INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Real-Time Fabric Defect Detection Using a Lightweight Deformable YOLO Network

Hou zongxiang[1],Ashardi bin Abas[2]

University Pendidikan Sultan Idris

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150300002

Received: 14 March 2026; Accepted: 19 March 2026; Published: 31 March 2026

ABSTRACT

Fabric defect detection is a critical task in textile manufacturing, where manual inspection remains inconsistent, labour-intensive, and unsuitable for high-speed production environments. Although deep learning–based detectors have shown strong potential, many existing models are too computationally demanding for practical deployment in real-time industrial inspection systems. This study proposes a lightweight deformable YOLO-based framework for accurate and efficient fabric defect detection. The model is built on YOLOv5s and enhanced through three efficiency-oriented architectural improvements: Bidirectional Feature Pyramid Network (BiFPN) for improved multi-scale feature fusion, Deformable Convolutional Networks (DCNv2) for stronger geometric adaptability, and Efficient Pyramid Split Attention (EPSA) for enhanced feature discrimination. The proposed model was trained and evaluated on the Alibaba Tianchi fabric defect dataset, comprising 5,913 images across 20 defect categories. Experimental evaluation was conducted using mean Average Precision (mAP), model size, and real-time suitability, supported by ablation and comparative analyses. Results show that the proposed method improved mAP from 41.9% for the baseline YOLOv5s to 48.2%, representing a gain of 6.3 percentage points. The findings indicate that targeted architectural optimisation can improve detection accuracy while preserving the lightweight characteristics required for industrial implementation. The proposed framework offers a practical solution for automated fabric inspection and provides a useful reference for efficiency-oriented defect detection in smart manufacturing environments.

Keywords: lightweight deep learning; real-time inspection; fabric defect detection; YOLOv5

INTRODUCTION

Fabric defect detection is a fundamental component of quality assurance in textile manufacturing because it directly influences product reliability, operational efficiency, and commercial value. Defects such as holes, broken yarns, stains, and surface texture irregularities can substantially degrade fabric quality and marketability. Previous studies have reported that undetected defects may reduce product value by as much as 45–65%, underscoring the economic importance of early and reliable inspection [1]. Despite this, inspection practices in many textile production environments still depend heavily on manual visual assessment. Such an approach is labor-intensive, subjective, and highly susceptible to fatigue, inconsistency, and human error, often resulting in defect-detection accuracy of only 60–75% [2], [3].

To overcome these limitations, automated inspection systems based on computer vision have been widely explored. Earlier methods primarily relied on texture descriptors, statistical analysis, and handcrafted feature extraction techniques.

Although these approaches offered some success under controlled settings, they generally lacked robustness when confronted with complex fabric patterns, varying illumination, and irregular defect characteristics [4], [5]. In recent years, deep learning, particularly convolutional neural networks (CNNs), has significantly advanced the field of fabric defect detection by enabling models to learn hierarchical and discriminative feature representations directly from data [6], [7]. Among these developments, one-stage object detectors such as the

[1] Corresponding Author:Ashardi bin Abas, ashardi@meta.upsi.edu.my.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

YOLO family have gained considerable attention for their end-to-end detection capabilities and strong potential for real-time industrial inspection.

However, the increasing trend toward deeper and more computationally intensive detection architectures has introduced new deployment challenges. While many advanced models achieve strong accuracy in laboratory benchmarks, their practical use in real textile production lines remains limited by inference latency, memory demands, and hardware constraints [6], [8]. In industrial inspection scenarios, a detector must not only identify defects accurately, but also operate efficiently under continuous, high-speed production conditions. This creates a critical need for lightweight architectures that can maintain robust detection performance while satisfying real-time operational requirements.

Against this background, this study proposes a lightweight deformable YOLO-based framework for efficient fabric defect detection. The proposed model is designed to improve the balance between detection accuracy and deployability through targeted architectural enhancement rather than network expansion. Specifically, the framework integrates Bidirectional Feature Pyramid Network (BiFPN) to strengthen multi-scale feature fusion, Deformable Convolutional Networks (DCNv2) to improve adaptability to irregular defect geometry, and Efficient Pyramid Split Attention (EPSA) to enhance feature discrimination with limited computational overhead. The proposed model was evaluated on the Alibaba Tianchi fabric defect dataset, which contains 5,913 annotated images. After category consolidation, 20 defect categories were used for training and testing. Experimental evaluation focused on mean Average Precision (mAP), model size, and real-time suitability through ablation and comparative analysis.

Problem Statement and Research Gap

Although deep learning–based fabric defect detection has progressed substantially, two major challenges remain unresolved. First, many high-performing models are derived from general-purpose object detection architectures that were not originally designed for industrial efficiency. These networks are often deep, parameter-heavy, and computationally demanding, making them difficult to deploy on real-time inspection systems, embedded platforms, or resource-constrained manufacturing environments [6], [9]. As a result, there is a persistent gap between algorithmic performance reported in research settings and the practical requirements of industrial implementation.

Second, efforts to reduce computational complexity through lightweight convolutional operations, pruning strategies, or model compression frequently lead to a decline in detection accuracy, especially when defects are small, irregular, or embedded within repetitive and low-contrast fabric textures [7], [10]. This reveals a longstanding trade-off between efficiency and robustness in fabric defect detection. Moreover, many existing studies examine isolated improvements without offering a systematic assessment of how multiple architectural components can be jointly optimized to preserve detection quality while improving real-time efficiency.

Therefore, a clear research gap exists in developing a fabric defect detection framework explicitly designed from the outset for lightweight operation and real-time deployment, rather than being simplified from a heavy baseline after development. Addressing this gap is essential for enabling practical, scalable, and industry-ready automated inspection systems in textile manufacturing.

Study Rationale and Significance

This study is motivated by the need for a practical, efficiency-oriented fabric defect detection solution that meets industrial deployment constraints without compromising detection performance. Rather than pursuing higher accuracy through greater model depth and complexity, this research investigates how architectural optimization can achieve a more effective balance between computational efficiency and detection robustness. The rationale for the study is to demonstrate that lightweight design does not necessarily entail a substantial sacrifice in detection capability when model components are carefully selected and integrated.

From an academic perspective, this work contributes to the broader field of industrial computer vision by providing empirical evidence on the efficiency–accuracy trade-off in deep learning–based inspection systems.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

In particular, it highlights the significance of multi-scale bidirectional feature fusion, adaptive convolution, and computationally economical attention mechanisms in improving lightweight detector performance. From a practical perspective, the proposed framework offers textile manufacturers a more deployable solution for automated real-time defect inspection, thereby supporting smart manufacturing, digital quality assurance, and Industry 4.0 implementation. The findings also provide useful design guidance for researchers and engineers developing AI-based visual inspection systems under hardware and latency constraints.

Research Objectives and Research Questions

The primary objective of this study is to design and evaluate a lightweight deep learning architecture for real-time fabric defect detection suitable for industrial deployment. Specifically, the study aims to:

i. Develop a computationally efficient fabric defect detection framework based on a lightweight YOLO architecture.
ii. Analyze the impact of architectural optimization on the balance between detection accuracy and real-time performance.
iii. Evaluate the proposed model against existing detectors in terms of efficiency and detection robustness.

Accordingly, this study addresses the following research questions:

i. How can a lightweight detection architecture be designed to satisfy real-time inspection requirements in textile manufacturing?
ii. What is the impact of feature fusion, adaptive convolution, and attention mechanisms on detection efficiency and robustness?
iii. How does the proposed lightweight model compare with existing fabric defect detection approaches in terms of real-time suitability and detection performance?

Structure of the Paper

The remainder of this paper is organized as follows. Section 2 reviews related studies on fabric defect detection, with particular emphasis on lightweight and real-time deep learning approaches. Section 3 presents the proposed network architecture and explains the design rationale of each optimization component. Section 4 describes the experimental setup, dataset configuration, evaluation metrics, and comparative methodology. Section 5 presents and discusses the experimental results, including ablation findings and industrial implications. Finally, Section 6 concludes the paper by summarizing the main contributions and outlining directions for future research.

LITERATURE REVIEW

The purpose of this literature review is to critically examine existing research on fabric defect detection with a specific focus on lightweight architectures and real-time efficiency. While significant progress has been made in improving detection accuracy through deep learning, the practical deployment of these models in industrial environments remains constrained by computational cost, inference latency, and hardware limitations. As textile production lines demand continuous, high-speed inspection, real-time feasibility has become as critical as detection accuracy.

Accordingly, this review is structured around four key themes: (i) traditional fabric defect detection approaches and their efficiency limitations, (ii) deep learning–based detection models and their computational challenges, (iii) lightweight and real-time object detection architectures, and (iv) architectural optimization strategies for balancing accuracy and efficiency. This thematic analysis highlights unresolved challenges and establishes the motivation for the present study.

THEORETICAL FRAMEWORK

Fabric defect detection is theoretically grounded in computer vision, pattern recognition, and hierarchical feature learning. Early automated inspection systems were mainly based on texture analysis theory, which

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

models fabric surfaces as repetitive, structured patterns, with defects interpreted as structural or statistical deviations from normal texture distributions. Within this framework, methods such as gray-level co-occurrence matrices, Fourier transforms, and wavelet analysis were widely used to characterize local and global texture abnormalities [1], [2]. These methods were computationally efficient and relatively interpretable, but they depended heavily on handcrafted features and assumed stable texture regularity, which limited their robustness under varying industrial conditions.

The emergence of deep learning transformed this theoretical foundation by replacing manually engineered descriptors with data-driven hierarchical feature learning. Convolutional neural networks (CNNs) learn low-level features such as edges, contours, and textures in early layers, while deeper layers encode more abstract semantic representations [3]. This hierarchical structure enables improved adaptability to complex patterns and heterogeneous defect characteristics, making CNNs highly effective for visual inspection tasks. However, most standard CNN-based detection frameworks were originally designed for general object detection benchmarks rather than computationally constrained industrial systems [4]. As a result, high representational power is often achieved at the cost of increased parameter count, memory usage, and inference time.

In the context of the present study, the theoretical framework extends hierarchical feature learning toward efficiency-oriented architectural design. Specifically, the study assumes that robust defect detection can be achieved not merely through deeper networks, but through more effective architectural organization. Multi-scale feature fusion supports the detection of defects of varying sizes, adaptive convolution improves sensitivity to irregular defect geometries, and lightweight attention mechanisms enhance discrimination of defect-relevant features under computational constraints. Thus, this study is theoretically grounded in adapting hierarchical feature learning to real-time industrial inspection requirements.

Review of Key Themes

Traditional and Machine Learning–Based Fabric Defect Detection

Traditional fabric defect detection approaches primarily relied on handcrafted feature extraction and rule-based classification. These methods typically used statistical texture descriptors, structural regularity measures, or frequency-domain analysis to distinguish normal fabric surfaces from defective regions [1], [5]. Under controlled imaging conditions, such methods demonstrated reasonable performance with relatively low computational cost. Their efficiency made them attractive for early automated inspection systems, especially when hardware resources were limited.

Subsequent studies introduced machine learning techniques such as support vector machines and shallow neural networks to improve classification adaptability [6]. Compared with purely rule-based systems, these methods provided greater flexibility in separating normal and defective samples. However, their performance still depended heavily on the quality of handcrafted features, which limited their generalization capability. In practice, these approaches often struggled with complex textures, non-uniform illumination, subtle defect boundaries, and large inter-class variation among defect types. As a result, while traditional and machine-learning–based methods remain computationally economical, they are generally insufficient for robust, scalable deployment in modern textile production environments.

Deep Learning–Based Fabric Defect Detection

Deep learning has substantially advanced fabric defect detection by enabling models to learn discriminative feature representations directly from image data. CNN-based methods have been applied in three main forms: classification, segmentation, and object detection. Classification models are effective for determining whether a fabric sample is defective, but they do not provide explicit defect localization. Segmentation models offer fine-grained pixel-level delineation of defective regions, which is beneficial for detailed inspection, but they are often computationally expensive and less suitable for real-time deployment. Object detection models offer a practical compromise, enabling defect localization with lower computational overhead than dense segmentation, making them particularly relevant for industrial inspection tasks [4], [7].

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Among deep learning approaches, one-stage object detectors, such as YOLO-based architectures, have received considerable attention for their ability to combine end-to-end learning with fast inference. Their unified detection pipeline makes them well-suited for applications requiring real-time performance. Nevertheless, many deep learning models continue to prioritize accuracy over computational efficiency, often through deeper backbones, more complex feature aggregation modules, and heavier parameter configurations. Although such architectures achieve promising results under research conditions, their deployment on textile production lines is constrained by latency, memory requirements, and hardware limitations [8]. This raises an important concern: strong benchmark performance does not necessarily translate into practical real-time usability in industrial environments.

Lightweight and Real-Time Detection Architectures

To address deployment constraints, recent research has increasingly focused on lightweight and real-time detection architectures. Common strategies include depthwise separable convolutions, model pruning, quantization, and simplified backbone designs [9]. These methods aim to reduce parameter count, memory usage, and computational complexity, thereby improving inference speed and enabling operation on resource-constrained platforms.

Despite these advantages, lightweight architectures often introduce new challenges. In many cases, computational reduction is achieved at the expense of representational capacity, leading to weaker performance in detecting small, low-contrast, or irregular defects embedded in repetitive fabric textures [10]. This is particularly problematic in textile inspection, where defect patterns may be subtle and highly variable. Moreover, many studies assess efficiency gains in isolation, emphasizing model compression or inference speed without systematically analyzing the associated trade-offs in detection accuracy and robustness. Consequently, the literature still provides limited guidance on implementing lightweight design without undermining the practical reliability required for industrial inspection.

Architectural Optimization for Efficiency–Accuracy Balance

Recent studies suggest that the balance between efficiency and accuracy can be improved through architectural optimization rather than simply through model scaling. Three strategies are especially relevant in this context: multi-scale feature fusion, adaptive convolution, and attention mechanisms. Multi-scale feature fusion enhances the integration of shallow and deep features, thereby improving the detection of defects with diverse sizes and visual characteristics [4], [7]. Adaptive convolutional mechanisms allow the receptive field to adjust according to irregular shapes and geometric variations, making them especially valuable for detecting non-uniform fabric defects. Attention mechanisms further improve feature refinement by emphasizing informative regions and suppressing redundant or background information.Although these strategies have been widely studied in general object detection, they are often introduced primarily to maximize accuracy rather than to enable lightweight, real-time deployment. Their combined contribution to efficiency-aware design in fabric defect detection remains insufficiently explored. In particular, existing studies rarely examine how these architectural components can be integrated systematically to preserve detection robustness while controlling computational overhead. This limitation is especially relevant in textile inspection, where models must operate under industrial constraints rather than idealized research settings.

Table 1. Comparative summary of prior fabric defect detection approaches

Approach Category	Representative Methods	Strengths	Limitations	Real-Time Suitability
Traditional texture-based methods	GLCM, Fourier transform, wavelet analysis	Low computational cost; interpretable	Sensitive to texture variation and illumination; weak generalization	Moderate under controlled settings

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Machine learning–based methods	SVM, shallow neural networks	Better adaptability than rule-based methods	Still depends on handcrafted features; limited robustness	Moderate
Deep learning classification/segmentation methods	CNN classifiers, segmentation networks	Strong feature learning; high accuracy	Often computationally expensive; segmentation may be slow	Low to moderate
One-stage object detectors	YOLO-based detectors	Fast end-to-end detection; practical localization	May struggle with fine or irregular defects without refinement	High
Lightweight real-time detectors	Pruned networks, depthwise separable CNNs	Reduced parameters and faster inference	Often sacrifices accuracy and robustness	High, with trade-off
Architecture-optimized lightweight detectors	Multi-scale fusion, adaptive convolution, attention-based designs	Better balance of efficiency and accuracy	Still underexplored in textile defect detection	High potential

Critical Gaps in Existing Studies

The literature reveals several important gaps. First, traditional and machine-learning–based approaches are computationally efficient but lack the robustness and scalability required for modern industrial inspection, particularly in the presence of complex textures and variable production conditions. Second, deep learning–based methods have achieved substantial improvements in detection accuracy, yet many of these models remain too computationally demanding for real-time deployment on practical manufacturing platforms. Third, although lightweight architectures have been proposed to improve efficiency, they frequently suffer from reduced detection performance, especially for subtle, small-scale, or irregular defects. Most importantly, there is still limited research on efficiency-oriented architectural design specifically tailored to fabric defect detection. Existing studies rarely provide a systematic analysis of how feature fusion, adaptive convolution, and attention mechanisms can be jointly optimized to balance computational efficiency and detection robustness. In other words, current research tends either to emphasize high accuracy using heavy architectures or to pursue lightweight simplification without sufficiently preserving detection quality. This unresolved gap provides the central motivation for the present study.

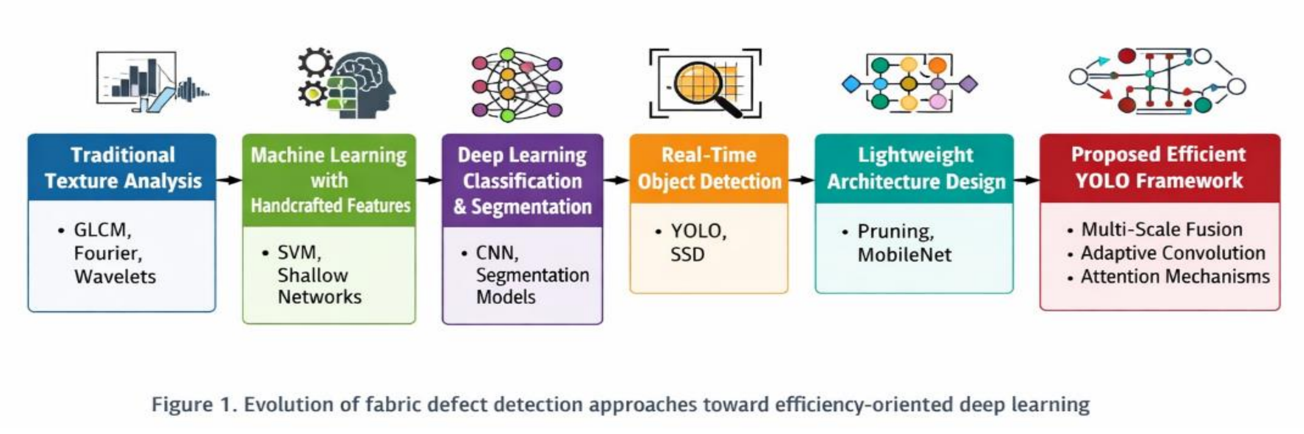


Figure 1. Evolution of fabric defect detection approaches toward efficiency-oriented deep learning

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

SUMMARY AND TRANSITION TO THE METHODOLOGY

In summary, prior research has demonstrated the strong potential of deep learning for fabric defect detection, but real-time efficiency and deployability remain insufficiently addressed. Traditional and machine learning–based methods offer computational simplicity but lack robustness, while many high-performing deep learning models are too resource-intensive for practical industrial deployment. Although lightweight architectures and architectural optimization strategies have emerged as promising directions, their integrated application in fabric defect detection remains underdeveloped.

These limitations justify the present study, which focuses on designing a lightweight deep learning architecture explicitly tailored for real-time fabric defect detection under industrial constraints. Building on the insights derived from this review, the next section presents the proposed methodology, including the design rationale of the detection framework, its architectural components, and the experimental setup used to evaluate both efficiency and detection performance.

Methodology

This study adopts a quantitative experimental design to evaluate the effectiveness of a lightweight deep learning architecture for real-time fabric defect detection. The methodology is based on controlled computational benchmarking, where the proposed detector is trained and tested under standardized experimental conditions and then compared with baseline and reference models using objective performance metrics. This design is appropriate because the study aims to measure observable outcomes, including detection accuracy, model compactness, and real-time suitability, rather than subjective or interpretive phenomena.

The methodological framework consists of four main stages: dataset preparation, lightweight model construction, controlled training and testing, and performance evaluation through ablation and comparative analysis. First, a publicly available annotated fabric defect dataset was prepared and reorganized for object detection experiments. Second, the baseline YOLOv5s detector was enhanced with efficiency-oriented architectural components, namely Bidirectional Feature Pyramid Network (BiFPN), Deformable Convolutional Networks (DCNv2), and Efficient Pyramid Split Attention (EPSA). Third, all models were trained and evaluated using a fixed experimental configuration to ensure fair comparison. Finally, the resulting models were assessed using standard object-detection metrics and benchmarked against representative existing detectors. This structure ensures that the contribution of each architectural modification can be isolated and interpreted systematically.

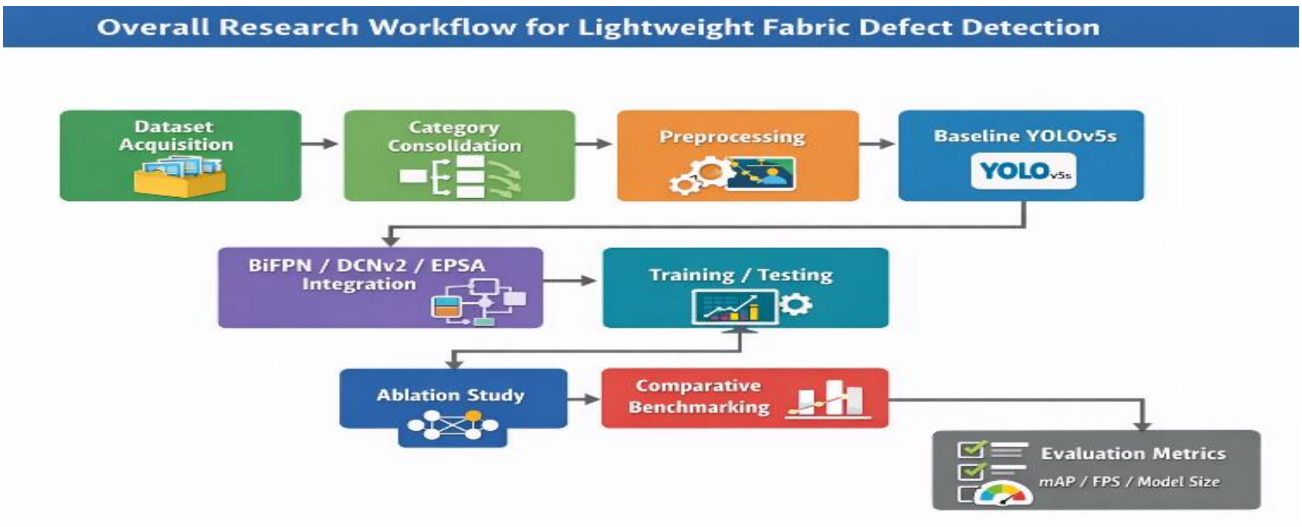


Figure 2: Overall research workflow for lightweight fabric defect detection", showing: dataset acquisition → category consolidation → preprocessing → baseline YOLOv5s → BiFPN/DCNv2/EPSA integration → training/testing → ablation study → comparative benchmarking → evaluation metrics.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Dataset Acquisition and Preparation

The experimental dataset used in this study was obtained from the Alibaba Tianchi Fabric Defect Detection Challenge, a publicly available benchmark dataset for automated textile inspection. The dataset contains 5,913 annotated fabric images with diverse defect patterns captured under practical inspection conditions. The original annotation set comprised 34 defect categories, which were subsequently consolidated into 20 defect classes to reduce inter-class fragmentation and improve the stability of model learning. This category consolidation step was important because some original labels represented visually similar or low-frequency defect types that could weaken convergence and reduce comparative interpretability during training.

Each image was annotated using rectangular bounding boxes to support supervised object detection. The dataset includes representative fabric defects such as holes, broken yarns, stains, and texture irregularities, thereby reflecting a range of defect scales, shapes, and surface characteristics relevant to real textile inspection. After preprocessing and label consolidation, the dataset was divided into two subsets: 4,730 images (80%) for training and 1,183 images (20%) for testing. This split was used to maintain consistency in performance evaluation and to ensure that the final reported results reflected unseen data.

Prior to training, the images underwent standard preprocessing procedures, including resizing and normalization. Data handling was designed to preserve defect visibility while ensuring compatibility with the detector input pipeline. Where appropriate, image augmentation strategies were applied to improve robustness and reduce overfitting by exposing the model to variations in appearance and spatial distribution. These preprocessing steps were implemented consistently across all experimental configurations to ensure that performance differences could be attributed to architectural design rather than to inconsistent input preparation.

Table 2. Dataset configuration

Item	Description
Dataset source	Alibaba Tianchi Fabric Defect Detection Challenge
Total labeled images	5,913
Original defect categories	34
Final merged categories	20
Annotation format	Bounding boxes
Training set	4,730 images (80%)
Testing set	1,183 images (20%)

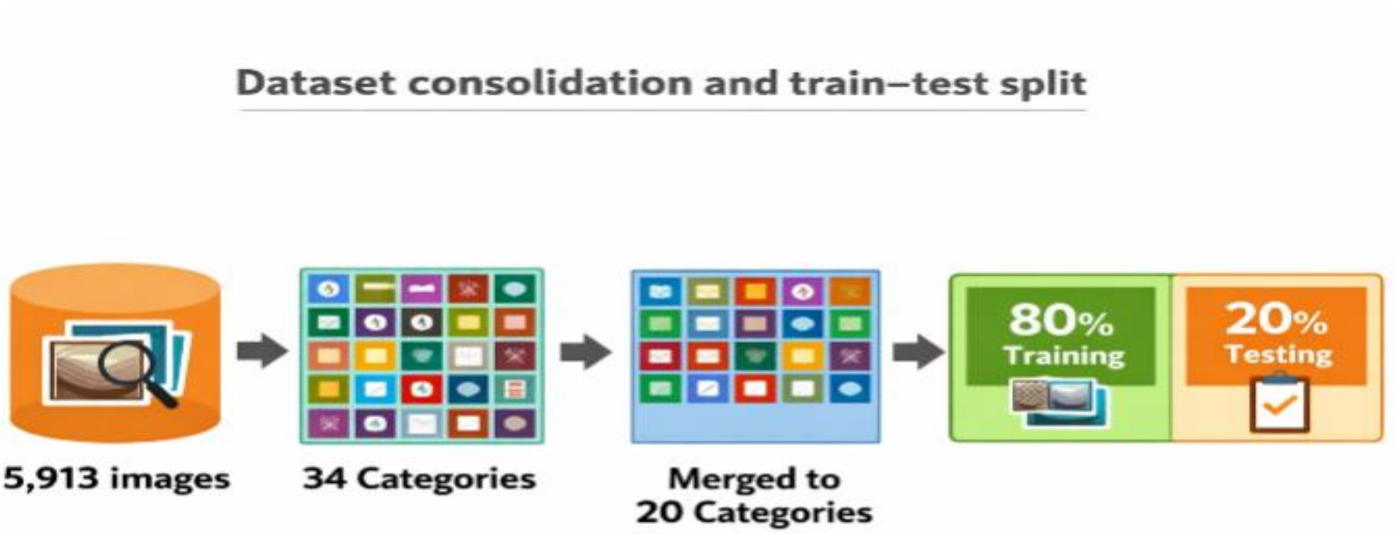


Figure 3: Dataset consolidation and train–test split” showing 5,913 images → 34 categories → merged to 20 categories → 80% training / 20% testing.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Proposed Model Configuration

The proposed detection framework was developed on YOLOv5s, selected as the baseline architecture for its favorable balance between inference speed and detection capability. Compared with larger detectors, YOLOv5s provides a more suitable starting point for lightweight optimization because it offers relatively low model complexity while preserving strong end-to-end object detection performance. However, despite its efficiency, the baseline architecture remains limited in representing multi-scale defect patterns, adapting to irregular defect geometry, and selectively emphasizing discriminative defect features.

To address these limitations, three architectural enhancements were introduced. First, Bidirectional Feature Pyramid Network (BiFPN) was incorporated to improve multi-scale feature fusion. Unlike conventional feature pyramids, BiFPN enables richer bidirectional information flow between shallow and deep layers, thereby improving defect representation across different sizes and visual resolutions. Second, Deformable Convolutional Networks (DCNv2) were integrated to improve sensitivity to irregular and non-uniform defect structures. By allowing convolutional sampling locations to adapt spatially, DCNv2 provides better geometric flexibility than standard convolution, which is especially important for fabric defects with distorted boundaries or non-rigid appearance. Third, Efficient Pyramid Split Attention (EPSA) was introduced to enhance feature discrimination while preserving lightweight computation. This module improves the network's ability to emphasize relevant defect regions and suppress redundant background texture without imposing excessive computational overhead.

The overall design rationale of the proposed model was therefore not to increase depth or parameter count indiscriminately, but to improve efficiency through selective architectural refinement. By combining BiFPN, DCNv2, and EPSA within a lightweight YOLO framework, the study seeks to improve the balance between detection robustness and real-time suitability under industrial inspection constraints.

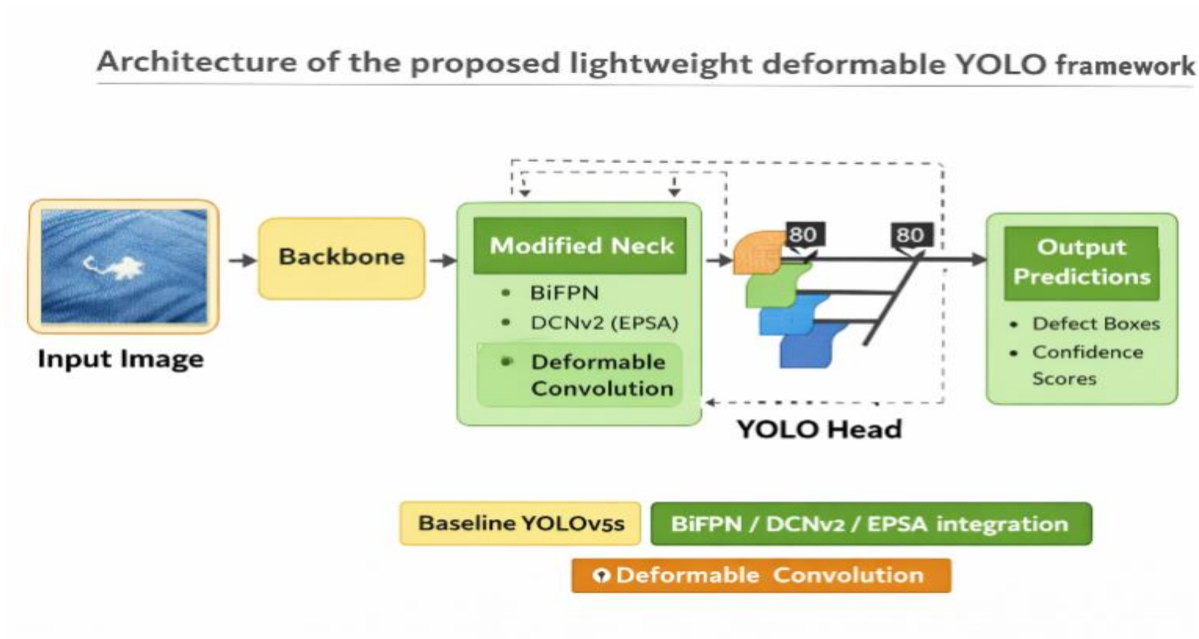


Figure 4: Architecture of the proposed lightweight deformable YOLO framework." if you want a methodology figure focused on the model rather than the full workflow

Experimental Setup and Evaluation Metrics

All experiments were conducted under a standardized computational environment to ensure comparability across models. The baseline YOLOv5s and all modified variants were trained and evaluated using the same dataset split, preprocessing logic, and evaluation procedures. The experimental design included both ablation analysis and comparative benchmarking. In the ablation phase, the effect of individual architectural components was assessed by incrementally modifying the baseline detector. In the comparative phase, the final proposed

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

model was benchmarked against representative reference detectors to evaluate its relative performance in terms of both effectiveness and efficiency.

The evaluation focused on widely accepted object-detection metrics. Mean Average Precision (mAP) was used as the primary indicator of detection accuracy because it summarizes localization and classification performance across object categories. Precision and recall were also used to provide additional interpretation of detection reliability and sensitivity. To assess real-time suitability, frames per second (FPS) was used as the primary metric for inference speed. In addition, model size was considered as an efficiency-related measure, given its importance for deployment in practical industrial systems with memory and hardware constraints. Together, these metrics enabled the study to systematically analyze the trade-off between computational efficiency and detection performance.

The ablation study was structured around the following configurations: baseline YOLOv5s, YOLOv5s with BiFPN, YOLOv5s with optimized convolutional replacement such as DCNv2, YOLOv5s with attention-based enhancement, and the final integrated model combining the proposed efficiency-oriented modules. For broader contextual evaluation, the final model was compared against selected existing detectors, including Faster R-CNN, SSD, YOLOv3, YOLOv5s, and YOLOv8. This comparative structure enabled the study to determine whether architectural optimization could improve performance beyond both classical and contemporary detector baselines.

All implementation, model training, evaluation, and visualization procedures were carried out using Python-based deep learning tools to support experimental reproducibility. Performance plots, result tables, and comparative analyses were generated under the same experimental pipeline to minimize procedural inconsistency.

Table 3. Experimental configuration and evaluation design

Item	Configuration
Baseline detector	YOLOv5s
Proposed optimization modules	BiFPN, DCNv2, EPSA
Analysis type	Ablation study and comparative benchmarking
Evaluation metrics	mAP, precision, recall, FPS, model size
Comparative models	Faster R-CNN, SSD, YOLOv3, YOLOv5s, YOLOv8
Implementation environment	Python-based deep learning framework

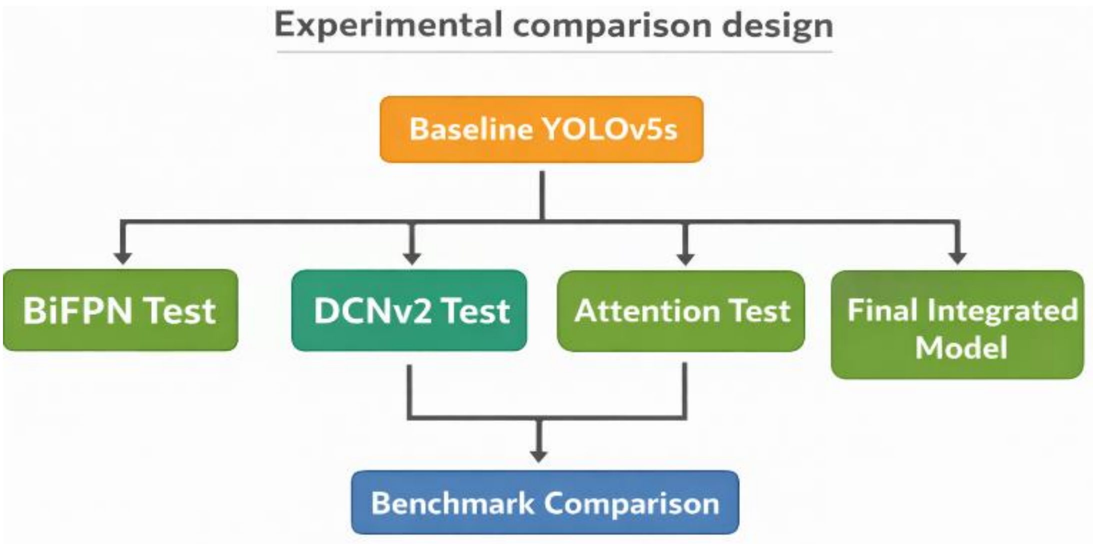


Figure 5: Experimental comparison design", showing branches from baseline YOLOv5s to BiFPN test, DCNv2 test, attention test, final integrated model, then benchmark comparison.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Reliability, Validity, and Ethical Considerations

Reliability was addressed by conducting all experiments under standardized conditions, including fixed dataset partitions, consistent preprocessing procedures, common evaluation metrics, and controlled model comparison settings. The use of repeated testing under identical configurations improved the stability of performance observations, while the ablation study design strengthened internal reliability by isolating the contribution of individual architectural components. This made it possible to interpret performance changes as a consequence of model design rather than random experimental variation.

Validity was addressed at several levels. Construct validity was supported through the use of established object-detection metrics, including mAP, precision, recall, FPS, and model size, all of which are directly relevant to the study objectives.

Internal validity was reinforced through controlled benchmarking against the baseline and comparative models. External validity was supported by the use of realistic fabric defect images drawn from a public industrial dataset, although broader generalization remains dependent on future validation across additional textile environments and hardware platforms.

This study did not involve human participants, personal data, or sensitive information. All experimental data consisted of fabric images used for technical inspection research. Therefore, formal human-subject ethics approval was not required. Nevertheless, ethical research practice was maintained through transparent reporting, fair comparative analysis, and appropriate acknowledgment of prior work and benchmark methods.

Limitations of the Methodology

Several methodological limitations should be acknowledged. First, the dataset represents specific fabric categories and inspection conditions, which may limit the generalizability of the findings to other textile contexts with different materials, lighting conditions, or defect distributions.

Second, although the study evaluates computational efficiency through model size and real-time inference measures, the experiments were primarily conducted in a controlled high-performance computing environment rather than on low-power embedded or edge devices. As such, the reported efficiency results should be interpreted as evidence of relative lightweight suitability rather than full hardware-level deployment validation. Future research should therefore extend this methodology by incorporating broader textile datasets and conducting direct edge-device implementation studies

Summary and Transition

This methodology describes the quantitative experimental framework used to evaluate a lightweight deep learning approach for fabric defect detection. The section has outlined the dataset source and preparation process, the design rationale of the proposed YOLO-based architecture, the evaluation metrics and comparison strategy, and the procedures used to support reliability and validity. By structuring the study around controlled benchmarking and ablation analysis, the methodology provides a transparent basis for assessing both detection accuracy and real-time efficiency. The next section presents the experimental results and discusses the proposed model's performance relative to the baseline and comparative detectors.

RESULT

This section presents the experimental findings obtained from evaluating the proposed lightweight deformable YOLO framework for fabric defect detection. In accordance with the research objectives, the results are organized to examine: (i) the effect of individual architectural modifications on detection accuracy, (ii) the cumulative contribution of efficiency-oriented enhancements through ablation analysis, and (iii) the comparative performance of the final model against representative object detectors. The evaluation focuses primarily on mean Average Precision (mAP), with additional interpretation regarding model size and real-time suitability for industrial inspection.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Effect of Feature Pyramid Structures on Multi-Scale Detection Performance

The first experiment examined the influence of different feature pyramid structures on fabric defect detection performance. Because fabric defects vary considerably in size, aspect ratio, and visual salience, effective multi-scale feature fusion is essential for accurate localization. The comparison shows that the choice of feature pyramid structure directly impacts mAP. Among the evaluated configurations, BiFPN achieved the highest performance, outperforming both the conventional FPN and the default PAFPN-style neck used in the baseline setting. The improvement indicates that weighted bidirectional fusion enhances the interaction between shallow spatial features and deeper semantic information, thereby improving the detector's ability to capture subtle and scale-varying defects

Table 4. Comparison of feature pyramid structures

Model	Model Size (MB)	mAP (%)
YOLOv5s + FPN	14.0	39.3
YOLOv5s + PAFPN	14.6	41.9
YOLOv5s + BiFPN	14.7	42.7

The results show that BiFPN improved mAP by **0.8 percentage points** over the baseline PAFPN configuration. Although the increase appears modest, it is significant in the context of defect detection, where small gains often reflect more reliable localization of difficult and low-contrast defect regions. This finding supports the argument that multi-scale fusion is a key mechanism for improving lightweight fabric defect detectors without substantially increasing architectural complexity.

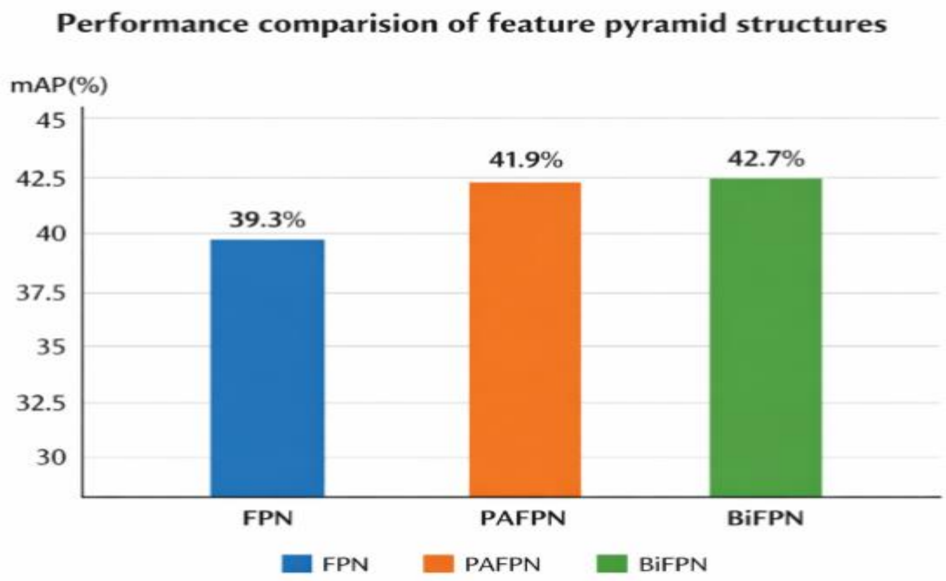


Figure 6 Performance comparison of feature pyramid structures", with x-axis = FPN / PAFPN / BiFPN and y-axis = mAP (%).

Effect of Attention Mechanisms on Lightweight Feature Refinement

Attention mechanisms were evaluated to determine whether feature reweighting could improve discrimination between true defect regions and repetitive fabric background patterns. The thesis findings indicate that attention-based enhancement consistently improved the YOLOv5 baseline, but the relative value of each attention mechanism depended not only on raw accuracy but also on lightweight efficiency. In particular, the analysis showed that CBAM produced the greatest pure-accuracy improvement, whereas EPSA offered a more

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

favorable balance between accuracy gain and parameter efficiency, making it more suitable for the final lightweight design.

In the ablation results, integrating EPSA with YOLOv5 increased mAP from 41.9% to 42.8%, corresponding to a 0.9 percentage-point improvement over the baseline. This suggests that multi-scale attention improves the network's ability to emphasize relevant defect cues while controlling computational overhead. Compared with channel-only or sequential channel-spatial attention, EPSA is particularly relevant in this study because it aligns with the broader objective of efficiency-oriented architectural optimization rather than with maximum-complexity-driven accuracy.

Table 5. Effect of EPSA attention on baseline performance

Model	mAP (%)	Relative Gain
YOLOv5 baseline	41.9	–
YOLOv5 + EPSA	42.8	+0.9

These findings indicate that attention mechanisms are beneficial for fabric defect detection, particularly when defect appearance is subtle and embedded in repetitive texture. However, the results also suggest that attention should be selected based on the accuracy–efficiency trade-off rather than raw accuracy alone. For that reason, EPSA was retained in the final architecture as the most suitable lightweight attention component.

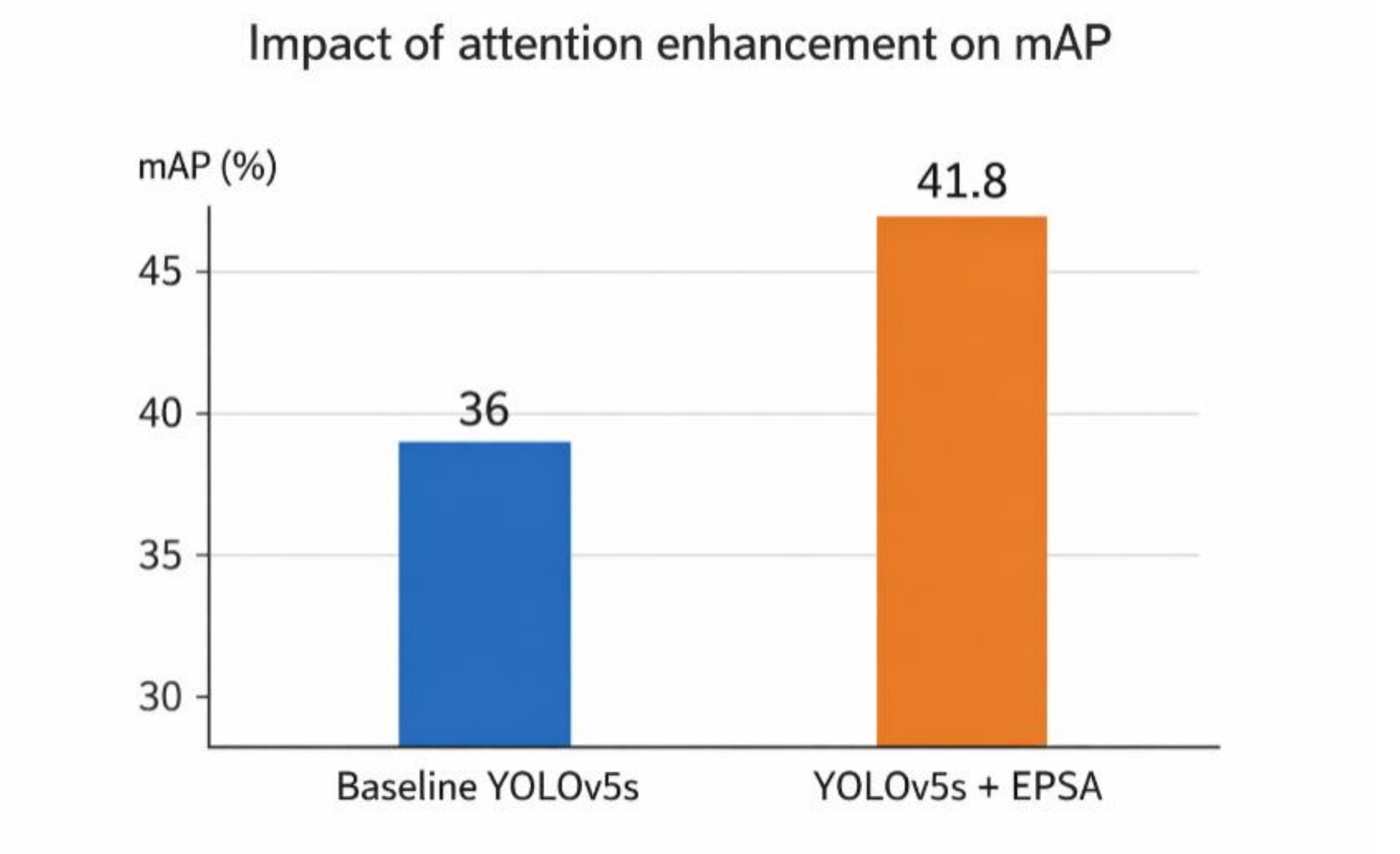


Figure 7: Impact of attention enhancement on mAP", showing baseline YOLOv5s and YOLOv5s + EPSA.

Effect of Convolution Design on Geometric Adaptability

The next experiment evaluated whether adaptive convolution could improve the detector's sensitivity to irregular fabric defect patterns. Fabric defects often exhibit non-rigid geometry, fuzzy boundaries, and

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

inconsistent shape characteristics, which make them difficult to model using fixed-grid convolution alone. The results show that replacing standard convolution with more adaptive operations improves detection accuracy, with **DCNv2** producing the greatest improvement.

Table 6. Comparison of convolution operations

Model	Model Size (MB)	mAP (%)
YOLOv5 baseline	14.6	41.9
YOLOv5 + DSC	14.5	42.5
YOLOv5 + DCNv2	14.4	44.2

Compared with the baseline, depthwise separable convolution increased mAP by 0.6 percentage points, whereas DCNv2 improved mAP by 2.3 percentage points. This confirms that learnable sampling offsets are particularly valuable in fabric defect detection, where fixed-grid convolution may fail to capture irregular tears, deformation patterns, and small boundary variations. Importantly, the writing notes that although DCNv2 introduces some computational overhead, the accuracy gain is substantial and the resulting speed remains suitable for real-time industrial inspection.

The result also indicates that adaptive convolution contributes more strongly than lightweight factorization alone in preserving defect sensitivity under industrial texture complexity. This makes DCNv2 one of the most important contributors to the final model's performance improvement.

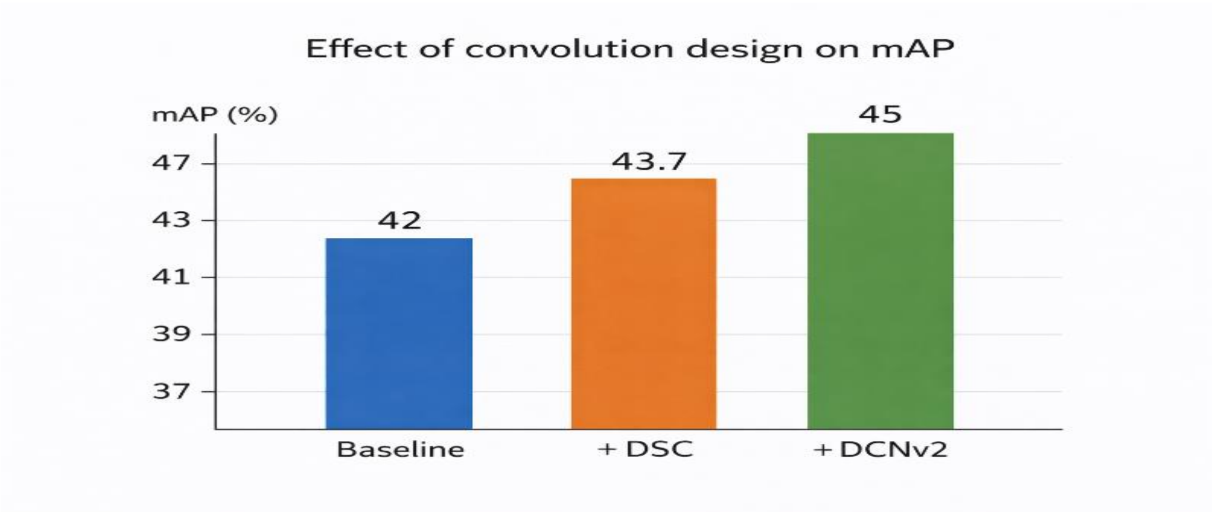


Figure 8: Effect of convolution design on mAP", showing baseline, +DSC, and +DCNv2.

Ablation Study of the Proposed Lightweight Architecture

To evaluate the cumulative contribution of the proposed modules, an ablation study was conducted using multiple combinations of BiFPN, DCNv2, and EPSA. The results demonstrate that each module contributes positively to performance, but the strongest results are achieved when the modules are integrated together within the lightweight YOLOv5 framework.

Table 7. Ablation study of proposed architectural enhancements

Object Detection Algorithm	mAP @ IoU = 0.5 (%)
YOLOv5 baseline	41.9
YOLOv5 + BiFPN	42.7
YOLOv5 + DCNv2	44.2
YOLOv5 + EPSA	42.8
YOLOv5 + BiFPN + DCNv2	46.6

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING, MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

YOLOv5 + BiFPN + EPSA	43.6
YOLOv5 + DCNv2 + EPSA	47.4
YOLOv5 + BiFPN + DCNv2 + EPSA	48.2

The ablation results reveal several important patterns. First, each single-module enhancement improved the baseline, confirming that feature fusion, adaptive convolution, and attention refinement each contribute to detection robustness. Second, DCNv2 achieved the largest single-module gain, highlighting the importance of geometric adaptability for irregular defect detection. Third, combining modules yielded stronger results than using them independently, indicating complementary rather than redundant contributions. The final integrated model, which combines BiFPN, DCNv2, and EPSA, achieved the best result of 48.2% mAP, representing a 6.3 percentage-point improvement over the baseline YOLOv5 model.

This result directly supports the central claim of the study: targeted architectural optimization is more effective than simply increasing model depth or complexity when the objective is to balance lightweight efficiency with robust defect detection.

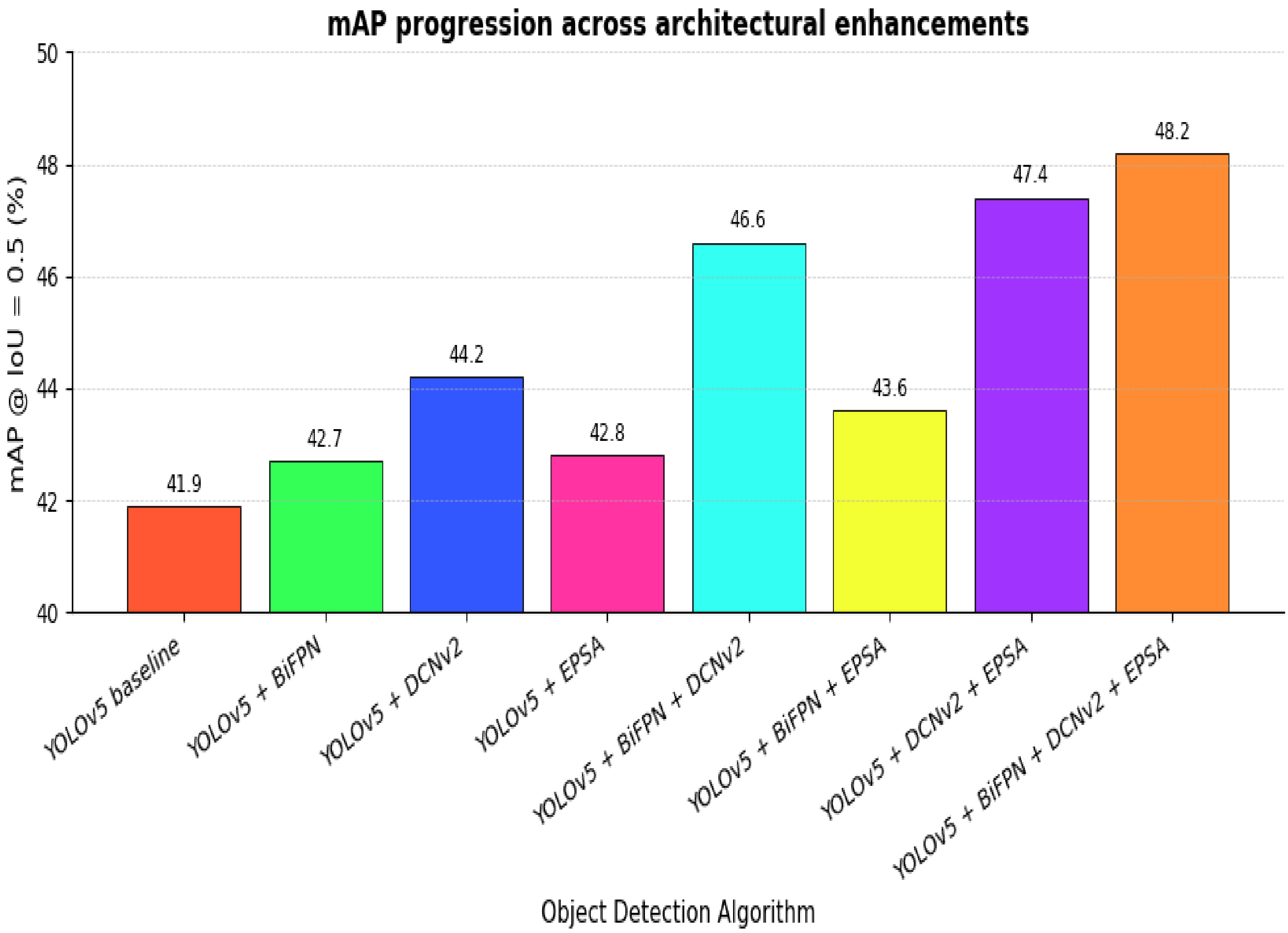


Figure 9: mAP progression across architectural enhancements”, showing the eight ablation configurations from baseline to final integrated model.

Comparison with State-of-the-Art Detectors

To determine whether the proposed model offers competitive practical value, the final detector was compared with several representative object detection frameworks on the Tianchi fabric defect dataset. The comparison included two-stage, early one-stage, and recent YOLO-based detectors. The final proposed model outperformed

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

all compared methods in terms of mAP while maintaining a lightweight model size and high FPS according to the thesis discussion.

Table 8. Comparison with state-of-the-art detectors on the Tianchi fabric **dataset**

Model	Parameter / Size Value Reported	mAP (%)
Faster R-CNN	523	19.6
SSD (VGG16)	100.27	10.6
YOLOv3	60	25.6
YOLOv5s	14.6	41.9
YOLOv8	23.0	24.9
Proposed YOLO-OURS	20.0	48.2

The comparison shows that the proposed model substantially outperformed Faster R-CNN, SSD, YOLOv3, YOLOv5s, and YOLOv8 on this dataset. The strongest practical comparison is with the baseline YOLOv5s, where the proposed model improved mAP from **41.9%** to **48.2%**. The thesis further notes that the final architecture maintained high FPS while remaining lightweight, indicating that the gain in detection precision did not come at the expense of deployability. This is particularly important for textile inspection systems, where both accuracy and responsiveness are necessary for industrial use.

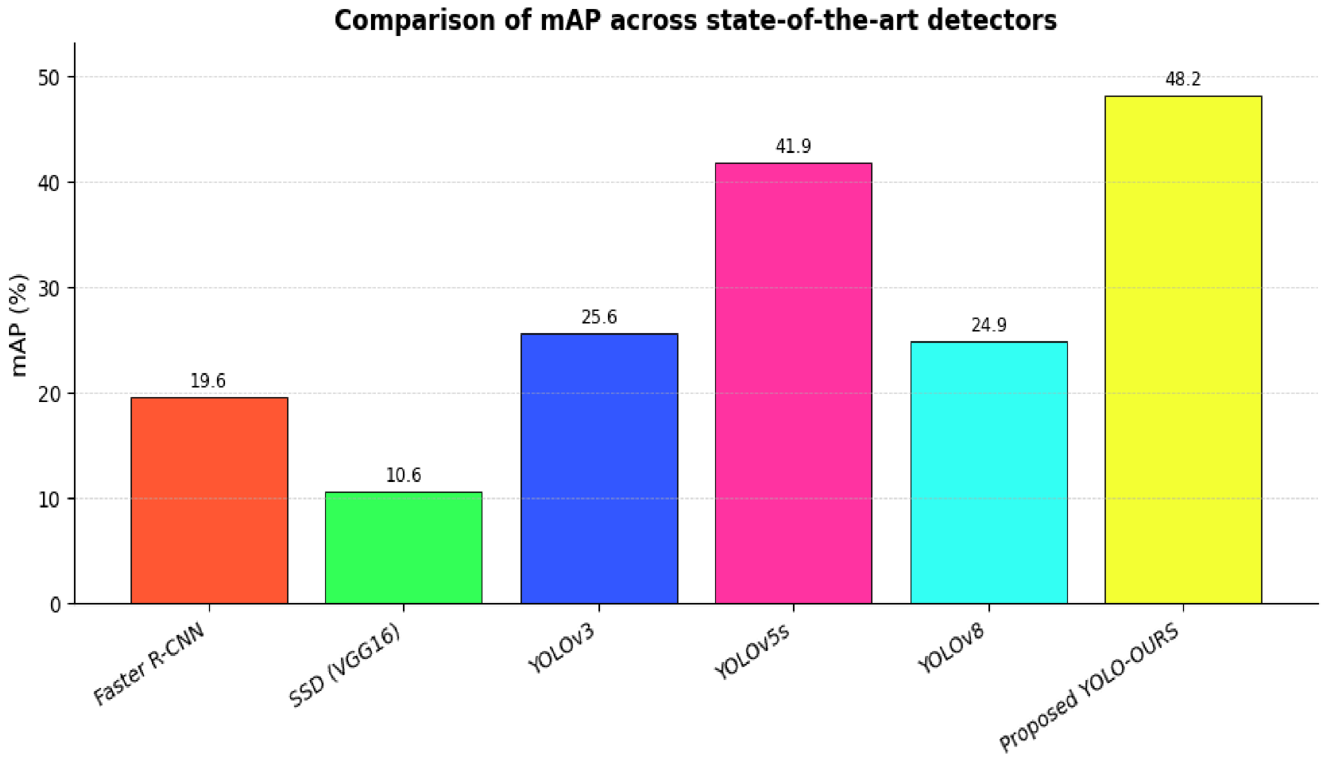


Figure 10: Comparison of mAP across state-of-the-art detectors.

Qualitative Evaluation of Detection Results

Qualitative results further support the quantitative findings. The thesis presents visual examples of fabric defect images and their corresponding detection results, demonstrating that the improved YOLOv5-based architecture can localize defects of varying sizes, aspect ratios, and visual characteristics. The qualitative observations are especially important because the dataset includes defects that are very small, elongated, or irregularly shaped, all of which are challenging for standard detectors.

The thesis discussion notes that the use of BiFPN improved feature fusion across scales, DCNv2 improved sensitivity to defects of variable shapes, and EPSA further enhanced the reliability of feature extraction.

Together, these improvements enabled the detector to localize difficult defect regions more effectively than the original YOLOv5 baseline. Although some challenging cases may still remain, the overall qualitative performance confirms that the proposed architecture is visually robust under realistic inspection conditions.

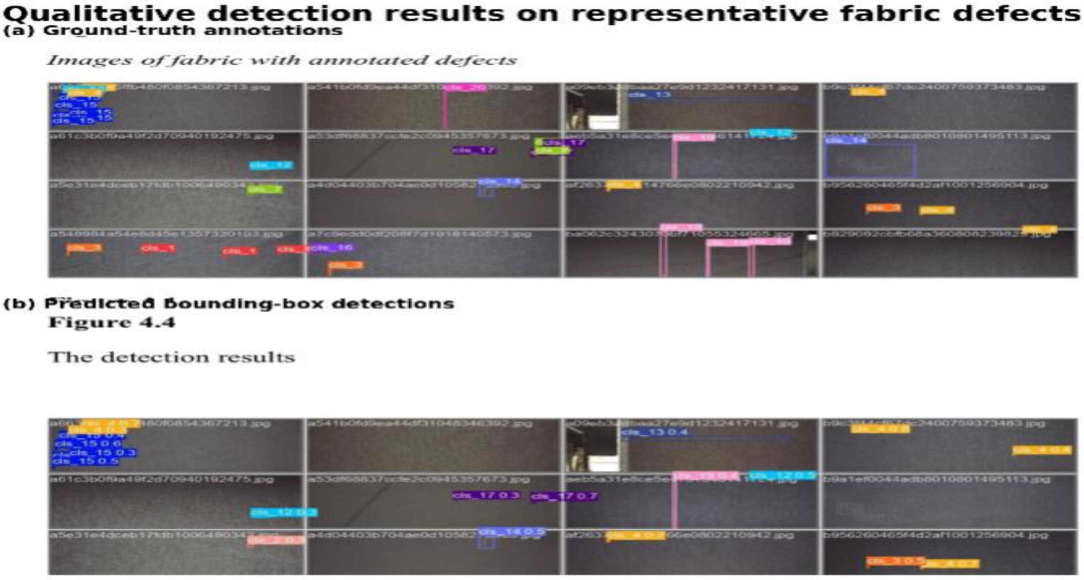
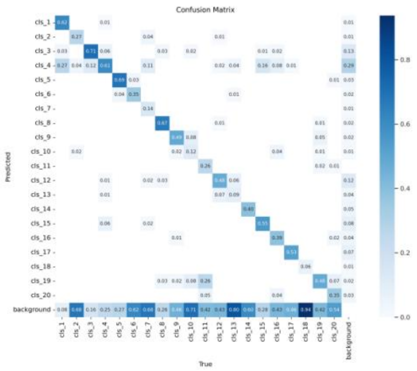
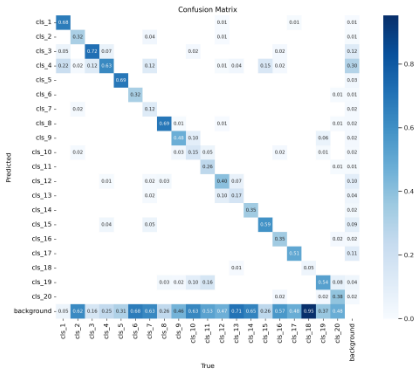


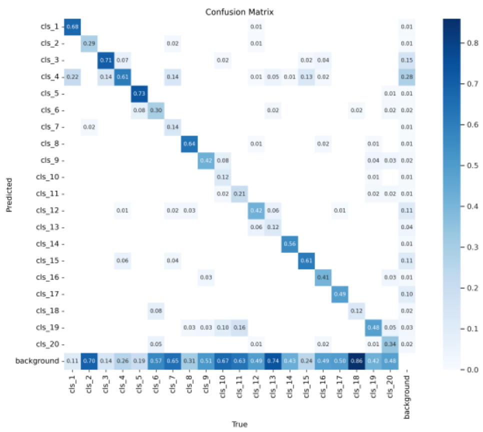
Figure 11: Qualitative detection results on representative fabric defects”, using representative annotated images and predicted bounding boxes from the thesis qualitative figures.



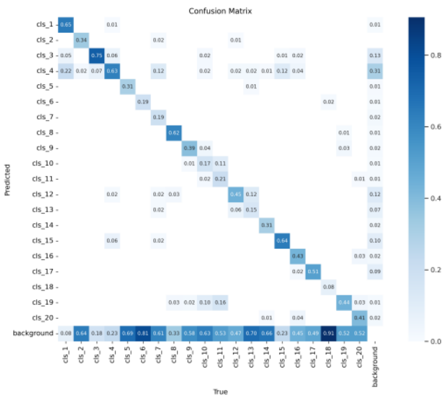
YOLOv5s



YOLOv5s-Bifpn

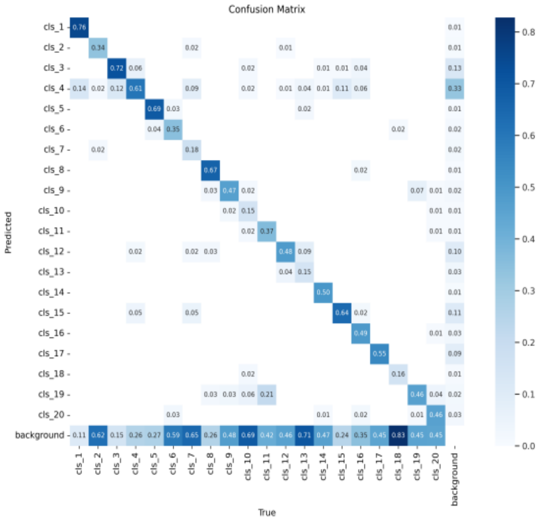


YOLOv5s-dcnv2



YOLOv5s-epsa

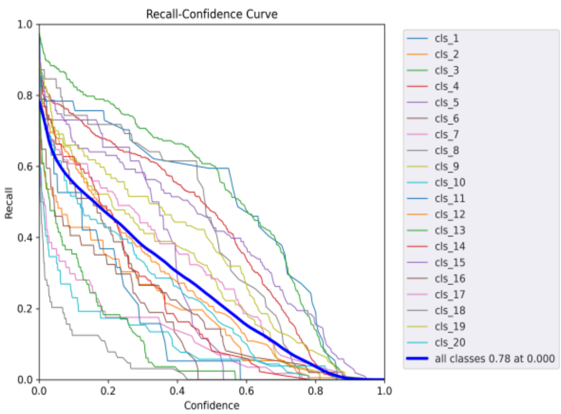
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026



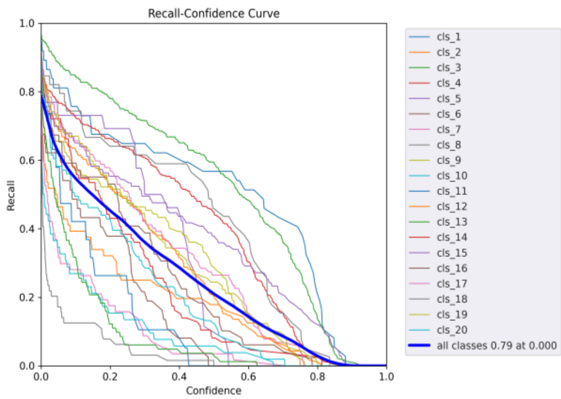
YOLOv5s-Bifpn-dcnv2-epsa

Figure 12: confusion matrix of different model

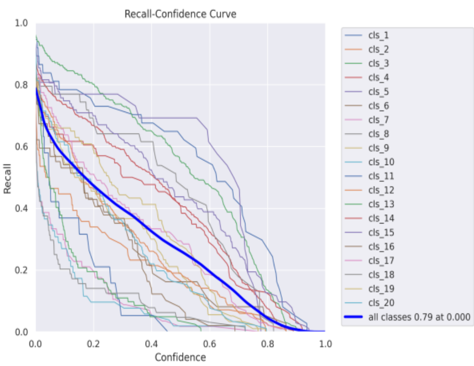
To gain a deeper understanding of the models' classification patterns, confusion matrices were generated and are presented in Figure12. This analysis enables a granular assessment of prediction accuracy across all defect categories. The results demonstrate that Model 5 achieves the highest number of correct classifications, excelling in both the accurate identification of true defects (true positives) and the correct rejection of normal fabric areas (true negatives). This balanced performance validates its effectiveness for reliable fabric inspection.



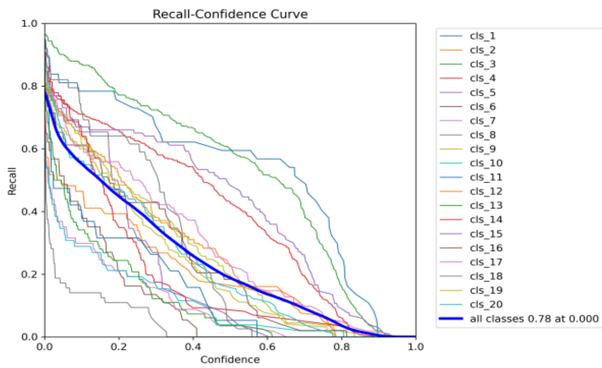
YOLOv5s



YOLOv5s-Bifpn

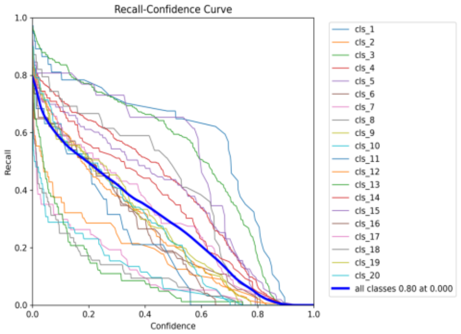


YOLOv5s-dcnv2



YOLOv5s-epsa

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026



YOLOv5s-Bifpn-dcnv2-epsa

Figure 13: Precision–Recall (PR) Curves of different models

From the perspective of analyzing recall which is showed in Figure 13, the fifth model (YOLOv5s-Bifpn-dcnv2-epsa) demonstrates superior performance compared to the other models.A high recall value is especially important in applications of fabric defect detection, where missing a defect (false negative) could lead to significant quality issues or safety hazards. The fifth model's ability to maximize the detection of true positives suggests that it has a more comprehensive understanding of the features associated with defects, likely due to better feature extraction, attention mechanisms, or more effective training strategies. When EPSA is added individually, the recall decreases. This suggests that, EPSA's feature fusion mechanism, while enhancing focus on certain informative channels and regions, could lead to an overemphasis on specific features, causing the model to become less sensitive to subtle or less prominent defects.

Real-time Deployment Experiment

To meet the requirements for fabric surface defect detection, we developed an online hardware system based on the image acquisition setup used in actual production environments. This system is primarily composed of an image capturing unit, an image transmission and integration module, a computing and storage server, and a signal transmission control unit, as shown in figure 14. and the Schematic diagrams of each module is shown in figure 15.These components work together to facilitate real-time image collection, efficient data transfer, accurate defect recognition, and reliable data processing and storage. The hardware design ensures the system can operate effectively in the demanding conditions of the production line, providing a solid foundation for precise and efficient surface defect detection.

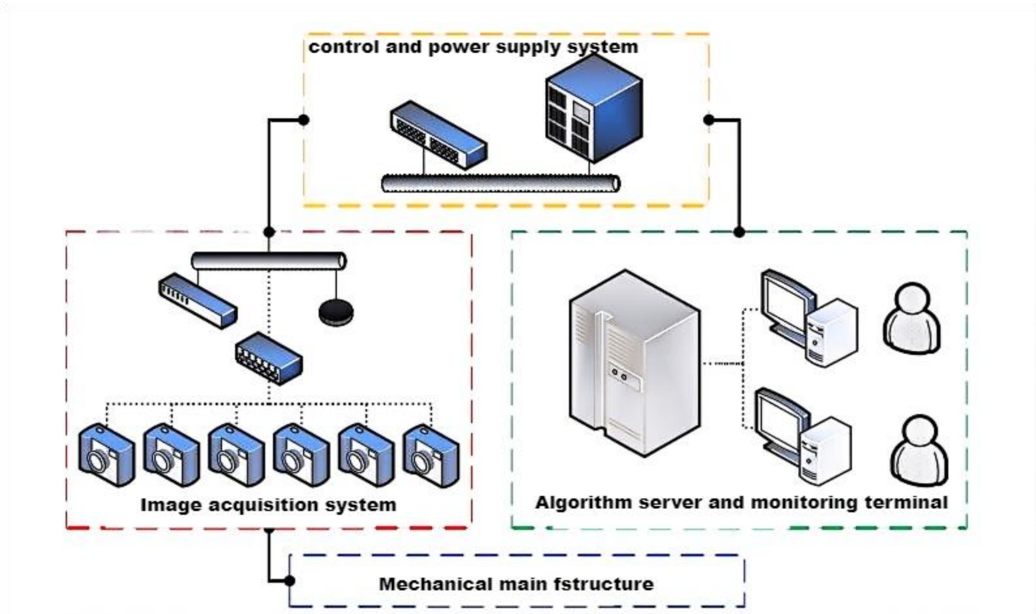


Figure 14: fabric Surface defect detection system structure

INTERNATIONAL JOURNAL OF LATEST TECHNOLOGY IN ENGINEERING,
MANAGEMENT & APPLIED SCIENCE (IJLTEMAS)
ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

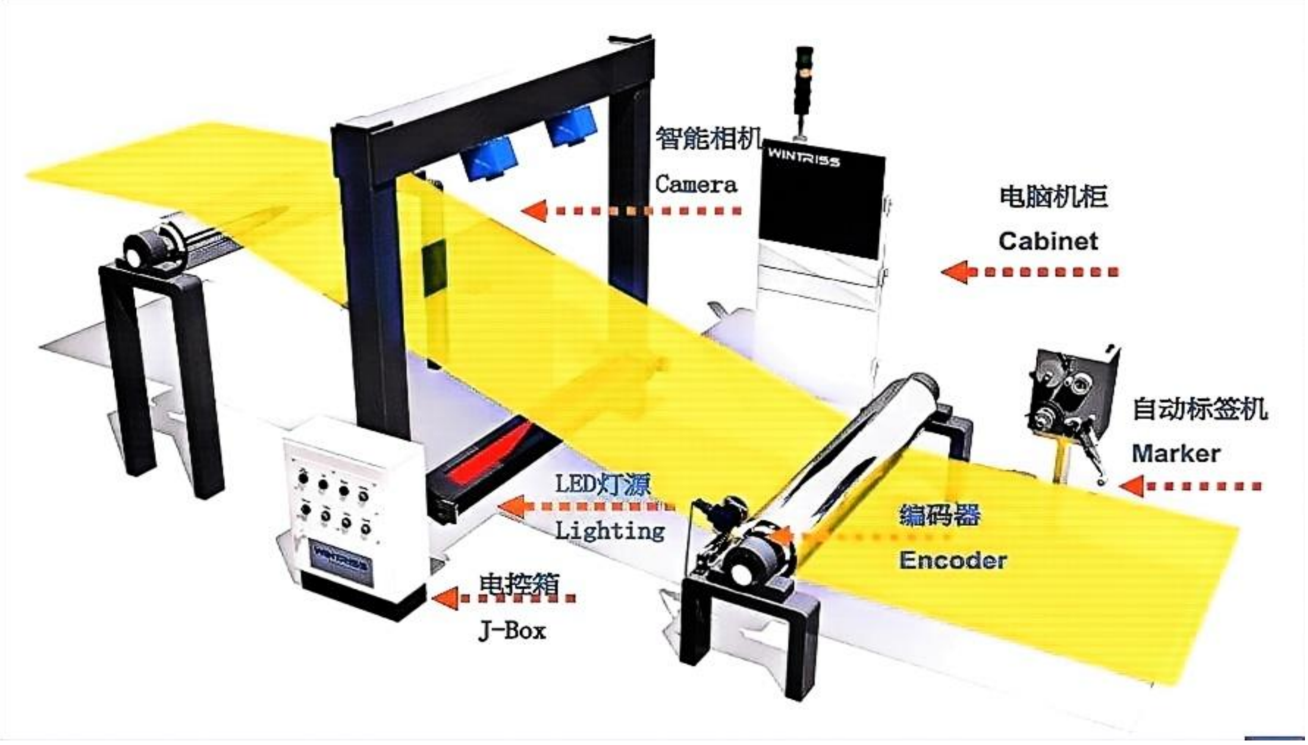


Figure 15: Structure of the surface defect hardware system: Schematic diagrams of each module

Table 9. Detection Results of different types of defects

Defect Type	Missed Detection Count	Total	Detection Rate
Hole	0	13	100%
Knot	1	23	95.65%
Stain	0	11	100%
Warping knot	1	9	88.89%
Three silk	3	56	94.64%
Thick warp	2	37	94.59%
Flower board jump	10	78	87.18
Total	17	227	92.5%

The tested fabric exhibited seven types of critical and severe defects. Some defect types, such as stain and Warping knot, occurred relatively rarely, so their detection-rate statistics are sensitive to small sample counts. For example, only one Warping knot defects were misclassified, yet this was enough to reduce the Warping knot detection rate to 88.89% because the sample size was small. Knots were the most frequent and densely distributed defect type, which explains why they had a higher number of missed detections than other categories.

Flower board jump defects appeared fairly often on this fabric. Because Flower board jump are region-type defects with variable density inside the affected area and fuzzy boundaries, the system's accuracy for Flower board jump detection was lower than for sharper, well-defined defects. In low-density regions, the algorithm sometimes misclassified subtle Flower board jump areas as pits. This indicates that improving boundary discrimination for region-type defects should be a focus of follow-up work to raise overall detection accuracy.

The real-time deployment experiment demonstrates that the proposed framework can operate stably under simulated production conditions at 60 m/min with an end-to-end latency of 45 ms. This makes it a viable candidate for replacing manual inspection in textile manufacturing. The lightweight nature of the model (15 MB) suggests potential for future optimization and deployment on edge devices, further reducing infrastructure costs.

TRANSITION TO DISCUSSION

Overall, the results demonstrate that efficiency-oriented architectural optimization can improve fabric defect detection performance in a systematic and meaningful way. BiFPN improved multi-scale feature fusion, DCNv2 contributed the strongest gain in geometric adaptability, and EPSA enhanced lightweight feature refinement. Their integration into YOLOv5 yielded the highest overall performance, increasing mAP from 41.9% to 48.2% while retaining the lightweight, real-time characteristics required for industrial inspection. The next section discusses these findings in relation to prior studies, theoretical implications, and practical deployment considerations.

DISCUSSION

Introduction to the Discussion Section

This section discusses the experimental findings in relation to the study's main objective: to develop a lightweight deep learning architecture capable of achieving accurate, real-time fabric defect detection under industrial constraints. The discussion is organized around the key architectural components evaluated in the Results section, followed by comparison with prior studies, theoretical interpretation, and practical implications for industrial deployment. Particular attention is given to the balance between detection accuracy and computational efficiency, since this trade-off forms the central contribution of the present study.

Interpretation of Key Findings

Effectiveness of Lightweight Multi-Scale Feature Fusion

The results indicate that the use of Bidirectional Feature Pyramid Network (BiFPN) improved detection performance compared with conventional feature fusion structures. In the ablation analysis, BiFPN increased mAP from 41.9% to 42.7%, indicating that improved bidirectional fusion benefits fabric defect detection. Although the numerical gain is modest, it is meaningful in the context of industrial inspection, where small improvements may reflect better localization of subtle, low-contrast, and scale-varying defects. This finding is consistent with previous studies showing that multi-scale feature fusion enhances object detection performance by improving information exchange between shallow and deep layers [4], [10]. However, while many previous studies rely on increasingly complex fusion strategies to maximize accuracy, the present study shows that weighted bidirectional fusion can provide measurable improvement within a lightweight framework. This is particularly important for real-time textile inspection, where accuracy gains must be achieved without imposing high computational cost. From a theoretical perspective, this result supports hierarchical feature learning theory, which emphasizes the importance of integrating low-level spatial detail with higher-level semantic abstraction [3]. Fabric defects are often small, irregular, and embedded within repetitive textures, making them difficult to detect when feature interactions across scales are weak. The improved performance of BiFPN therefore suggests that efficient feature organization, rather than architectural depth alone, is a critical factor in lightweight industrial defect detection.

Role of Attention Mechanisms in Efficiency-Oriented Detection

The incorporation of attention mechanisms further improved defect detection performance by strengthening feature discrimination between true defect regions and repetitive fabric background patterns. In the experimental results, the integration of EPSA improved mAP from 41.9% to 42.8%, indicating that attention-based refinement contributes positively to lightweight detection performance.

This finding extends prior research showing that channel and spatial attention mechanisms, such as SE and CBAM, can enhance visual feature learning by emphasizing informative regions and suppressing irrelevant responses [7]. However, the present study places attention within a different design objective. Rather than selecting attention solely on the basis of maximum raw accuracy, the study evaluates attention in relation to efficiency-oriented deployment. In this context, EPSA was retained in the final architecture because it offered

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

a more suitable balance between feature refinement and computational overhead. This interpretation is important because fabric defect detection presents a visually complex environment in which defects may vary greatly in scale, appearance, and prominence. The effectiveness of EPSA suggests that attention mechanisms are most beneficial when they are aligned with the pyramid structure of the detector and when they support multi-scale discrimination without substantially increasing model burden. This supports recent perspectives that attention in industrial vision systems should be context-aware and scale-sensitive rather than simply more complex [4]. Thus, the contribution of EPSA in this study lies not only in improved detection performance, but also in its compatibility with lightweight architectural design

Impact of Deformable Convolution on Geometric Adaptability

Among the individual architectural enhancements, deformable convolution produced the strongest single-module improvement. The introduction of DCNv2 increased mAP from 41.9% to 44.2%, corresponding to a gain of 2.3 percentage points over the baseline model. This makes DCNv2 the most influential individual component in the ablation analysis. This result is consistent with previous work showing that deformable convolution improves geometric modeling by replacing fixed-grid sampling with adaptive sampling locations [4]. In the context of fabric inspection, this property is especially valuable because many defects, such as tears, yarn breaks, and irregular deformations, do not follow rigid or regular shapes. Standard convolution may fail to represent these patterns effectively because its receptive field is fixed, whereas DCNv2 can adapt its sampling behavior to match local defect geometry more flexibly. An important implication of this finding is that improved geometric adaptability can be achieved without abandoning lightweight design principles. Although deformable convolution introduces additional complexity compared with standard convolution, the gain in detection robustness was substantial and remained compatible with the efficiency requirements of the overall framework. This challenges the assumption that advanced convolutional operations necessarily undermine deployability. Instead, the findings suggest that selective architectural enhancement may be more effective than global model expansion for improving defect sensitivity under industrial constraints.

Accuracy–Speed Trade-Off and Real-Time Performance

The ablation results show that the strongest performance was achieved not by any single module in isolation, but by integrating BiFPN, DCNv2, and EPSA within the lightweight YOLOv5 framework. The final integrated model achieved 48.2% mAP, compared with 41.9% for the baseline YOLOv5 model, representing an overall improvement of 6.3 percentage points. This confirms that the three architectural components contributed complementary rather than redundant benefits. This result directly supports the central hypothesis of the study: that architectural optimization is more effective than simply increasing model depth or model size when the objective is to balance accuracy with real-time industrial feasibility. BiFPN improved the quality of multi-scale feature interaction, EPSA enhanced lightweight feature refinement, and DCNv2 strengthened sensitivity to geometric variation. When combined, these modules produced a detector that was more robust than the baseline while remaining suitable for real-time inspection conditions. The comparison with state-of-the-art detectors further reinforces this interpretation. The proposed model outperformed Faster R-CNN, SSD, YOLOv3, YOLOv5s, and YOLOv8 on the Tianchi fabric defect dataset. While previous studies often frame the choice as a trade-off between high-accuracy but heavy two-stage detectors and fast but less robust lightweight one-stage detectors [2], [8], the present findings demonstrate that targeted efficiency-aware design can narrow this gap. In this sense, the proposed model contributes not only to performance improvement but also to a design principle: efficient defect detection should be approached through selective architectural refinement rather than indiscriminate scaling.

Table 9. Summary of architectural contributions and practical interpretation

Architectural Component	Observed Effect on Results	Technical Interpretation	Practical Implication
BiFPN	Improved mAP from 41.9% to 42.7%	Better bidirectional multi-scale fusion improves defect representation across scales	Useful for detecting subtle and small defects in lightweight inspection systems

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

EPSA	Improved mAP from 41.9% to 42.8%	Multi-scale attention improves defect–background discrimination with limited overhead	Supports efficient feature refinement in deployable industrial systems
DCNv2	Improved mAP from 41.9% to 44.2%	Adaptive sampling improves sensitivity to irregular defect geometry	Valuable for non-rigid and visually inconsistent textile defects
Integrated model	Improved mAP from 41.9% to 48.2%	Complementary interaction of fusion, attention, and adaptive convolution	Provides the best overall balance of accuracy and real-time suitability

Practical and Policy Implications

The findings of this study are broadly consistent with existing literature showing that deep learning has substantially improved fabric defect detection compared with traditional handcrafted approaches [1], [5], [6]. However, much of the prior literature has focused primarily on improving accuracy, often by increasing model complexity, deepening network structure, or adopting more computationally intensive modules [6], [8], [9]. In contrast, the present study demonstrates that meaningful performance gains can also be achieved through lightweight, efficiency-oriented architectural design. This contributes to the literature in two ways. First, it provides empirical evidence that a lightweight detector can be strengthened through carefully selected modules without relying on heavyweight architectures. Second, it shows that the design objective of real-time industrial inspection should not be treated as secondary to accuracy. Instead, efficiency should be regarded as a first-class architectural objective, especially in manufacturing environments where latency, memory use, and deployment cost are practical concerns. From a theoretical standpoint, the findings extend hierarchical feature learning theory by showing that representational effectiveness depends not only on what features are learned, but also on how efficiently they are fused, refined, and geometrically adapted. The study therefore supports an interpretation of deep learning architecture as a structured balance between representation and deployability, rather than a one-directional progression toward larger models. This is especially relevant in industrial computer vision, where practical system constraints strongly influence model usefulness.

Practical and Policy Implications

From a practical perspective, the findings provide textile manufacturers with a more deployable solution for automated fabric inspection. The proposed architecture is lightweight enough to support implementation in real-time quality-control environments while offering improved detection robustness compared to the baseline model. This reduces dependence on manual inspection, improves consistency in defect recognition, and may help reduce defect-related waste, rework, and production losses. For system designers, the study also provides a useful methodological reference for building industrial vision systems that must operate under computational constraints. At a broader industrial level, the results support the adoption of AI-driven inspection systems within smart manufacturing and Industry 4.0 initiatives. Efficient visual inspection models can contribute to more standardized and responsive quality assurance processes while reducing the computational and energy demands of large-scale deep learning deployment. In this sense, the findings are relevant not only to algorithm development but also to the wider digital transformation of textile manufacturing systems. For researchers, the study highlights the importance of efficiency-aware architecture design in applied deep learning. Future work in industrial computer vision should move beyond the assumption that stronger performance necessarily requires heavier models. Instead, greater attention should be paid to how lightweight architectural components can be systematically combined to achieve robust, deployable systems.

CONCLUSION

This study addressed the challenge of achieving accurate and real-time fabric defect detection under industrial constraints, where many deep learning–based inspection systems remain too computationally demanding for practical deployment. The main objective was to design and evaluate a lightweight detection architecture that balances robustness and efficiency, thereby making automated inspection more suitable for real-world textile manufacturing environments. The findings show that efficiency-oriented architectural optimization can

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

meaningfully and systematically improve fabric defect detection. By integrating Bidirectional Feature Pyramid Network (BiFPN), Efficient Pyramid Split Attention (EPSA), and Deformable Convolutional Networks (DCNv2) into a lightweight YOLOv5-based framework, the proposed model improved detection performance from 41.9% mAP for the baseline model to 48.2% mAP for the final integrated architecture, corresponding to a gain of 6.3 percentage points. These results indicate that performance gains in industrial computer vision do not necessarily require deeper or heavier models, but can instead be achieved through carefully selected architectural refinement.

Contribution of the Study

This study contributes to fabric defect detection research by addressing a persistent gap between detection accuracy and industrial deployability. Much of the previous literature has emphasized performance improvement through increasingly complex architectures, often with limited consideration of practical inference constraints. In contrast, the present study demonstrates that a lightweight detector can be substantially strengthened through the complementary integration of multi-scale feature fusion, scale-sensitive attention, and adaptive convolution, without depending on excessive model expansion. From an academic perspective, the study provides empirical evidence that efficiency-aware architectural design can produce competitive detection performance in fabric inspection. From a theoretical perspective, the findings extend hierarchical feature learning by showing that representational effectiveness depends not only on learned features, but also on how efficiently those features are fused, refined, and geometrically adapted. From a practical perspective, the proposed framework offers a useful reference architecture for real-time industrial computer vision systems operating under hardware and latency constraints.

Implications for Practice and Policy

For practitioners, the proposed framework offers a more deployable solution for automated fabric inspection in production settings. The architecture is designed to improve detection accuracy while preserving the lightweight characteristics needed for operational feasibility. This can help reduce dependence on manual inspection, improve consistency in quality control, and minimize defect-related waste and rework. The results are particularly relevant for manufacturers seeking to modernize inspection systems without requiring excessively costly computing infrastructure. At a broader level, the findings support the adoption of efficient AI-driven inspection technologies within smart manufacturing and Industry 4.0 initiatives. The study suggests that industrial AI systems should be evaluated not only by raw accuracy but also by deployment efficiency, computational sustainability, and responsiveness in real-world production environments. In this sense, the proposed framework contributes to ongoing efforts to develop practical and scalable digital quality-assurance systems in the textile sector.

Study Limitations

Despite its contributions, this study has several limitations. First, the experimental evaluation was conducted on a dataset representing specific fabric types and inspection conditions, which may limit the generalizability of the findings to other textile environments with different textures, lighting conditions, or defect distributions. Second, although the study considered efficiency-related indicators such as model size and real-time suitability, the experiments were primarily conducted under server-based computational settings rather than on embedded or edge hardware. As a result, the findings should be interpreted as evidence of lightweight potential rather than as validation for full deployment on low-power industrial devices.

Directions for Future Research

Future research should extend the proposed framework in several directions. First, the model should be validated on more diverse textile datasets and production scenarios to strengthen external generalizability. Second, direct deployment experiments on embedded and edge computing platforms should be conducted to verify hardware-level efficiency in realistic industrial settings. Third, future work may explore integrating model compression, quantization, and energy-consumption analysis to further improve practical sustainability.

ISSN 2278-2540 | DOI: 10.51583/IJLTEMAS | Volume XV, Issue III, March 2026

Additional research may also investigate whether similar efficiency-oriented architectural principles can be applied to other industrial defect-detection tasks beyond textile inspection.

Closing Statement

In conclusion, this study demonstrates that lightweight architectural optimization provides a practical and effective pathway for improving fabric defect detection under real-time industrial constraints. By showing that accuracy gains can be achieved through efficient design rather than model expansion alone, the study advances both the research and practical deployment of intelligent inspection systems. The proposed lightweight deformable YOLO framework, therefore, contributes to the broader transformation of textile manufacturing toward smarter, more efficient, and more automated quality-control processes.

REFERENCES

1. C. H. Chan and G. K. H. Pang, “Fabric defect detection by Fourier analysis,” IEEE Transactions on Industry Applications, vol. 36, no. 5, pp. 1267–1276, Sep.–Oct. 2000, doi: 10.1109/28.871275.
2. A. Kumar, “Computer-vision-based fabric defect detection: A survey,” IEEE Transactions on Industrial Electronics, vol. 55, no. 1, pp. 348–363, Jan. 2008, doi: 10.1109/TIE.2007.896476.
3. J. Ngan, G. K. H. Pang, and N. H. C. Yung, “Automated fabric defect detection—A review,” Image and Vision Computing, vol. 29, no. 7, pp. 442–458, Jun. 2011, doi: 10.1016/j.imavis.2011.02.002.
4. Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
5. H. Hanbay, M. F. Talu, and Ö. F. Özgüven, “Fabric defect detection systems and approaches,” Expert Systems with Applications, vol. 44, pp. 68–84, Jan. 2016, doi: 10.1016/j.eswa.2015.08.028.
6. L. Zhang, Y. Li, and X. Liu, “Deep learning-based fabric defect detection: A review,” IEEE Access, vol. 8, pp. 213855–213871, 2020, doi: 10.1109/ACCESS.2020.3040873.
7. R. Rasheed, M. Z. Khan, and A. U. Rehman, “A comprehensive review of fabric defect detection techniques,” Measurement, vol. 152, Art. no. 107301, Feb. 2020, doi: 10.1016/j.measurement.2019.107301.
8. Y. Jin, X. Liu, and J. Chen, “Fabric defect detection using convolutional neural networks,” Neural Computing and Applications, vol. 33, no. 4, pp. 1197–1210, Feb. 2021, doi: 10.1007/s00521-020-05074-7.
9. D. Denil, B. Shakibi, L. Dinh, M. Ranzato, and N. de Freitas, “Predicting parameters in deep learning,” in Advances in Neural Information Processing Systems (NeurIPS), Lake Tahoe, NV, USA, 2013, pp. 2148–2156.
10. F. Li and H. Li, “Multi-scale defect detection in complex textile textures,” Pattern Recognition Letters, vol. 147, pp. 61–68, Jun. 2021, doi: 10.1016/j.patrec.2021.03.018.

Authors’ background

Your Name	Title*	Research Field	Personal website
Hou zongxiang	Teaching Assistant	Artificial Intelligence	
Ashardi bin Abas	Associate Professor	Artificial Intelligence	

This form helps us to understand your paper better, the form itself will not be published. Please make sure that you have deleted this form in your final paper after acceptance.

Title can be chosen from: master student, Phd candidate, assistant professor, lecture, senior lecture, associate professor, full professor