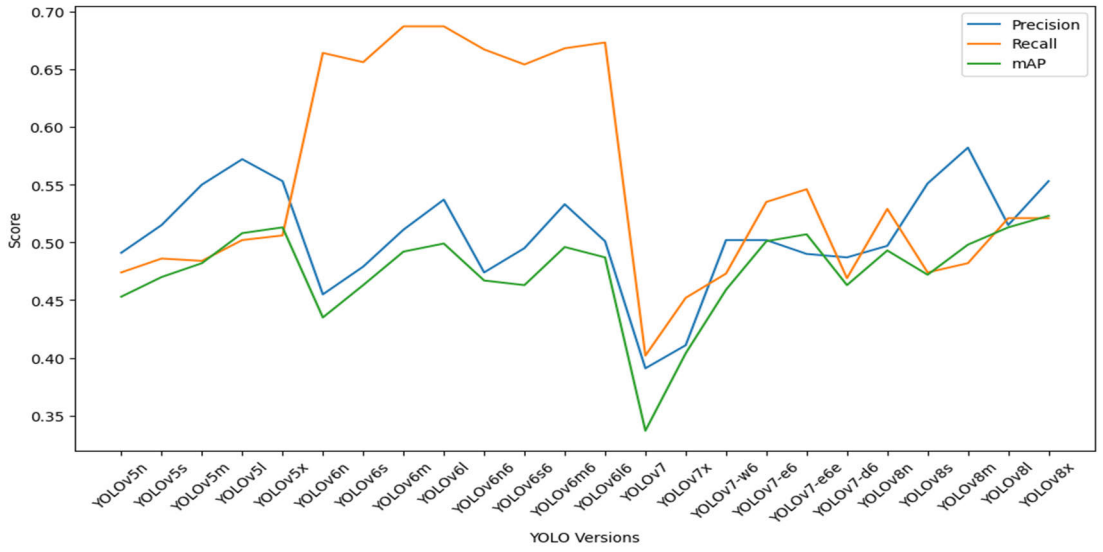


FIGURE 14. Comparison chart of metrics for the YOLO family of models for Scheme II.

observed. As shown in Table 6, It is noticeable that the detection accuracy for most defect types has increased, and the overall rate of missed detections for all defects has decreased. Furthermore, some defect types have shown a significant

TABLE 6. Detailed experimental results based on scheme II.

Model	Metrics	all	scratch	fold	growth line	healed injury	hole	bacterial injury	rotten surface	pinhole
YOLOv5n	map	0.44	0.212	0.316	0.543	0.424	0.646	0.394	0.415	0.676
	recall	0.479	0.236	0.349	0.545	0.47	0.627	0.414	0.456	0.694
YOLOv5s	map	0.47	0.179	0.377	0.553	0.456	0.681	0.393	0.435	0.683
	recall	0.486	0.244	0.405	0.525	0.488	0.671	0.414	0.483	0.661
YOLOv5m	map	0.482	0.256	0.414	0.565	0.475	0.679	0.37	0.433	0.667
	recall	0.484	0.301	0.451	0.552	0.453	0.653	0.365	0.459	0.638
YOLOv5l	map	0.508	0.283	0.4	0.592	0.501	0.724	0.396	0.496	0.672
	recall	0.502	0.276	0.477	0.584	0.552	0.67	0.38	0.46	0.648
YOLOv5x	map	0.513	0.289	0.403	0.594	0.507	0.731	0.392	0.508	0.685
	recall	0.506	0.27	0.49	0.605	0.5	0.657	0.393	0.488	0.645
YOLOv6n	map	0.435	0.197	0.486	0.496	0.379	0.655	0.401	0.409	0.688
	recall	0.664	0.398	0.417	0.685	0.577	0.731	0.633	0.659	0.799
YOLOv6s	map	0.463	0.182	0.35	0.524	0.501	0.615	0.4	0.39	0.625
	recall	0.656	0.417	0.663	0.783	0.655	0.802	0.553	0.692	0.831
YOLOv6m	map	0.492	0.27	0.439	0.576	0.375	0.691	0.432	0.478	0.69
	recall	0.687	0.537	0.611	0.732	0.634	0.881	0.502	0.667	0.82
YOLOv6l	map	0.499	0.336	0.397	0.588	0.49	0.729	0.406	0.477	0.694
	recall	0.687	0.421	0.658	0.723	0.699	0.842	0.51	0.619	0.835
YOLOv7	map	0.337	0.143	0.377	0.376	0.214	0.597	0.259	0.218	0.516
	recall	0.402	0.244	0.582	0.448	0.36	0.596	0.432	0.227	0.324
YOLOv7x	map	0.404	0.199	0.384	0.514	0.292	0.597	0.296	0.313	0.641
	recall	0.452	0.325	0.431	0.571	0.436	0.526	0.388	0.27	0.669
YOLOv7w6	map	0.459	0.249	0.43	0.566	0.441	0.678	0.322	0.336	0.654
	recall	0.473	0.309	0.542	0.522	0.497	0.629	0.402	0.297	0.588
YOLOv7e6	map	0.501	0.234	0.452	0.584	0.531	0.712	0.431	0.419	0.648
	recall	0.535	0.301	0.588	0.609	0.589	0.676	0.537	0.372	0.608
YOLOv7e6e	map	0.507	0.259	0.448	0.579	0.509	0.731	0.4	0.452	0.675
	recall	0.546	0.316	0.536	0.613	0.608	0.737	0.512	0.366	0.678
YOLOv7d6	map	0.463	0.227	0.469	0.528	0.474	0.672	0.362	0.328	0.64
	recall	0.469	0.301	0.536	0.491	0.52	0.631	0.427	0.237	0.612
YOLOv8n	map	0.493	0.223	0.423	0.568	0.495	0.68	0.389	0.462	0.707
	recall	0.529	0.28	0.584	0.586	0.51	0.681	0.416	0.517	0.657
YOLOv8s	map	0.472	0.217	0.386	0.555	0.46	0.71	0.316	0.43	0.701
	recall	0.474	0.236	0.484	0.526	0.463	0.648	0.337	0.465	0.637
YOLOv8m	map	0.498	0.301	0.424	0.582	0.492	0.666	0.372	0.441	0.709
	recall	0.482	0.285	0.484	0.516	0.445	0.606	0.391	0.458	0.669
YOLOv8l	map	0.513	0.263	0.447	0.603	0.509	0.707	0.399	0.453	0.721
	recall	0.521	0.293	0.497	0.609	0.541	0.706	0.411	0.442	0.669
YOLOv8x	map	0.523	0.246	0.496	0.601	0.521	0.71	0.401	0.475	0.731
	recall	0.521	0.333	0.529	0.59	0.522	0.685	0.404	0.424	0.681

improvement in their detection results. This indicates that smaller-resolution images of leather surface defects are more helpful for the model's training process. Notably, the recall metric of YOLOv6l was the best among all evaluated YOLO series models, indicating its ability to better identify real objects and minimize false negatives. This success signifies that the model's improvements for industrial applications were highly effective.

In item of the detection speed (frames per second, or fps), all models were further evaluated. Larger models generally provide a certain degree of improvement in detection accuracy but at the cost of increased training complexity, which can lead to a decrease in detection speed. Among them, the YOLOv8n model achieves the fastest detection speed at 101fps. The evaluation results indicate that the YOLO series models are competent for real-time detection tasks in the

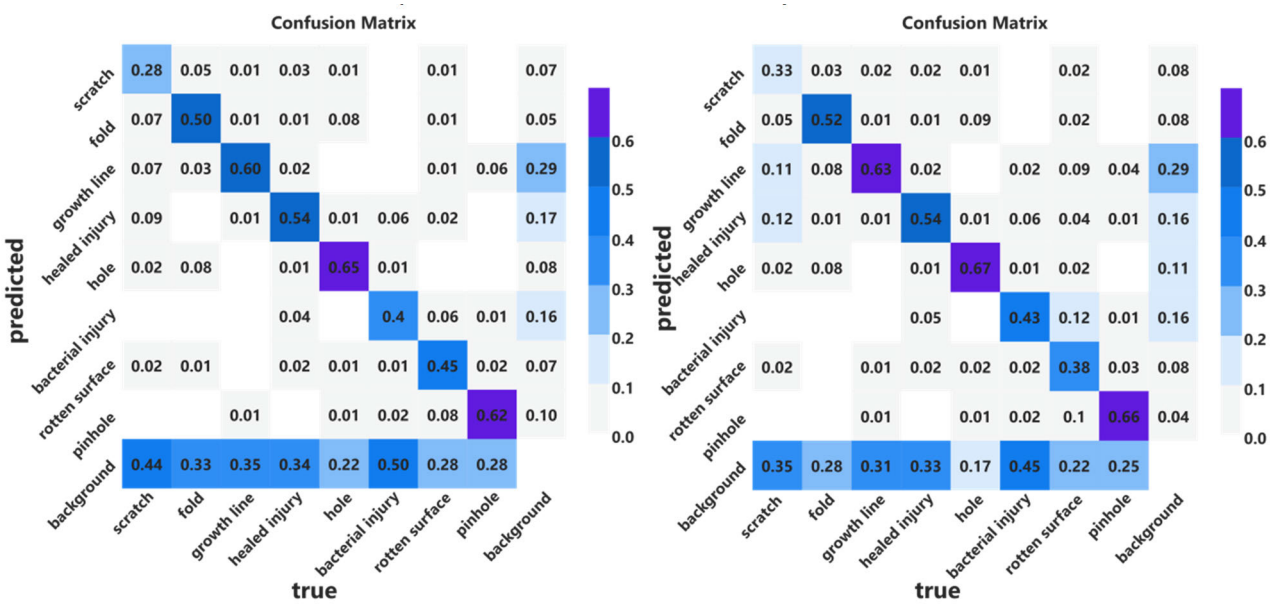


FIGURE 15. Confusion matrix for YOLOv5l and YOLOv8x models on Dataset II.

leather inspection domain, providing valuable assistance to human inspection processes.

To sum up, compared with Scheme 1, Scheme 2 is significantly improved. However, the overall performance of the model series did not turn out to be as outstanding as expected.

C. PERFORMANCE EVALUATION FOR EXPERIMENTAL SCHEME III

Scheme I and II both aim to simultaneously detect eight types of leather surface defects. Numerous experiments have shown that the YOLO series models are effective in detecting multiple leather surface defects simultaneously, but there is still considerable room for performance improvement, as the challenges are quite high. In industrial applications, if reliable intelligent detection can be achieved for a specific type of defect, it can greatly reduce costs for the industry.

Therefore, Scheme III was designed specifically for single defect detection on leather surfaces. When detecting a particular type of defect, all other defects are considered as background, allowing us to assess the YOLO model's ability for single defect detection.

Due to the large number of models, in this round of evaluation, the best-performing model (YOLOv5x, YOLOv6l, YOLOv7x, and YOLOv8x) of each series of YOLOv5-v8 in terms of overall performance is selected to showcase the results. Figure 16 presents a comparison of the YOLOv5x, YOLOv6l, YOLOv7x, and YOLOv8x of this round of experiments. Table 7 provides a detailed display of the experimental results of the best-performing models in each series. The experiments show that among the leather surface defect categories, ''scratch,'' ''hole,'' ''Healing injury,'' ''growth line,'' and ''pinhole'' are the categories with higher detection scores, achieving scores of over 65% for all three metrics. Particularly, the YOLOv8x model achieves a remarkable detection accuracy of 90.9% and an mAP of 85% on the ''pinhole'' defect dataset. The YOLOv7x model demonstrates excellent performance on the ''hole'' defect dataset, achieving a detection accuracy of 89.8% and an mAP of 85.1%. ''scratch'' are challenging defect type on both datasets, but the YOLOv7x model achieves a single defect detection accuracy of 78.9% and an mAP of 60.7%. On the other hand, the ''crease'' and ''bacterial injury'' defects pose significant challenges in this dataset, with most models achieving mAP50 and recall scores in the range of 40% to 60%. Notably, the YOLOv8x model only achieved a recall rate of 27.7% and an mAP of 29.7% on the ''Crease'' defect detection.

D. DISCUSSION

This work aims to explore the optimal scheme of leather surface defect detection based on the YOLO model, and then to lay the foundation for in-depth research and engineering applications. Therefore, three schemes were designed from different angles. Three rounds of evaluation on datasets I to III were performed for not only the detection performance of each model on individual defect types but also on multi-class defect detection.

Figure 17 shows the performance of different models in the three schemes in the form of a bar chart (with mAP as the indicator). YOLOv5 and YOLOv8 did not perform well enough in Scheme 1, the gap between mAP of different defect types was large, and the total mAP was too low. As can be seen from Figure 17, the detection performance of the model in Scheme 2 is more stable for different defects. The defect types with good performance in Scheme 1 still have

TABLE 7. Detailed experimental results based on scheme III.

Model	Metrics	scratch	fold	growth line	healed injury	hole	bacterial injury	rotten surface	pinhole
YOLOv5x	map	0.512	0.585	0.727	0.652	0.833	0.468	0.77	0.828
	recall	0.552	0.655	0.706	0.65	0.758	0.459	0.705	0.745
YOLOv6l	map	0.358	0.479	0.735	0.632	0.807	0.452	0.752	0.665
	recall	0.529	0.535	0.71	0.6	0.655	0.515	0.686	0.818
YOLOv7x	map	0.607	0.535	0.622	0.676	0.851	0.494	0.742	0.834
	recall	0.483	0.581	0.602	0.69	0.752	0.548	0.792	0.722
YOLOv8x	map	0.504	0.279	0.745	0.669	0.743	0.49	0.759	0.85
	recall	0.513	0.277	0.619	0.591	0.682	0.514	0.679	0.704

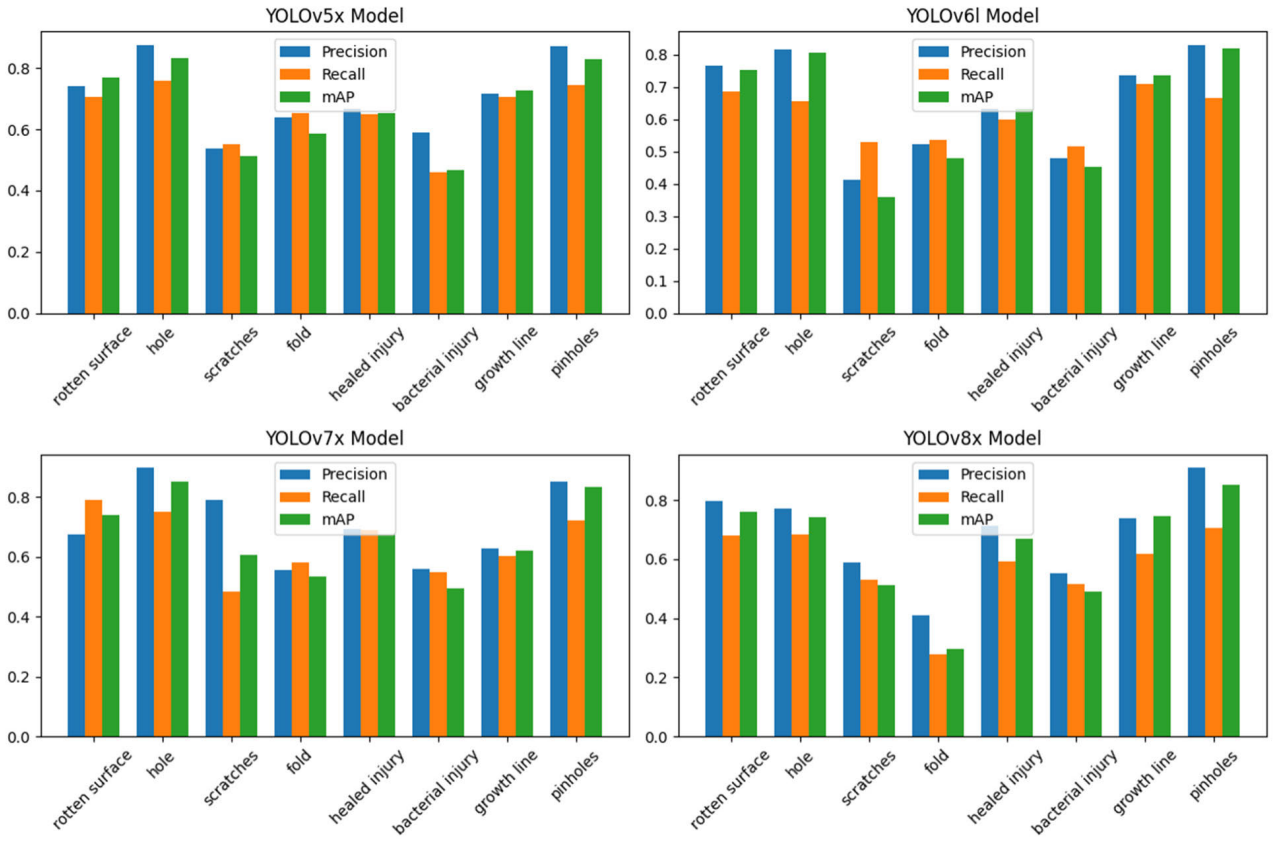


FIGURE 16. Comparison of the best performing models in each series of YOLOv5-v8 on Scheme III.

similar detection accuracy in the detection of Scheme 2, while the defect types with poor performance in Scheme 1 have a significant improvement in the detection accuracy. Scheme 3 shows a stable and excellent single defect detection performance, which has great application potential. It can be observed that, as the dataset changes, most of the defect types exhibit an increasing trend in detection performance. Particularly, there are significant improvements in the detection performance for ''pinhole'' and ''healed injury'' defects, with a substantial decrease in the missed detection rate and a significant increase in the detection accuracy. However, the detection performance for ''bacterial injury'' and ''hole'' defects shows a decline, with better performance on Dataset I compared to the other datasets.

The experimental findings signify the substantial enhancement provided by more refined datasets to the models. These models are capable of achieving an average improvement of 5% mAP on images with lower resolutions. YOLOv8 exhibits commendable overall performance.

Furthermore, from the experimental data, it can be seen that the proficiency and enormous potential of the YOLO model in single defect detection tasks are commendable. Certain

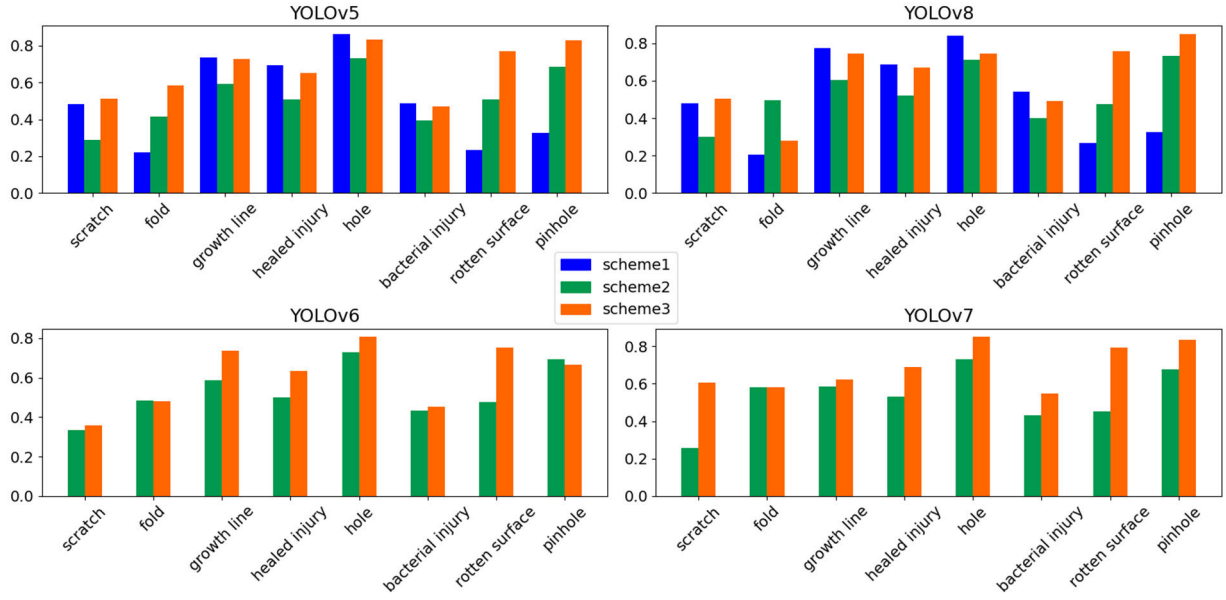


FIGURE 17. The graph compares the accuracy and miss detection rate among three different schemes.

specific confidences can attain around 90% mAP, whereas multi-defect detection tasks present significant challenges and extensive research opportunities.

Additionally, The performance of the Faster R-CNN model was also evaluated on Dataset II, which is employed to compare with the YOLO series. It only achieved an AP of 32%, which is much lower compared to the detection effectiveness of the YOLO series. Moreover, the Faster R-CNN model exhibited slower detection speed and higher training cost. This can be attributed to its weaker capabilities in abstract object detection compared to the YOLO series models.

By comprehensively comparing the above experimental results, the feasibility of employing the YOLO series models for intelligent defect detection on leather surfaces were assessed, leading to the following conclusions:

(1) The YOLO series of real-time detection models demonstrate excellent responsiveness and high accuracy in detecting leather surface defects. In single defect detection tasks, the categories of rotten, hole, healing, growth, and pinhole exhibit higher detection scores, with all three indicators achieving scores of 65% or higher. Particularly, the pinhole defect achieved a detection accuracy of 90.9% and an mAP of 85%.

(2)Defects such as wrinkles and fungal injuries present significant challenges, and their detection performance still requires improvement.

(3) The YOLO series models can simultaneously detect multiple types and multiple instances of defects in a single leather image, with the highest achieving an mAP of 52.3%. There is ample room for performance improvement and significant challenges in this regard.

(4) The YOLOv5 model is more environmentally friendly, and easier to deploy and train, but its performance is not particularly outstanding. The YOLOv6 model's improvements in industrial scenarios make it more suitable for industrial use, with a much higher recall rate compared to other models. YOLOv7 excels in single defect detection tasks, while YOLOv8 demonstrates stronger overall performance in multi-defect detection tasks.

(5) A large number of small and dense defect targets are not easy to be detected by YOLO model, which is the main challenge restricting the detection performance.

(6) In response to the above challenges, some improvements to the YOLO model have been developed. By adding lightweight attention mechanism to the neck network structure of YOLOv5, the feature extraction ability of the YOLOv5 model are enhanced, which do not generate too much additional training cost. The novel neck network structure shows a better feature fusion ability. In addition, a detection head with auxiliary positioning also have been proposed, which can improve the positioning accuracy of the detection box. The focus of this work is to evaluate the specific performance of YOLO series models on the intelligent detection task of leather surface defects. Due to the limitations of the paper layout, these improvements will be elaborated on in detail in our another paper.

VI. CONCLUSION

This work presented a systematic and in-depth experimental evaluation of the YOLO series model for the recognition and localization of surface defects on leather, which was based on three schemes designed from different angles. A thorough review of the state-of-the-art real-time object detection algorithms, particularly the YOLO series models, was conducted. Through experimental validation, the feasibility of employing the YOLO series models for intelligent defect detection on

leather surfaces was assessed. A large number of experimental evaluations showed YOLO models can significantly assist in the leather trimming process, reducing manual labor and enhancing efficiency. The multi-defect synchronous detection performance shows some positive significance, but there is a lot of room for performance improvement, requiring further improvement of the YOLO model or the development of new models. In contrast to multi-defect detection tasks, single-defect detection tasks have achieved a high detection accuracy, which appears relatively simpler and more feasible within industrial production environments. These works laid a solid foundation for the design and development of new solutions for leather defect detection. Some improvements to the existing YOLO family of models have been made so that their performance can meet the requirements of practical applications in the relevant fields. Future efforts will construct a more comprehensive dataset and evaluation system, including tasks such as generating leather surface defect images using adversarial networks and contrastive learning techniques, to further enrich the dataset.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and suggestions, which strengthened this article a lot.

REFERENCES

[1] O. Omoloso, K. Mortimer, W. R. Wise, and L. Jraisat, "Sustainability research in the leather industry: A critical review of progress and opportunities for future research," *J. Cleaner Prod.*, vol. 285, Feb. 2021, Art. no. 125441.

[2] Z. Chen, J. Deng, Q. Zhu, H. Wang, and Y. Chen, "A systematic review of machine-vision-based leather surface defect inspection," *Electronics*, vol. 11, no. 15, p. 2383, Jul. 2022, doi: 10.3390/electronics11152383.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[4] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6517–6525.

[5] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[6] A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[7] G. Jocher. (2020). *YOLOv5 by Ultralytics*. Accessed: Feb. 30, 2023. [Online]. Available: https://github.com/ultralytics/yolov5

[8] C. Li, L. Li, H. Jiang, K. Weng, Y. Geng, L. Li, Z. Ke, Q. Li, M. Cheng, W. Nie, Y. Li, B. Zhang, Y. Liang, L. Zhou, X. Xu, X. Chu, X. Wei, and X. Wei, "YOLOv6: A single-stage object detection framework for industrial applications," 2022, *arXiv:2209.02976*.

[9] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[10] G. Jocher, A. Chaurasia, and J. Qiu. (2023). *YOLO by Ultralytics*. Accessed: Feb. 30, 2023. [Online]. Available: https://github.com/ultralytics/

[11] A. F. L. Serafim, "Multiresolution pyramids for segmentation of natural images based on autoregressive models: Application to calf leather classification," in *Proc. Int. Conf. Ind. Electron., Control Instrum. (IECON)*. IEEE, 1991, pp. 1842–1847.

[12] C. Yeh and D.-B. Perng, "A reference standard of defect compensation for leather transactions," *Int. J. Adv. Manuf. Technol.*, vol. 25, nos. 11–12, pp. 1197–1204, Jun. 2005.

[13] K. Krastev, L. Georgieva, and N. Angelov, "Leather features selection for defects recognition using fuzzy logic," *Energy*, vol. 2, no. 3, pp. 1–6, 2004.

[14] A. Branca, "Automated system for detection and classification of leather defects," *Opt. Eng.*, vol. 35, no. 12, p. 3485, Dec. 1996.

[15] M. Jawahar, N. K. C. Babu, and K. Vani, "Leather texture classification using wavelet feature extraction technique," in *Proc. IEEE Int. Conf. Comput. Intell. Comput. Res.*, Dec. 2014, pp. 1–4.

[16] G. Liu, N. Cai, P. Xiao, and J. Lin, "Leather defect detection based on photometric stereo and saliency object detection," *Comput. Eng. Appl.*, vol. 55, no. 8, pp. 215–219, 2019.

[17] C. Kwak, J. A. Ventura, and K. Tofang-Sazi, "Automated defect inspection and classification of leather fabric," *Intell. Data Anal.*, vol. 5, no. 4, pp. 355–370, Nov. 2001.

[18] H. Pistori, "Defect detection in raw hide and wet blue leather," in *Computational Modelling of Objects Represented in Images. Fundamentals, Methods and Applications*. Boca Raton, FL, USA: CRC Press, 2006.

[19] E. Q. S. Filho, P. H. F. de Sousa, P. P. R. Filho, G. A. Barreto, and V. H. C. de Albuquerque, "Evaluation of goat leather quality based on computational vision techniques," *Circuits, Syst., Signal Process.*, vol. 39, no. 2, pp. 651–673, Feb. 2020.

[20] Y. S. Gan, S.-S. Chee, Y.-C. Huang, S.-T. Liong, and W.-C. Yau, "Automated leather defect inspection using statistical approach on image intensity," *J. Ambient Intell. Humanized Comput.*, vol. 12, no. 10, pp. 9269–9285, Oct. 2021.

[21] S.-T. Liong, Y. S. Gan, Y.-C. Huang, K.-H. Liu, and W.-C. Yau, "Integrated neural network and machine vision approach for leather defect classification," 2019, *arXiv:1905.11731*.

[22] M. Aslam, T. M. Khan, S. S. Naqvi, G. Holmes, and R. Naffa, "On the application of automated machine vision for leather defect inspection and grading: A survey," *IEEE Access*, vol. 7, pp. 176065–176086, 2019.

[23] S.-T. Liong, Y. S. Gan, Y.-C. Huang, C.-A. Yuan, and H.-C. Chang, "Automatic defect segmentation on leather with deep learning," 2019, *arXiv:1903.12139*.

[24] S.-T. Liong, Y. S. Gan, K.-H. Liu, T. Q. Binh, C. T. Le, C. A. Wu, C.-Y. Yang, and Y.-C. Huang, "Efficient neural network approaches for leather defect classification," 2019, *arXiv:1906.06446*.

[25] S.-T. Liong, D. Zheng, Y.-C. Huang, and Y. S. Gan, "Leather defect classification and segmentation using deep learning architecture," *Int. J. Comput. Integr. Manuf.*, vol. 33, nos. 10–11, pp. 1105–1117, Nov. 2020.

[26] Y. S. Gan, S.-T. Liong, S.-Y. Wang, and C. T. Cheng, "An improved automatic defect identification system on natural leather via generative adversarial network," *Int. J. Comput. Integr. Manuf.*, vol. 35, no. 12, pp. 1378–1394, Dec. 2022.

[27] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[28] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[29] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[32] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50× fewer parameters and <0.5 MB model size," 2016, *arXiv:1602.07360*.

[33] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6848–6856.

[34] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inform. Process. Syst. (NIPS)*, 2017, pp. 5998–6008.

[35] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[36] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 936–944.
[37] J. Terven and D. Cordova-Esparza, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," 2023, *arXiv:2304.00501*.
[38] C. Feng, Y. Zhong, Y. Gao, M. R. Scott, and W. Huang, "TOOD: Task-aligned one-stage object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 3490–3499.
[39] C.-Y. Wang, H.-Y. Mark Liao, and I.-H. Yeh, "Designing network design strategies through gradient path analysis," 2022, *arXiv:2211.04800*.
[40] Z. Chen, D. Xu, J. Deng, Y. Chen, and C. Li, "Comparative study on deep-learning-based leather surface defect identification," *Meas. Sci. Technol.*, vol. 35, no. 1, Jan. 2024, Art. no. 015402, doi: 10.1088/1361-6501/acfb9f.

ZHIQIANG CHEN received the B.S. degree from Wuhan University of Water Conservancy and Electric Power, Wuhan, China, in 2001, the M.S. degree from Chongqing University, Chongqing, China, in 2004, and the Ph.D. degree from the University of Fukui, Japan, in 2011. He is currently a Professor with the School of Electrical and Information Engineering, Quzhou University. His research interests include data mining, machine vision-based surface defect detection and prognostics, and health management.

QIRUI ZHU received the B.S. degree from Anhui University of Finance and Economics, China, in 2021. He is currently pursuing the master's degree with Hangzhou Dianzi University. His research interest includes leather defect detection.

XIAOFAN ZHOU received the B.S. degree from Nanchang Hangkong University, Jiangxi, China, in 2007, and the M.S. and Ph.D. degrees from the University of Fukui, Fukui, Japan, in 2010 and 2013, respectively. He is currently the Director of the Advanced Technology Research and Development Center, Zhejiang Qianjiang Robot Company Ltd. His research interests include electrical engineering, process automation, and industrial robotics applications.

JIEHANG DENG received the B.S. and M.S. degrees from Xi'an University of Technology, China, in 2002 and 2005, respectively, and the Ph.D. degree from the University of Fukui, Japan, in 2009. He is currently an Associate Professor with the School of Computers, Guangdong University of Technology, China. His research interests include image processing and pattern recognition.

WEI SONG was born in 1981. She received the master's and Ph.D. degrees from Fukui University, Japan, in 2006 and 2009, respectively. She is currently with the School of Mechatronic Engineering and Automation, Shanghai University. She is also an Associate Professor. Her research interests include machine vision, 3D pose measurement, and robot control.