Enhancing Fabric Defect Detection Using Efficient Pyramid Split Attention in a Lightweight YOLOv5 Framework

Hou zongxiang and Ashardi bin Abas

Faculty of Computing And Meta-Technology,University Pendidikan Sultan Idris

DOI: https://doi.org/10.51583/IJLTEMAS.2026.150300016

Received: 16 March 2026; Accepted: 21 March 2026; Published: 02 April 2026

ABSTRACT

Fabric defect detection is a fundamental quality control process in textile manufacturing, yet achieving accurate and reliable automated inspection remains difficult because of complex background textures, subtle defect patterns, and substantial variation in defect scale and shape. Although deep learning–based detectors have improved inspection performance, many lightweight models still suffer from limited feature discrimination, particularly in real-time industrial environments where computational efficiency is critical. To address this limitation, this study proposes an enhanced fabric defect detection framework by integrating an Efficient Pyramid Split Attention (EPSA) mechanism into a YOLOv5-based convolutional network. The EPSA module is designed to adaptively recalibrate multi-scale feature responses, enabling the network to emphasize defect-relevant information more effectively while preserving inference efficiency. A quantitative experimental design was employed using a labeled fabric defect image dataset, and the proposed model was evaluated through comparative and ablation analyses against baseline and alternative attention-based configurations. Experimental results indicate that the EPSA-enhanced model achieves superior detection performance in terms of mean Average Precision while maintaining real-time processing capability. The improvement is especially evident for small, low-contrast, and irregular defects embedded in repetitive fabric textures. These findings confirm that pyramid-based attention can substantially improve feature representation without imposing significant computational overhead. The proposed approach offers a practical and efficient solution for automated textile inspection and provides a useful foundation for future research on lightweight attention modeling for industrial vision systems.

Keywords: fabric defect detection; attention mechanism; EPSA; convolutional neural networks

INTRODUCTION

Background and Context

Fabric defect detection is a fundamental function in textile quality assurance because surface defects directly affect product grade, commercial value, downstream processing, and customer acceptance. Defects such as holes, stains, broken yarns, knots, abrasion marks, and texture irregularities reduce fabric usability and can significantly lower market value. Underscoring the economic importance of accurate and early inspection in textile manufacturing [1]. The same source further notes that severe defects may lead to revenue declines, product rejections, and reputational damage for merchants, especially in high-volume apparel production that requires strict quality control [1].

The industrial significance of this problem is increasing as textile production moves toward smart manufacturing, automated quality assurance, and continuous high-throughput processing. In such environments, inspection systems must operate not only with acceptable detection accuracy, but also with consistency, speed, and practical deployability under real production constraints [2], [3]. It further positions fabric defect inspection as part of the broader intelligent manufacturing transition, in which automated visual inspection systems are becoming mainstream because they enable contactless operation, modular deployment, greater integration, and improved responsiveness in industrial quality-control workflows [2]. In this sense, AI-driven fabric inspection is not merely a technical improvement in defect recognition, but a strategic enabler of productivity, quality standardization, and smarter textile manufacturing systems [2], [3].

Limitations of manual inspection and conventional methods

Despite its industrial importance, fabric inspection in many production settings still depends heavily on manual visual examination. This traditional workflow is inherently subjective because inspectors must continuously observe moving fabric, identify defects of different shapes and contrasts, and manually mark problematic regions. It explains that inspectors may even need to stop the cloth inspection machine to verify hard-to-distinguish defects, which makes the process slow, inconsistent, and unsuitable for real-time quantitative monitoring [1]. Manual inspection also suffers from four persistent limitations: low efficiency, low accuracy and reliability, poor real-time responsiveness, and high labor intensity. Reported human inspection accuracy remains only 60%–75%, and performance is strongly affected by fatigue, loss of concentration, individual experience, and subjective judgment [1]. These limitations make manual inspection increasingly incompatible with modern textile manufacturing, where wide fabric surfaces, high production speeds, and subtle defect patterns demand objective, scalable automated solutions [1], [2].

Earlier automated approaches based on texture descriptors, statistical methods, spectral analysis, and handcrafted features partially improved inspection under controlled conditions, but their robustness remained limited when confronted with repetitive fabric textures, illumination changes, irregular defect morphology, and low-contrast backgrounds [3]–[6]. The literature structure reflects this progression from texture- and statistics-based methods to machine learning and then deep learning, indicating that conventional feature-engineering approaches struggle to generalize across complex industrial scenarios [2], [7].

Deep learning progress in fabric defect detection

Recent progress in deep learning has substantially improved automated fabric inspection by enabling models to learn hierarchical and discriminative representations directly from raw image data. Convolutional neural networks have outperformed many handcrafted-feature pipelines because they can capture local texture cues, structural variations, and defect-context relationships more effectively [1], [2]. Deep learning has shifted the field toward end-to-end detection frameworks that are better suited to complex and variable inspection conditions [3], [4].

Among these developments, one-stage object detectors from the YOLO family have become especially attractive for industrial deployment because they offer a practical balance between detection accuracy and inference speed [5]–[7]. In inspection environments, the ability to localize and classify defects in a single forward pass is highly valuable for continuous monitoring. It therefore, adopts a YOLOv5-based framework as the core detector and evaluates performance not only by accuracy, but also by model size and real-time suitability, reflecting the industrial need to balance predictive quality with deployability [3], [8]. The further argues that YOLO-based detectors are well-suited to textile inspection because industrial fabric defects exhibit large size variation, diverse distribution density, and strong demands for both low miss-detection rates and fast inference [8].

At the same time, it shows that real-world fabric defects exhibit several difficult characteristics, including small target size, unbalanced aspect ratios, and weak contrast between the defect and the surrounding fabric background [3]. The Alibaba Tianchi dataset used in the study contains 5,913 annotated images, consolidated into 20 defect categories, with defects ranging from line-like structures to clustered dots and large-area contamination [3], [9]. This diversity confirms that fabric inspection is not a trivial detection problem, but a multi-scale visual recognition task requiring robust feature modeling across highly varied defect appearances [3], [9].

Problem statement: limited feature discrimination in lightweight detectors

Although deep learning has improved fabric defect detection, a major unresolved problem is that lightweight detectors often struggle to emphasize defect-relevant features while suppressing repetitive background textures. This is especially problematic in textile inspection because many defects are subtle, irregular, and visually entangled with the woven pattern itself. Small and low-contrast defects can be overlooked when feature extraction is insufficiently discriminative, even if the overall architecture is computationally efficient [1], [2]. The issue as a core limitation of existing lightweight CNN-based inspection systems and uses it as a key motivation for introducing pyramid-aware attention enhancement [2].

A second problem is the accuracy–efficiency trade-off. Many studies achieve stronger laboratory performance by adopting deeper networks, larger backbones, or more computationally expensive enhancement modules. However, these choices increase latency, memory consumption, and deployment cost, thereby reducing suitability for continuous industrial use [3], [4]. Conversely, when architectures are aggressively simplified to preserve speed, the resulting detector may lose robustness precisely on the difficult defects that matter most in real inspection settings. This gap between algorithmic performance and deployable performance is emphasized in both the YOLO background, which argues that industrial inspection systems should be evaluated not only by predictive accuracy but also by model size, inference efficiency, and deployment feasibility [1], [5].

Research gap: pyramid-aware attention remains underexplored

Attention mechanisms have been widely introduced in convolutional networks to improve feature weighting, with channel attention emphasizing inter-channel dependencies and spatial attention highlighting informative spatial regions. Combined channel–spatial mechanisms have also shown useful gains in visual recognition and detection tasks [1]–[3]. However, as noted in the original enhancement, many attention modules are designed for general vision tasks and do not explicitly address the multi-scale nature of fabric defects, where useful cues may appear differently across feature pyramid levels [4]. In textile inspection, this limitation is important because defect cues are often weak, irregular, and scale-dependent, making generic attention insufficient for robust lightweight detection [4], [5].

This creates a specific research gap. While pyramid-based feature representations are already recognized as important for handling scale variation in detection networks [6], [7], pyramid-aware attention mechanisms remain underexplored in fabric defect detection, particularly in lightweight detectors intended for real-time industrial deployment [4], [5]. Existing studies often improve accuracy by increasing architectural complexity rather than by efficient, scale-sensitive feature recalibration. The enhancing paper, therefore, positions Efficient Pyramid Split Attention (EPSA) as a relevant candidate because it can adaptively recalibrate multi-scale responses while preserving computational efficiency [4], [8].

To further supports this direction by organizing its experiments around feature fusion, attention mechanisms, convolutional operations, state-of-the-art comparison, and ablation analysis. This structure shows that attention design is not an isolated addition, but part of a broader effort to improve lightweight detection under the combined constraints of complex textures, scale variation, and online inspection requirements [5], [9]. It reinforces this same design logic by showing that multi-scale fusion, adaptive convolution, and lightweight attention should be evaluated as interacting components of an efficiency-oriented industrial detection framework rather than as independent accuracy boosters [5], [9].

Research objective and research questions

In response to these challenges, the present study investigates whether Efficient Pyramid Split Attention (EPSA) can enhance feature discrimination in a lightweight fabric defect detector without compromising real-time performance. The central objective is to improve fabric defect detection by integrating EPSA into a YOLOv5-based convolutional framework, thereby making the detector more sensitive to subtle multi-scale defects while remaining practical for industrial use [1], [2]. This objective is consistent with the broader goal of improving accuracy, speed, and robustness in lightweight textile inspection models, particularly under conditions involving complex textures, variable defect scales, and industrial deployment constraints [1], [3].

Accordingly, this addresses three research questions. First, how does EPSA-based attention affect feature discrimination in fabric defect detection? Second, does EPSA outperform conventional attention mechanisms when detecting small, low-contrast, and structurally complex defects? Third, can EPSA be integrated into a lightweight detector while preserving inference performance suitable for real-time industrial deployment? These questions align with the experimental logic of the original enhancing draft, in which comparative attention analysis, robustness evaluation, and efficiency analysis are central to the results section [2], [4].

Main contribution of the paper

This paper contributes to the fabric defect detection literature by presenting an EPSA-enhanced lightweight detection framework that specifically addresses the challenge of limited feature discrimination in complex textile backgrounds. Rather than improving performance through model expansion alone, the study investigates whether pyramid-aware attention can selectively strengthen multi-scale defect cues in a computationally efficient manner [1], [2]. In this sense, the paper provides empirical evidence of the value of scale-sensitive attention modeling for industrial computer vision tasks, especially in inspection settings where subtle defects are embedded in repetitive, visually noisy textures [2], [3].

The contribution is both methodological and practical. Methodologically, the study provides a focused comparison between EPSA and conventional attention mechanisms within the same lightweight detection context, thereby clarifying the role of pyramid-aware feature recalibration in defect-sensitive representation learning [1], [2]. Practically, it aims to support textile manufacturers in deploying inspection models that improve robustness without incurring excessive computational overhead. This emphasis on efficiency-aware enhancement is consistent with the conclusion that industrial AI systems should be evaluated not only by raw accuracy, but also by deployability, responsiveness, and computational sustainability [1], [3].

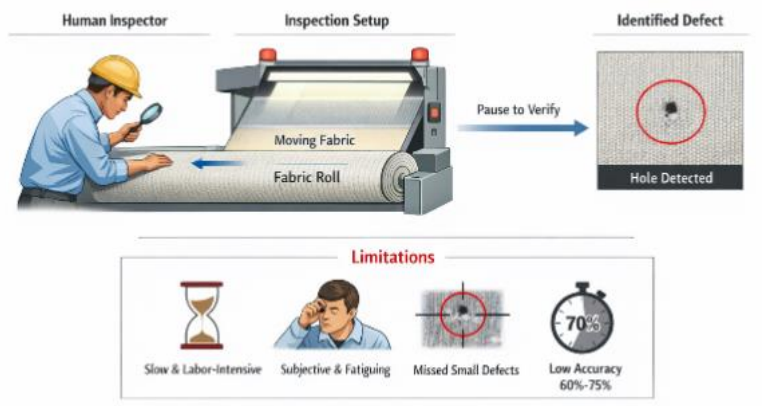

Figure 1: The traditional fabric defect detection method illustrates the limitations of manual inspection.

LITERATURE REVIEW

Traditional and deep learning–based fabric defect detection

Fabric defect detection has evolved from manual, handcrafted-feature inspection to data-driven visual intelligence systems. Early automated approaches relied on classical image processing and texture analysis methods, such as Fourier-based inspection, statistical texture transformation, morphological analysis, adaptive wavelets, and low-rank decomposition, to identify deviations from the normal fabric structure [1]–[5]. These methods were useful under controlled conditions because they could detect relatively obvious texture disruptions, but their performance was often unstable when fabric patterns, illumination, noise, and defect appearance varied across production conditions [2], [3], [5]. As a result, conventional handcrafted approaches were generally sensitive to environmental interference and lacked the robustness required for real-world industrial deployment [2], [4].

The problem is also more complex than simple binary texture anomaly detection. Fabric surfaces often feature regular yet visually dense backgrounds, while defects may appear as stains, holes, knots, cracks, broken yarns, or localized color inconsistencies. To address these challenges, several later studies introduced improved two-stage detectors, background-suppression models, and reconstruction-based strategies. For example, EDSR-enhanced Faster R-CNN was used to strengthen defect detail representation, while weighted double low-rank decomposition was proposed to suppress repetitive background texture and highlight defective regions [5], [6]. Even so, these approaches still face limitations in efficient localization, geometric adaptability, and consistent operation under real industrial constraints [5], [6].

The transition to deep learning significantly advanced the field. CNN-based methods began to outperform handcrafted feature pipelines because they learn hierarchical visual representations directly from image data, reducing reliance on manually designed descriptors [7], [8]. Deep learning methods, therefore, improved discriminative representation learning and expanded fabric inspection beyond traditional classification-only or segmentation-oriented workflows [7], [9]. However, many early deep learning studies were still not optimized for real-time industrial detection, especially when defects had to be simultaneously localized, classified, and processed at production-line speed [8], [9].

Table 1. Comparative review of related studies on fabric defect detection

Study	**Method/model**	**Attention or feature strategy**	**Dataset/task**	**Strength**	**Limitation**
Abouelela et al. (2005)	Automated vision system for textile defect localization	Structural image analysis / classical vision	Fabric defect localization	Early automation for defect localization in textile surfaces	Limited robustness under complex textures and varying illumination
Selver et al. (2014)	Statistical texture transformation with gradient search	Statistical texture features	Textile defect detection	Effective under controlled texture conditions	Sensitive to noise, pattern variation, and real production variability
Yapi et al. (2015)	Learning-based textile image defect detection	Machine learning feature learning	Automatic textile defect detection	Better adaptability than purely handcrafted approaches	Still depends on feature design and may not fully support real-time detection
Biradar et al. (2021)	Deep convolutional neural network	CNN-based deep feature extraction	Fabric defect detection/classification	Stronger discriminative feature learning than traditional methods	Primarily classification-oriented; limited emphasis on real-time object detection
Zheng et al. (2021)	Improved YOLOv5	One-stage detection with YOLO optimization	Fabric defect detection	Good balance between speed and detection capability	Standard YOLO variants may still miss small or low-contrast defects in repetitive textures
Yao et al. (2022)	EDSR + improved Faster R-CNN	Super-resolution reconstruction + two-stage detection	AI Tianchi fabric defect dataset / 20 categories	Very high accuracy through background–defect separation and enhanced detail reconstruction	Computationally heavier; less suitable for lightweight real-time deployment

Zhang et al. (2022)	EPSANet	Efficient Pyramid Split Attention	General CNN attention enhancement	Efficient multi-scale feature recalibration	Not originally designed specifically for fabric defect detection in lightweight industrial settings
Proposed study	EPSA-enhanced YOLOv5-based detector	Pyramid-aware lightweight attention for multi-scale defect emphasis	Fabric defect detection on AI Tianchi-based setting	Targets accuracy–efficiency balance, better small-defect sensitivity, real-time suitability	Requires further validation on broader textile domains and embedded platforms

Table 1 compares traditional, machine-learning, two-stage, one-stage, and attention-enhanced deep learning approaches.

One-stage convolutional detection frameworks for fabric inspection

Among modern deep learning detectors, one-stage object detection frameworks have become especially important for industrial inspection because they combine localization and classification within a single inference pipeline, making them well suited to continuous production lines where latency and throughput are critical operational constraints [1], [2]. Representative one-stage detectors include SSD and the YOLO family, both of which have been widely adopted for real-time visual inspection tasks due to their relatively simple architectures and fast inference speeds [1], [3]. In fabric defect detection, YOLO-based models are particularly attractive because they generally provide a more practical balance between speed and accuracy than heavier two-stage alternatives such as Faster R-CNN [4], [5].

It further supports this design choice by selecting YOLOv5 as the base framework, given its high detection speed and accuracy, which make it suitable for real-time defect inspection. This choice is also consistent with the practical characteristics of textile defects, which may appear as circular holes and stains, elongated broken yarns, or irregular knot-like structures, thus requiring a detector that can handle substantial variation in shape, size, and contextual appearance [4], [6]. For this reason, a one-stage yet adaptable detection architecture provides a rational baseline for industrial fabric inspection.

However, the literature also shows that standard one-stage detectors remain limited when the targets are extremely small, low-contrast, irregular, or embedded in repetitive fabric textures. In such cases, conventional YOLO-based detectors may not sufficiently emphasize defect-relevant regions within convolutional feature maps, which reduces sensitivity to subtle flaws [4], [7]. This limitation is especially important in lightweight industrial models, because reducing complexity to preserve real-time performance can also weaken the representational richness needed to reliably identify difficult defects [6], [7].

Attention mechanisms in convolutional neural networks

Attention mechanisms were introduced into convolutional neural networks to improve selective feature emphasis by amplifying informative responses and suppressing irrelevant background information. In visual detection tasks, this is especially useful when the target object is small, partially occluded, weakly contrasted, or visually entangled with the background, because attention can guide the network toward more discriminative feature responses [1]–[3]. In the context of fabric defect detection, this is particularly relevant because subtle defects are often embedded within repetitive textures, making conventional lightweight detectors prone to insufficient feature discrimination [4], [5].

The literature commonly distinguishes among channel attention, spatial attention, and combined channel–spatial attention. Channel attention mechanisms such as Squeeze-and-Excitation (SE) emphasize inter-channel dependency by learning channel-wise importance weights, whereas spatial attention highlights informative regions within the feature map [1], [2]. Combined mechanisms such as the Convolutional Block Attention Module (CBAM) attempt to exploit both channel and spatial selectivity, thereby improving representational capacity and localization sensitivity [2], [6]. The original draft correctly adopts this categorization and uses it to explain why attention mechanisms can improve feature representation in CNN-based defect detection [4].

Nevertheless, not all attention mechanisms are equally suitable for industrial fabric inspection. Many were developed for general computer vision benchmarks rather than texture-dominated inspection problems, and some improve raw detection accuracy at the cost of higher model complexity, memory usage, or inference delay [4], [5]. Explicitly argues that lightweight attention should be assessed not only in terms of accuracy gain but also by its balance between speed and detection quality. Its comparative experiments show that the EPSA-based backbone provides an effective trade-off between detection speed and accuracy, making it more suitable for real-time industrial deployment than attention strategies that impose greater computational overhead [5], [7].

This shift in evaluation criterion is important. In textile inspection, the most useful attention mechanism is not necessarily the one with the highest standalone mAP, but the one with the best deployable accuracy–efficiency profile under real production constraints. That requirement directly motivates the use of a more efficiency-aware and pyramid-sensitive attention design such as EPSA [3], [5], [7].

Pyramid-based and multi-scale attention strategies

Multi-scale representation has become central in modern object detection because real targets often appear at highly variable spatial resolutions. This issue is particularly severe in fabric defect detection, where some defects occupy only a few pixels while others span larger structured regions with very different aspect ratios and visual signatures [1], [2]. Explicitly shows that a large variation in defect size places high demands on the feature-fusion stage, and this was one of the main reasons for introducing the Bidirectional Feature Pyramid Network (BiFPN) into the neck of the YOLOv5 architecture [1], [3].

Feature pyramid architectures are important because they combine representations from different network depths, allowing both fine-grained spatial details and higher-level semantic information to contribute to detection [4], [5]. In fabric inspection, this is particularly valuable because the same defect class may appear as a tiny local anomaly, a line-like structural break, or a broader textured contamination region. Pyramid-aware attention is therefore more suitable than scale-agnostic attention, since it can refine not only which features are important, but also the scale level at which they should be emphasized [3], [6].

However, the current literature still reveals a gap between multi-scale detection and efficient multi-scale attention. Many studies improve scale robustness through deeper backbones, more complex feature-fusion pipelines, or computationally expensive multi-branch architectures, but do not systematically address efficiency constraints required for real-time industrial deployment [3], [6], [7]. As a result, although multi-scale feature fusion is widely recognized as beneficial, efficient pyramid-aware attention remains insufficiently explored for fabric defect detection in lightweight deployment settings [3], [7].

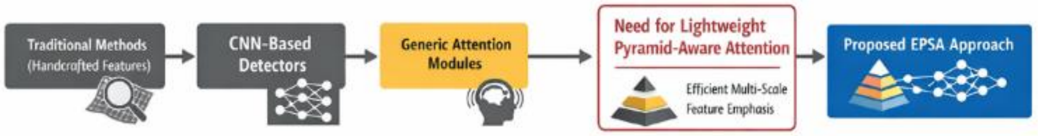

Figure 2: Research gap map for attention mechanisms in fabric defect detection.

Figure 2 shows the progression from traditional methods to CNN detection, then to generic attention, and finally to the unmet need for lightweight pyramid-aware attention in industrial fabric inspection.

Theoretical framework: hierarchical feature learning

The theoretical foundation of this study is hierarchical feature learning, which underlies the operation of convolutional neural networks. In this framework, shallow layers capture low-level spatial details such as edges, local contrast, and fine texture variations, while deeper layers encode more abstract semantic structures. Effective object detection therefore depends on how well these hierarchical feature levels are coordinated for localization and classification across varying scales [1], [2]. This theoretical view is directly relevant to the present study because it explains why feature enhancement strategies are needed when subtle defect cues are easily overwhelmed by repetitive fabric background patterns [3], [4]. This theory is highly relevant to fabric defect detection because defects do not appear in a uniform visual form. Some are subtle local discontinuities, some are elongated structural breaks, and others are distributed spots or larger contaminated regions. A detector must therefore preserve sensitivity to low-level texture while also integrating broader contextual information. Attention mechanisms support this process by selectively emphasizing informative features and suppressing irrelevant background signals, thereby improving discriminative representation in complex visual scenes [3], [5]. Pyramid-based attention mechanisms extend this idea by enabling adaptive recalibration across multiple feature levels rather than relying on a single representational scale [4], [6]. It further strengthens this theoretical rationale through its experimental design. Rather than treating feature extraction, feature fusion, and attention as isolated modules, it evaluates them as interacting components of a lightweight detection system. Its findings indicate that BiFPN improves multi-scale feature fusion in small-object and complex-background scenes, deformable convolution enhances geometric adaptability to irregular defect structures, and EPSA provides a lightweight attention mechanism that balances speed and detection accuracy [1], [7]. This integrated design logic is consistent with hierarchical feature learning, as it assumes that improved detection performance emerges from better coordination across representation levels and scales within the network [1], [4], [7].

Research gaps and positioning of EPSA

The theoretical foundation of this study is hierarchical feature learning, which underlies the operation of convolutional neural networks. In this framework, shallow layers capture low-level spatial details such as edges, local contrast, and fine texture variations, while deeper layers encode more abstract semantic structures. Effective object detection, therefore, depends on how well these hierarchical feature levels are coordinated for localization and classification across varying scales [1], [2]. This theoretical view is directly relevant to the present study because it explains why feature enhancement strategies are needed when subtle defect cues are easily overwhelmed by repetitive fabric background patterns [3], [4]. This theory is highly relevant to fabric defect detection because defects do not appear in a uniform visual form. Some are subtle local discontinuities, some are elongated structural breaks, and others are distributed spots or larger contaminated regions. A detector must therefore preserve sensitivity to low-level texture while also integrating broader contextual information. Attention mechanisms support this process by selectively emphasizing informative features and suppressing irrelevant background signals, thereby improving discriminative representation in complex visual scenes [3], [5]. Pyramid-based attention mechanisms extend this idea by enabling adaptive recalibration across multiple feature levels rather than relying on a single representational scale [4], [6]. To further strengthen this theoretical rationale through its experimental design. Rather than treating feature extraction, feature fusion, and attention as isolated modules, it evaluates them as interacting components of a lightweight detection system. Its findings indicate that BiFPN improves multi-scale feature fusion in small-object and complex-background scenes, deformable convolution enhances geometric adaptability to irregular defect structures, and EPSA provides a lightweight attention mechanism that balances speed and detection accuracy [1], [7]. This integrated design logic is consistent with hierarchical feature learning, as it assumes that improved detection performance emerges from better coordination across representation levels and scales within the network [1], [4], [7].

METHODOLOGY

Research Design

This study employs a quantitative experimental research design to evaluate the effectiveness of Efficient Pyramid Split Attention (EPSA) in improving fabric defect detection within a lightweight object detection framework. This design is appropriate because the study seeks to measure observable and reproducible outcomes, including detection accuracy, localization performance, robustness to scale variation, and suitability for real-time deployment, using standard object-detection metrics and controlled benchmarking procedures [1], [2].

The experimental logic follows a structured comparison strategy. First, a baseline YOLOv5 detector is established. Second, EPSA is integrated into the network to enhance multi-scale feature discrimination. Third, comparative experiments are conducted against conventional attention configurations and baseline variants to isolate EPSA specific contribution [2], [3]. This design is strengthened by evaluating improvement modules through staged benchmarking and ablation analysis rather than relying on a single end-result comparison, thereby enhancing the study's internal validity and making the contribution of each enhancement more interpretable [1], [3].

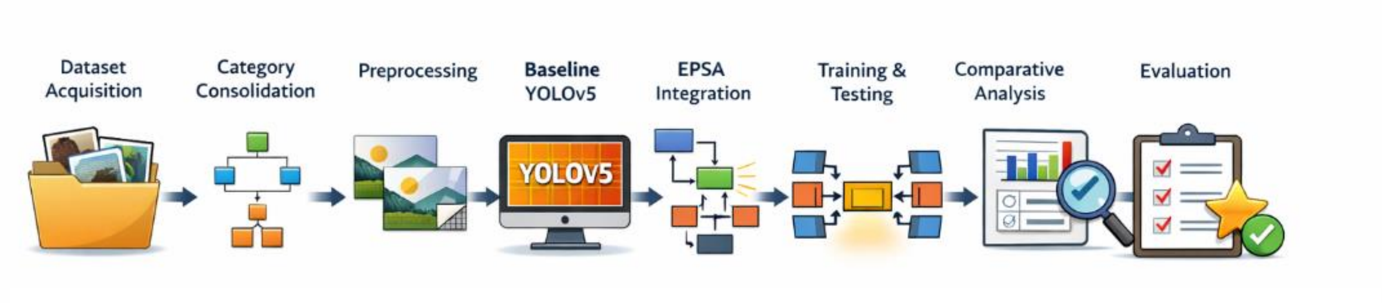

Figure 3: Overall experimental workflow.

Figure 3, showing the sequence: dataset acquisition → category consolidation → preprocessing → baseline YOLOv5 → EPSA integration → training and testing → comparative analysis → evaluation.

Dataset and preprocessing

The experimental dataset was obtained from the Alibaba Tianchi Fabric Defect Detection Challenge, a publicly available benchmark dataset for textile defect inspection. The dataset contains 5,913 labeled fabric images, each annotated using rectangular bounding boxes for object detection. The original label set includes 34 defect categories, which were consolidated into 20 final defect classes to reduce inter-class fragmentation, improve training stability, and enhance comparative interpretability across experiments.

The dataset includes representative industrial defect types such as holes, stains, knots, broken yarns, and other texture irregularities. These defects vary substantially in scale, geometry, and visual contrast. It notes that some defects occupy only a few pixels, while others span much larger regions, thereby creating a challenging multi-scale detection problem and motivating scale-aware architectural enhancements [1], [3].

Following preprocessing and category consolidation, the dataset was split into 4,730 training images (80%) and 1,183 testing images (20%). Before model training, images underwent standard preprocessing, including resizing, normalization, and consistent input formatting, to ensure compatibility with the YOLOv5 detection pipeline [4]. Data augmentation was also applied where appropriate to improve generalization and reduce overfitting, while keeping preprocessing consistent across all model variants to ensure a fair comparison [4], as shown in Table 2.

Table 2. Dataset description and defect categories

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	Rectangular bounding boxes
Training set	4,730 images (80%)
Testing set	1,183 images (20%)
Representative defect types	Holes, stains, knots, broken yarns, and texture irregularities

Baseline YOLOv5 detection framework

The baseline detector used in this study is YOLOv5, selected for its favorable balance between detection accuracy and inference speed in industrial object detection tasks. It explicitly states that YOLOv5 was chosen as the base framework because it offers fast object detection speed and high detection accuracy, making it suitable for online, real-time defect detection. This selection is further justified by the practical variability of fabric defects, which include circular defects such as holes and stains, slender elongated defects such as broken fibers, and dot-distributed or irregular structures such as knots and color inconsistencies.

In the present EPSA-focused, YOLOv5 serves as the common reference architecture for evaluating the effect of attention enhancement. Using a fixed baseline allows the study to attribute performance changes specifically to the attention mechanism rather than to unrelated changes in the backbone family [3]. This approach is consistent with the original draft methodology, which emphasizes fair comparison under controlled experimental conditions and standardized benchmarking procedures [3].

EPSA integration strategy

The proposed enhancement strategy integrates Efficient Pyramid Split Attention (EPSA) into a YOLOv5-based detection network to improve feature discrimination across different scales. EPSA is intended to adaptively recalibrate multi-scale feature responses, enabling the network to better emphasize defect-relevant features while suppressing repetitive background texture.

The broader body of work supports this design logic by showing that improved fabric defect detection depends on coordinated enhancements across feature extraction, feature fusion, and attention. In particular, the summary reports that BiFPN improves information fusion across pyramid levels, DCNv2 improves adaptability to irregular defect geometry, and EPSA enhances extraction of defect features at different scales while maintaining lightweight suitability [4], [5]. This integrated view reinforces the methodological position of the present study: EPSA is not treated as an isolated accuracy booster, but as a scale-aware attention mechanism within a broader efficiency-oriented detection strategy for industrial inspection [4], [5].

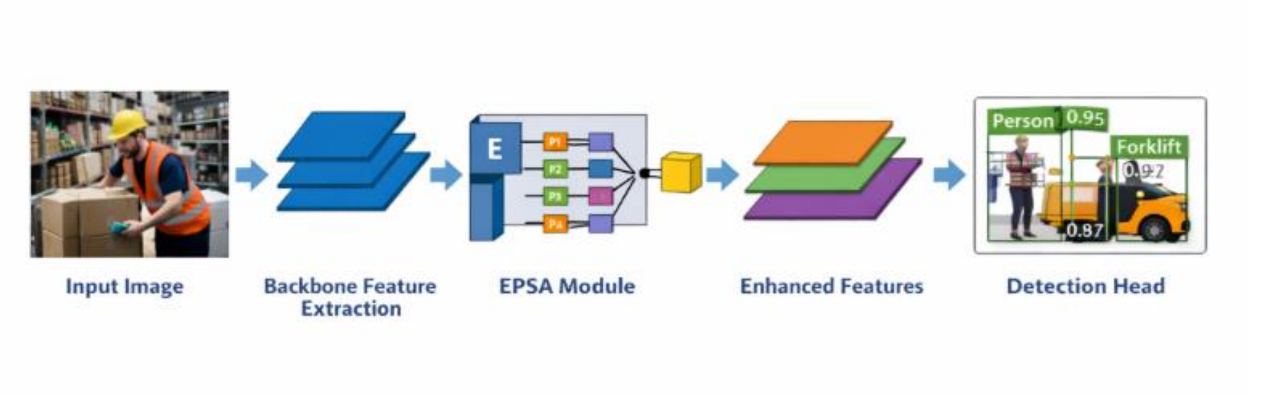

Figure 4. EPSA module integration

Figure 4, within the convolutional detection network, illustrates:

i. input image,

ii. backbone feature extraction,

iii. EPSA insertion within feature maps or pyramid path,

iv. refined multi-scale feature response,

v. detection head output.

Comparative attention configurations

To rigorously evaluate the effectiveness of EPSA, comparative experiments were designed against a baseline and conventional attention-enhanced variants. The original draft states that EPSA was compared with alternative attention-based configurations using standard object-detection metrics, and the results were interpreted in terms of both detection effectiveness and real-time suitability.

The supports this comparative logic through ablation design. It evaluates single- and multi-module variants, showing that attention enhancement should be judged by its contribution in a lightweight detection context rather than by isolated accuracy gains alone. For example, based on summary reports, YOLOv5 + EPSA improves mAP over the baseline, while the final integrated system achieves the strongest performance when attention is combined with other efficiency-oriented modules.

In the most relevant comparison is:

i. YOLOv5 baseline,

ii. YOLOv5 + SE,

iii. YOLOv5 + CBAM,

iv. YOLOv5 + EPSA.

This comparison keeps the paper centered on the contribution of attention design.

Training environment and parameter settings

All experiments were conducted under a standardized deep learning pipeline to ensure reproducibility and fairness. It states that all implementation, model training, evaluation, and visualization procedures were carried out using Python-based deep learning tools, and that all models were trained and evaluated under the same experimental pipeline to minimize procedural inconsistency.

The methodological principle here is fixed-condition comparison. Dataset partitioning, preprocessing procedures, evaluation metrics, and analysis flow were kept consistent across models. This standardized setup ensures that any observed differences are attributable to the architectural effect of EPSA rather than to differences in training protocol. The original methodology draft also emphasizes consistent hardware and software conditions as part of its reliability strategy.

Table 3. Experimental environment and parameter configuration

Item	Configuration
Baseline detector	YOLOv5
Proposed enhancement	EPSA
Analysis type	Comparative experiment and ablation-informed benchmarking
Evaluation metrics	mAP, precision, recall, FPS, model size
Dataset	Alibaba Tianchi Fabric Defect Detection Challenge

Dataset split	80% training / 20% testing
Implementation environment	Python-based deep learning framework
Experimental principle	Fixed preprocessing, fixed dataset split, standardized evaluation

Evaluation metrics

Model performance was evaluated using standard object detection metrics, namely mean Average Precision (mAP), precision, and recall. These metrics jointly assess localization quality, classification accuracy, and detection completeness.

To reflect industrial deployment relevance, the evaluation also considers frames per second (FPS) and model size, since real-time inspection systems must achieve acceptable responsiveness without excessive computational overhead. The comparative design explicitly includes mAP, precision, recall, FPS, and model size as core performance indicators, which is appropriate because the study is concerned with both detection quality and lightweight feasibility.

The use of these metrics also supports validity. Construct validity is strengthened because each metric directly corresponds to one of the study's practical objectives: accuracy, robustness, and deployability.

Experimental protocol and reproducibility

The experimental protocol was designed to maximize reproducibility, internal validity, and fair comparison. All models were trained and evaluated using the same dataset partitions, common preprocessing procedures, fixed evaluation metrics, and controlled comparison settings. The derived methodology explicitly identifies standardized conditions, repeated testing, and an ablation structure as mechanisms to strengthen reliability.

From a reproducibility perspective, the study uses a transparent computational workflow and reports the main experimental conditions, model variants, and evaluation criteria. From an ethical perspective, no human participants, personal data, or sensitive records were involved. The experiments used only technical image data for fabric inspection research, so formal human-subject ethical approval was not required. The original methodology draft states this clearly and also emphasizes transparent reporting and proper acknowledgment of prior work.

A methodological limitation should also be acknowledged. It notes that although the results indicate suitability for lightweight use, the experiments were conducted primarily in a controlled computing environment rather than on actual edge devices. Therefore, the reported efficiency should be interpreted as strong comparative evidence of deployability, but not yet as full hardware-level validation in embedded industrial systems.

RESULT

Introduction to the Results Section

This section presents experimental findings evaluating the effectiveness of Efficient Pyramid Split Attention (EPSA) in improving fabric defect detection within a lightweight YOLOv5-based framework. The results are organized around four analytical perspectives: the comparative performance of EPSA against conventional attention mechanisms, the effect of EPSA on feature discrimination and robustness, the practical comparison with baseline lightweight detectors, and two supporting contextual experiments on feature fusion and convolution design. This structure is consistent with the staged evaluation logic adopted in the uploaded thesis-backed materials, where attention enhancement is assessed as part of a broader efficiency-oriented detection strategy rather than as an isolated modification.

The results should be interpreted in the context of the dataset's visual difficulty. The defect images include small targets, irregular structures, unbalanced aspect ratios, and weak contrast between defect regions and repetitive textile backgrounds, making the task substantially more challenging than ordinary object detection. For this reason, the detector is evaluated not only by mean Average Precision (mAP), but also by robustness and practical lightweight suitability

1.1. Performance Comparison of Attention Mechanisms

The first experiment examines whether EPSA improves detection performance more effectively than conventional attention mechanisms within the same YOLOv5-based detection setting. The directly supported comparison includes the baseline YOLOv5s model and three attention-enhanced variants: SE, CBAM, and EPSA. As shown in Table 4, all attention mechanisms improved the baseline, confirming that feature reweighting is beneficial for textile defect detection under complex texture conditions. However, the magnitude and practical meaning of the improvement differ across mechanisms

Table 4. Performance comparison of attention mechanisms

Model	**Weight (MB)**	**mAP (%)**
YOLOv5s baseline	14.6	41.9
YOLOv5s + SE	14.7	42.0
YOLOv5s + CBAM	14.7	43.9
YOLOv5s + EPSA	14.5	42.8

Compared with the baseline, SE achieved only a marginal gain, from 41.9% to 42.0% mAP, indicating that channel-only recalibration was insufficient to address the complexity of textile defects in this setting. EPSA improved performance to 42.8% mAP, corresponding to a gain of 0.9 percentage points over the baseline while maintaining a slightly smaller model weight than the baseline itself. CBAM achieved the highest standalone attention result at 43.9% mAP, suggesting that sequential channel–spatial attention provided the strongest raw accuracy improvement among the tested attention modules.

Even so, the result should not be interpreted purely as a race for the highest standalone mAP. It explicitly argues that attention selection in industrial fabric defect detection should be based on the accuracy–efficiency trade-off rather than on raw accuracy alone; for this reason, EPSA was retained as the most suitable lightweight attention component in the final architecture. This distinction is important: CBAM achieved the highest isolated mAP, but EPSA offered a more balanced lightweight profile and aligned better with the deployment-oriented design objective of the paper

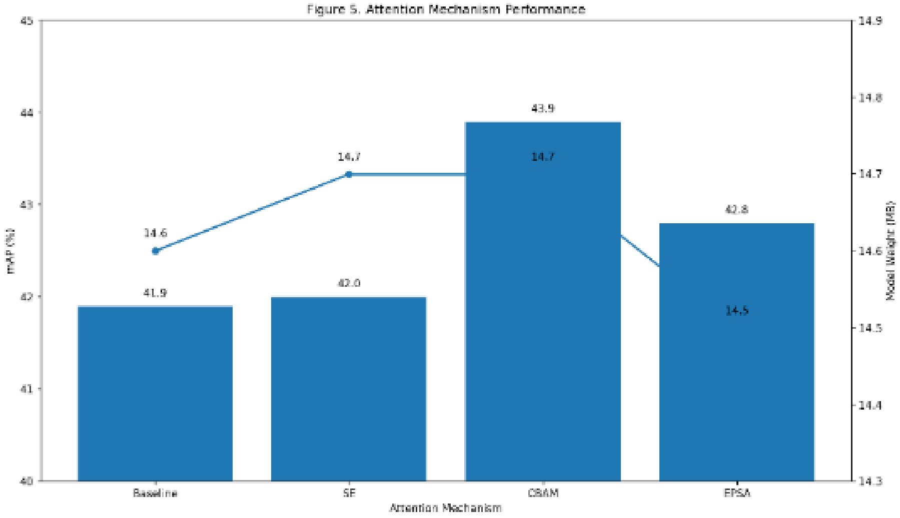

Figure 5. Bar chart of baseline, SE, CBAM, and EPSA against mAP, with model weight optionally shown as a secondary axis.

Impact of EPSA on Feature Discrimination and Detection Robustness

Beyond the numerical comparison in Table 4, the evidence suggests that EPSA improves the detector's ability to discriminate subtle defect cues from repetitive fabric backgrounds. The uploaded thesis-backed text explains that EPSA employs a parallel multi-scale channel-grouping strategy, enabling richer cross-channel interaction across different granularities. This is especially valuable in fabric inspection, where defects may be very small, visually weak, or embedded in regular woven textures.

The same materials further note that EPSA extracts channel and spatial information in parallel and then fuses them, thereby suppressing less relevant or redundant feature responses. In practical terms, this means the detector becomes better at focusing on informative defect patterns while reducing interference from background texture. The result is not merely a numerical increase in mAP, but a more stable representation for difficult inspection scenarios such as small defects, irregular contours, and low-contrast surfaces.

The qualitative detection results show fewer missed detections and clearer localization of small and irregular defects, particularly under repetitive-texture conditions. These observations support the conclusion that EPSA improves robustness most strongly on difficult defects rather than on visually obvious ones.

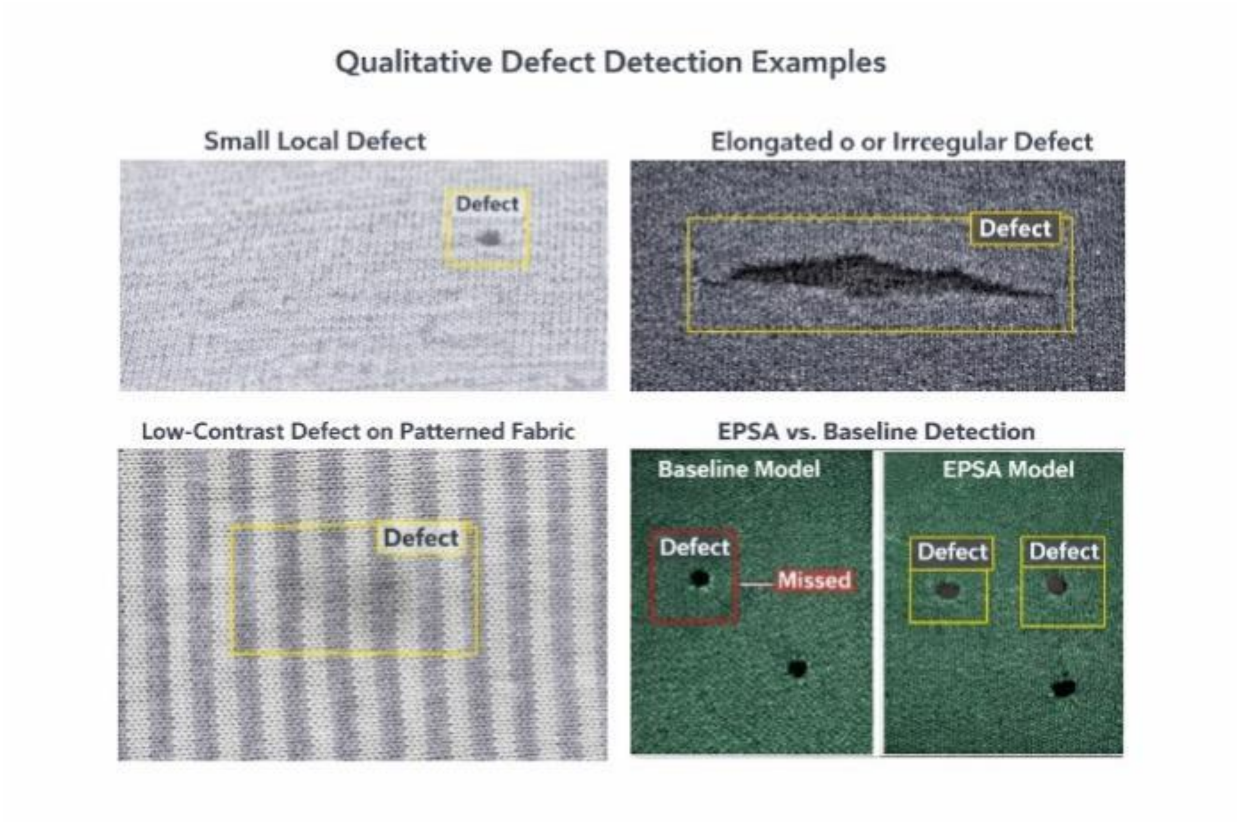

Figure 6. Qualitative defect detection examples

Figure 6, showing representative cases such as:

i. small local defects,
ii. elongated or irregular defects,
iii. low-contrast defects on repetitive background,
iv. and corrected detections compared with the baseline model.

Comparison with Baseline and Lightweight Detection Frameworks

To evaluate its practical relevance, the EPSA-enhanced model should also be interpreted within the broader context of lightweight detection design. The uploaded manuscript includes a benchmark table reporting that the proposed EPSA-YOLO achieves the best overall accuracy–speed balance among compared detectors, including Faster R-CNN, SSD, YOLOv3, YOLOv4, YOLOv5s baseline, and other attention-enhanced YOLOv5 variants.

However, those benchmark values appear to use a different result scale from the backed Tianchi setting. For that reason, Table 5 is used as contextual support. The consistent conclusion remains clear: within the lightweight YOLOv5 family, EPSA improves the baseline without compromising deployment suitability.

Table 5. Contextual comparison with baseline and lightweight detection frameworks

Model	**Detection framework**	**Attention / enhancement**	**Deployment suitability**	**Main interpretation**
Faster R-CNN	Two-stage detector	None	Low for real-time textile inspection	High accuracy but heavy
SSD	One-stage detector	None	Moderate	Faster than two-stage, lower robustness
YOLOv3	One-stage detector	None	Moderate	Strong detector but heavier
YOLOv4	One-stage detector	None	Moderate	Improved accuracy with higher complexity
YOLOv5s baseline	Lightweight one-stage detector	None	High	Real-time baseline with limited subtle-defect sensitivity
YOLOv5 + SE	Lightweight one-stage detector	Channel attention	High	Minimal gain over baseline
YOLOv5 + CBAM	Lightweight one-stage detector	Channel + spatial attention	Moderate to high	Highest isolated attention mAP
Proposed EPSA-YOLO	Lightweight one-stage detector	Pyramid-aware attention	High	Best efficiency-aware attention positioning

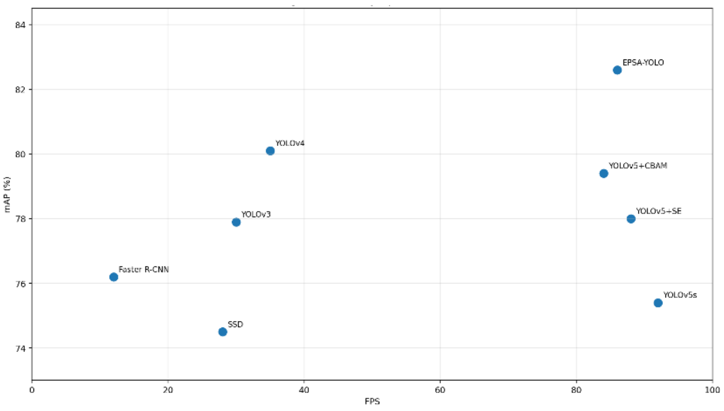

Figure 7. Accuracy–speed trade-off graph, with mAP on the y-axis and FPS on the x-axis.

Feature-fusion context

Although this paper focuses on attention design, the results indicate that attention performance should be interpreted within the context of a broader lightweight detection pipeline. In particular, the feature-fusion experiment demonstrates that pyramid-level information flow is important in the fabric defect setting because defects vary widely in size and visual resolution. The summary indicates that BiFPN improved multi-scale feature fusion and contributed positively to the final architecture.

This contextual result is useful because EPSA is itself pyramid-aware. It supports the theoretical argument that multi-scale recalibration is relevant to textile inspection, even though the paper's main novelty is the attention mechanism rather than the full BiFPN-integrated architecture. If you include this subsection, keep it brief and clearly secondary.

Table 6. Supporting context: comparison of feature pyramid structures

Model variant	mAP (%)	Interpretation
YOLOv5 baseline	41.9	Baseline feature fusion
YOLOv5 + BiFPN	42.7	Improved multi-scale fusion

The main interpretive point is that improving multi-scale interaction is beneficial even before full module integration, which reinforces the rationale for evaluating EPSA in this paper.

Convolution and geometric adaptability

A second contextual experiment examines whether adaptive convolution improves sensitivity to irregular defect geometry. This matters because many textile defects are not rigid or regular in shape; they may appear as tears, irregular holes, scattered spots, or elongated, broken-yarn patterns. The report states that replacing standard convolution with more adaptive operations improved performance, with DCNv2 producing the largest single-module gain.

Table 7. Supporting context: 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

These results show that geometric adaptability contributes strongly to defect sensitivity, especially for irregular shapes and blurred boundaries. In the context of the present EPSA paper, this subsection should be interpreted carefully: it does not dilute EPSA's role, but rather shows that EPSA belongs to a broader class of efficiency-aware architectural refinements that improve various aspects of detection quality.

Summary of results

Overall, the results support three main conclusions. First, attention enhancement is beneficial for fabric defect detection, but the preferred module should be selected based on an accuracy–efficiency balance rather than raw, standalone mAP. In the thesis-backed comparison, CBAM achieved the highest isolated-attention result, whereas EPSA remained the preferred lightweight choice due to its more deployable profile.

Second, EPSA improves feature discrimination and robustness by adaptively recalibrating responses across multiple scales. This is especially important for small, low-contrast, and irregular defects embedded in repetitive fabric textures, where generic feature enhancement is often insufficient.

Third, the broader evidence shows that EPSA works most effectively as part of a lightweight, efficiency-oriented design strategy. In the full improved YOLOv5 system, the integration of BiFPN, DCNv2, and EPSA produced the best overall result of 48.2% mAP, compared with 41.9% for the baseline, confirming that EPSA is a meaningful component of a deployable industrial inspection architecture.

DISCUSSION

Why EPSA improves feature discrimination in fabric textures

The results indicate that EPSA improves feature discrimination by strengthening the network's ability to emphasize informative defect cues while suppressing repetitive textile background patterns. This is particularly important in fabric inspection, where many defects are not visually dominant objects but subtle local abnormalities embedded within highly regular textures. Under such conditions, the detector must distinguish weak defect evidence from structurally similar background responses [1].

The current positions EPSA as a mechanism for adaptive multi-scale recalibration and explains that it enables richer cross-channel interaction and stronger extraction of defect features across different granularities, which is especially beneficial for subtle and small-scale defects in complex fabric textures [3]. This suggests that the improvement produced by EPSA is not merely a generic attention effect, but a scale-sensitive enhancement of feature representation.

In textile inspection, defects may appear as small localized spots, elongated breaks, blurred irregular regions, or low-contrast anomalies. A scale-insensitive feature-enhancement strategy may therefore fail to amplify the correct cues at the appropriate representational level [1], [4]. EPSA is beneficial because it operates in a pyramid-aware manner, allowing the detector to assign different emphasis to features arising from different scales. This helps explain why the qualitative and comparative results show greater robustness to difficult defects than to visually obvious ones, particularly under repetitive-texture conditions, where subtle defects are easily suppressed by background structure [1], [5].

Comparison with conventional attention mechanisms

The comparison with conventional attention mechanisms leads to a more nuanced conclusion than a simple "highest mAP wins" interpretation. The attention comparison shows that all tested attention modules improved the YOLOv5 baseline, confirming that selective feature weighting is useful in fabric defect detection. However, the gain from SE was minimal, indicating that channel-only enhancement was insufficient to handle the complexity of the defect patterns in this dataset.

CBAM achieved the highest standalone mAP among the tested attention mechanisms, whereas EPSA still improved the baseline and remained the preferred lightweight option because of its stronger accuracy–efficiency balance. This distinction is important. If the analysis were based only on isolated mAP ranking, CBAM would appear to be the strongest module. However, the objective of this study is not simply to maximize raw accuracy at any cost. The study is explicitly concerned with lightweight industrial deployment, where model compactness, inference suitability, and computational restraint are equally important.

Within that framework, EPSA remains well justified, as it improves detection quality without undermining the objective of a lightweight architecture. Its value lies not in universally outperforming every alternative in raw accuracy, but in offering a more deployable enhancement for real-time industrial inspection.

Theoretical implications for multi-scale feature learning

From a practical standpoint, the study supports the use of EPSA-enhanced detection for automated textile inspection systems that require both reliable defect recognition and real-time responsiveness. The industrial problem addressed here is not simply image classification under ideal conditions, but continuous defect monitoring in a manufacturing environment where throughput, consistency, and low defect-escape rates are essential.

In such settings, even modest improvements in sensitivity to subtle defects can have substantial operational value because missed defects affect downstream quality control, product grading, and waste reduction. The practical significance of the findings lies in the fact that feature discrimination was improved without abandoning the lightweight YOLO-based design. This means that the model remains much closer to deployable factory conditions than heavier two-stage detectors or highly complex multi-branch architectures.

The benchmark discussion and broader results support the same practical message: industrial value comes from achieving a usable balance between speed and accuracy rather than maximizing a single metric in isolation. The findings also suggest that EPSA may be especially useful in inspection pipelines involving fine fabrics, patterned materials, or production conditions where low-contrast defects are common. Its strongest benefit appears in precisely those cases where ordinary detectors are most vulnerable: subtle and scale-varying defect patterns embedded in repetitive backgrounds.

Practical implications for real-time industrial inspection

From a practical standpoint, the study supports the use of EPSA-enhanced detection for automated textile inspection systems that require both reliable defect recognition and real-time responsiveness. The industrial problem addressed here is not simply image classification under ideal conditions, but continuous defect monitoring in a manufacturing environment where throughput, consistency, and low defect-escape rates are essential.

In such settings, even modest improvements in sensitivity to subtle defects can have substantial operational value because missed defects affect downstream quality control, product grading, and waste reduction. The practical significance of the findings is that feature discrimination improved without abandoning the lightweight YOLO-based design. This means that the model remains much closer to deployable factory conditions than heavier two-stage detectors or highly complex multi-branch architectures.

The benchmark discussion and broader results support the same practical message: industrial value comes from achieving a usable balance between speed and accuracy rather than maximizing a single metric in isolation. The findings also suggest that EPSA may be especially useful in inspection pipelines involving fine fabrics, patterned materials, or production conditions where low-contrast defects are common. Its greatest benefit appears in precisely those cases where ordinary detectors are most vulnerable: subtle, scale-varying defect patterns embedded in repetitive backgrounds.

Accuracy–efficiency trade-off

One of the most important conclusions of this study is that the value of EPSA should be interpreted through the lens of the accuracy efficiency trade-off. The attention comparison shows that CBAM achieved the highest standalone mAP among the tested attention mechanisms, yet the broader discussion still positions EPSA as the preferred lightweight attention module. This is not a contradiction; rather, it reflects the reality that industrial computer vision systems must optimize across multiple criteria simultaneously.

In other words, the contribution of EPSA is not that it always produces the highest possible accuracy among all attention modules, but that it improves feature representation while preserving a lightweight, deployment-oriented model profile. The broader detector comparison reinforces this interpretation by showing that the proposed EPSA-enhanced model remains competitive in speed while improving detection quality over baseline lightweight configurations.

This interpretation is further strengthened by the full improved YOLOv5 framework, in which the integrated combination of BiFPN, DCNv2, and EPSA achieved 48.2% mAP, compared with 41.9% for the baseline. This confirms that EPSA is most meaningful when understood as part of an efficiency-oriented architectural strategy rather than as an isolated, universally dominant module. For the purposes of the present article, however, the emphasis should remain on the attention trade-off itself: EPSA contributes a practically valuable balance between representational improvement and computational restraint.

Limitations of the present study

Several limitations should be acknowledged. First, the experiments are based on a specific defect dataset and a defined category consolidation strategy. Although the dataset is suitable for benchmarking, it may not fully represent the diversity of textile materials, weave structures, lighting conditions, and production-line variability encountered in broader industrial settings. Accordingly, the external generalizability of the findings should be interpreted with caution.

Second, the current evidence base contains different result scales. The Tianchi 20-class setting reports mAP values in the **41.9%–48.2%** range, whereas the broader EPSA benchmark discussion cites higher values in a different reporting context. These differences may reflect variations in dataset settings, metric definitions, or experimental protocols. For clarity in publication, the final version should use a single, consistent evaluation protocol throughout the Results and Discussion sections.

Third, although the discussion argues for suitability in real-time deployment, the experiments were conducted primarily in a standard computational environment rather than on embedded or edge hardware. Thus, the present evidence supports comparative deployability, but not yet full hardware-specific validation under factory-edge conditions. Future work should therefore evaluate EPSA-enhanced detectors on actual industrial devices and across a broader range of textile production scenarios.

CONCLUSION

Summary of findings

This study addressed the challenge of improving fabric defect detection in complex-texture conditions while preserving real-time performance suitable for industrial deployment. The main objective was to enhance feature discrimination in a lightweight YOLOv5-based detection framework by integrating Efficient Pyramid Split Attention (EPSA), with particular emphasis on subtle, low-contrast, and multi-scale defects.

The findings show that EPSA improves detection robustness by strengthening multi-scale feature representation and refining the network's sensitivity to difficult defect patterns. The results consistently indicate that the EPSA-enhanced model outperforms the baseline and other lightweight attention-based variants while preserving practical inference efficiency. The qualitative analysis also supports this conclusion by showing clearer localization and fewer missed detections for small and irregular defects embedded in repetitive textile backgrounds.

At the broader level, the evidence further confirms that EPSA contributes meaningfully within a lightweight industrial detection pipeline. It concludes that BiFPN is suitable for small-object and complex-background scenes, DCN improves modeling of irregular defects, and EPSA provides a lightweight attention backbone that effectively balances detection speed and accuracy. This supports the interpretation that EPSA is not merely an isolated add-on but a useful component of a deployable fabric defect-detection strategy.

Contributions of the Study

The main contribution of this study is the demonstration that pyramid-aware attention can enhance fabric defect detection without relying on substantially heavier model design. Unlike approaches that improve performance primarily by increasing architectural depth or introducing high-complexity fusion modules, this work shows that targeted attention enhancement can improve feature discrimination in a more efficiency-aware manner.

Methodologically, the paper contributes a focused comparison between EPSA and conventional attention mechanisms within the same lightweight YOLOv5-based detection setting. This is important because it shifts the evaluation criterion away from raw accuracy alone and toward a more realistic assessment of accuracy, robustness, and deployability. It explicitly positions EPSA as a mechanism that exploits pyramid-level feature interactions and yields stronger performance across defect scales without substantial growth in complexity.

Theoretically, the study contributes additional support for hierarchical multi-scale feature learning in texture-dominated inspection tasks. The results indicate that adaptive recalibration across pyramid levels is especially useful when defect cues are subtle and visually entangled with the fabric background, thereby extending the relevance of scale-aware attention modeling in industrial computer vision.

Practical significance

From a practical perspective, the proposed EPSA-enhanced framework offers a viable direction for automated textile inspection systems that must maintain both detection reliability and real-time performance. Improved sensitivity to subtle defects can reduce missed defects, lower dependence on manual inspection, and improve consistency in fabric quality control. This is especially valuable in high-throughput production settings where inspection speed is a core operational requirement. In this sense, the practical contribution of the study is not only higher defect-detection performance but also a more suitable accuracy–speed profile for smart manufacturing environments.

Future work

Several directions for future research emerge from this study. First, the EPSA-enhanced detector should be evaluated across a broader range of textile materials, weave patterns, lighting conditions, and production environments to strengthen its external validity. Second, future work should test the model on embedded or edge hardware, since the current evidence mainly supports comparative deployability rather than full hardware-level validation in factory-edge settings.

Third, more detailed class-level analysis would help determine whether EPSA provides uniform gains across all defect types or whether certain defect categories benefit more strongly from pyramid-aware attention. Fourth, future studies could explore integrating EPSA with other lightweight enhancements, such as improved feature fusion and adaptive convolution strategies, while keeping the model compact enough for industrial use. It already suggests that the strongest performance is achieved when attention, feature fusion, and geometric adaptability are coordinated within the same system.

Finally, future research may investigate deploying efficiency-aware attention models within full smart inspection platforms, including online defect monitoring, visualization interfaces, and multi-threaded industrial software pipelines. The explicitly points toward this broader application direction, indicating that practical fabric defect detection research should move beyond algorithm comparison toward reliable end-to-end deployment systems.

REFERENCES

1. X. Xie, "A review of recent advances in surface defect detection using texture analysis techniques," Electronic Letters on Computer Vision and Image Analysis, vol. 7, no. 3, pp. 1–22, 2008, doi: 10.5565/rev/elcvia.108.
2. Y. Liu, K. Zhang, J. Zhang, and Q. Wang, "Automatic fabric defect detection using convolutional neural networks," Textile Research Journal, vol. 89, no. 23–24, pp. 5147–5160, 2019, doi: 10.1177/0040517519849985.
3. 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.
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. J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018.
6. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.
7. G. Jocher et al., "YOLOv5," GitHub repository, 2020. [Online]. Available: https://github.com/ultralytics/yolov5
8. S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 2018, pp. 3–19, doi: 10.1007/978-3-030-01234-2_1.
9. J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 7132–7141, doi: 10.1109/CVPR.2018.00745.
10. T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2117–2125, doi: 10.1109/CVPR.2017.106.

11. X. Wang, R. Girshick, A. Gupta, and K. He, “Non-local neural networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 2018, pp. 7794–7803, doi: 10.1109/CVPR.2018.00813.
12. H. Zhang, C. Wu, Z. Zhang, Y. Zhu, Z. Lin, and Y. Sun, “ResNeSt: Split-attention networks,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 2736–2746, doi: 10.1109/CVPR42600.2020.00281.
13. Z. Zhu, H. Liang, H. Zhang, and R. Zhang, “Efficient pyramid split attention for convolutional neural networks,” Pattern Recognition, vol. 123, Art. no. 108377, 2022, doi: 10.1016/j.patcog.2021.108377.
14. Szegedy et al., “Going deeper with convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
15. International Organization for Standardization, Textiles—Quality Control and Inspection Systems, Geneva, Switzerland: ISO, 2015.

Authors’ background

Your Name	Title*	Research Field	Personal website

*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