



## YMaskNet: A Deep Learning Based Handloom Fabric Defect Detection Technique

Anindita Das<sup>ID</sup>, Aniruddha Deka<sup>\*ID</sup>

Computer Science and Engineering, Assam down town University, Assam 781026, India

Corresponding Author Email: [aniruddha.deka@adtu.in](mailto:aniruddha.deka@adtu.in)

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/jesa.581008>

### ABSTRACT

**Received:** 3 August 2025

**Revised:** 7 September 2025

**Accepted:** 21 September 2025

**Available online:** 31 October 2025

#### **Keywords:**

*handloom fabric, defect detection, deep learning, defective handloom fabric dataset, defect classification*

Quality control is crucial for handloom fabric manufacturers, as undetected defects can cause financial losses and harm their reputation. Traditional inspection methods, achieving only 60–75% accuracy, are inadequate for maintaining high-quality standards. To overcome this limitation, we propose YMaskNet, an automated defect detection system specifically designed for handloom fabrics, operable with or without human supervision. YMaskNet integrates YOLOv4 with a MobileNetV2 backbone, Bidirectional Feature Pyramid Network (BiFPN), and Mask R-CNN to deliver efficient and precise defect identification. MobileNetV2 enables lightweight and effective feature extraction, while BiFPN enhances multi-scale feature fusion, improving the detection of defects of varying sizes. Mask R-CNN adds further precision by segmenting defect regions accurately. Our proprietary YMask dataset supports training and evaluation of the model. Experimental results demonstrate that YMaskNet significantly outperforms traditional methods in both precision and recall while maintaining real-time performance. The system effectively detects small and complex fabric defects, ensuring higher reliability and operational efficiency. Overall, YMaskNet provides a robust, scalable, and automated solution for maintaining product quality in textile manufacturing and strengthening the competitiveness of the handloom industry through intelligent, data-driven quality control.

## 1. INTRODUCTION

The fashion industry is a key contributor to the global economy, with the textile sector constantly evolving to keep up with growing consumer demand and to drive innovation. Fashion trends evolve rapidly, and people across all economic backgrounds invest in clothing to keep pace with changing trends. Consumer preferences shape these trends, attracting diverse audiences to various fabric types and styles. The textile industry generally remains stable and resilient to financial market fluctuations. In India, Assam is a major producer of handloom textiles, deeply rooted in its ethnic heritage. However, the industry faces challenges as it increasingly relies on seasonal tourism rather than steady commercial growth. Like any other market product, fabrics are valued based on their quality, influencing both consumer satisfaction and producer profitability. While manufacturers aim for high-quality production in minimal time, handloom textiles are prone to defects caused by human errors, mechanical issues, or material inconsistencies [1]. Manual defect detection is still the most commonly used method, but it is labor-intensive and prone to errors, with inspectors accurately identifying only around 70% of defects. In Assam, limited testing facilities further delay production, often exceeding a month. To enhance efficiency and uphold quality standards, the demand for automated defect detection systems in the handloom industry is steadily increasing.

The textile industry has identified over 70 different types of fabric defects. Typically, these defects appear either in the

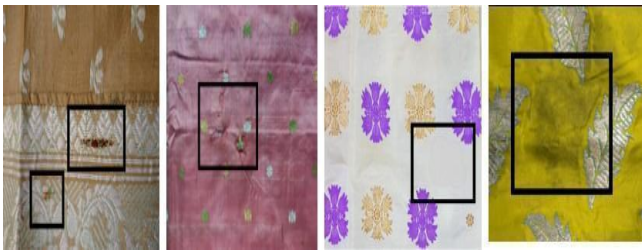
direction of the weave or perpendicular to it. Regarding quality standards, defects in fabrics are classified into two properties: changes in surface color and irregularities in texture. Six common types of fabric defects include float, caused by needle breaks; weft curling, resulting from highly twisted weft threads; slubs, caused by thick yarn spots or fly waste; holes from mechanical faults; stitching defects due to undesired loom movements; and rust stains from lubricants or rust. These significant defects can render the fabric unsellable and lead to revenue loss [2]. Some samples of different defective handloom fabrics are shown in Figure 1. To preserve quality and craftsmanship, the demand for automated fabric defect detection is increasing, ensuring greater efficiency and improved production standards. However, applying deep learning to defect detection in handloom products presents several challenges that must be addressed to enhance accuracy and effectiveness. The distinct textures, intricate patterns, and vibrant colors of handloom fabrics create complexities for conventional deep learning models, making it difficult for them to generalize reliably across different fabric types. Additionally, defects vary in size, with some being minuscule and challenging to detect, necessitating high-resolution image processing. The scarcity of labeled datasets specific to handloom textiles further hampers model training, leading to inaccuracies in defect classification. Computational complexity also poses a challenge, as deep learning models need processing power, making real-time detection difficult. Moreover, an imbalance in defect samples, where certain flaws occur less frequently, can cause biased predictions.

These issues can be addressed by advanced deep learning techniques to enhance defect detection, classification, and processing speed.

Detecting fabric defects is essential for maintaining quality control in textile manufacturing. Traditionally, fabric inspection on production lines relies on workers operating circular knitting machines. They use a central light source to spot defects, pausing the machine whenever an issue is detected. However, this process is physically demanding, time-consuming, and prone to errors, especially as production demands continue to rise. To improve accuracy, researchers have explored various automated visual inspection methods; with defect detection using wavelet transform emerging as a widely adopted approach for analyzing textural features. In this paper, we present our proposed approach:

- We assembled a dataset of handloom fabrics named YMask, categorizing it into five labeled classes: Float, Slub, Stain, Selvedge, and non-defective, to facilitate further research in this area.
- We developed an innovative approach named YMaskNet for detecting defects in handloom fabrics and created a comprehensive benchmark by analyzing our method against other state-of-the-art techniques using evaluation metrics.

In our work, an advanced defect detection model is proposed for handloom fabrics by designing a hybrid YOLO that ensures efficient processing for multi-scale feature extraction, significantly improving detection accuracy. The inclusion of masking enables precise defect segmentation, making it highly effective in identifying and classifying textile imperfections. This paper is structured as follows: Section 2 explores existing methods and their associated limitations. Section 3 introduces the proposed methodology, including details on dataset development. Section 4 presents the experimental results along with an analysis comparing them to peer competitors. Section 5 discusses our proposed method along with its limitations. Finally, Section 6 provides a summary of the findings and conclusions.



**Figure 1.** Samples of different defects in Handloom Fabrics

## 2. LITERATURE REVIEW

A machine vision technique utilizing the Gabor filter for defect detection based on image processing is proposed by Kumar et al. [3], which showed poor detection results for some special defect types. The Wiener filter [4] is employed to classify defective images by converting red-green-blue (RGB) images into a binary format to enhance detection effectiveness. Moreover, various alternative approaches are available for detecting fabric defects. For instance, Yıldız et al. [5] employed a thermal-based defect classification method using the K-nearest neighbor algorithm [6] and dimensionality reduction for textile defect classification. Both image processing and thermal imaging are similarly used in defect

detection, though they mainly address classification issues. With these methods, defects in images are apparent, but they can only be identified, not precisely located.

Song et al. [7] utilized the gray-level co-occurrence matrix (GLCM) to extract key features from images, such as contrast, correlation, cluster shade, and energy. They applied three different backpropagation neural network (BPNN) learning algorithms to assess performance. Their findings revealed that gradient descent with an adaptive learning rate achieved the best results. Traditional image processing algorithms often struggle with complex backgrounds and small object images. Consequently, neural network-based methods are under investigation by some researchers. Ouyang et al. [8] utilized a CNN with an embedded activation layer to detect defects. Li et al. [9] adopted focal loss [10] in ResNet50 [11] to tackle the challenge of imbalanced positive and negative samples. Despite the feasibility of the aforementioned algorithms, some drawbacks remain, such as slow recognition speeds and low precision. Wang et al. [12] introduced an improved non-maximum suppression (NMS) that considers interclass similarities during detection. The variation of Faster R-CNN is widely adopted by researchers to bring efficiency in detecting small targets. Liu et al. [13] introduced a single multi-box detector (SSD) is a notable single-stage detector that has demonstrated strong performance in object detection. Wang et al. [14] introduced an optimized algorithm called the efficient CNN, which replaces the conventional visual geometry group (VGG-16) with a deep residual network for feature extraction. They enhanced the model by improving the feature pyramid module and expanding the number of anchor frames.

To address the issues of uneven defect sample distribution and limited diversity in color fabric defect samples during fabric image acquisition, Zhao and Zhang [15] suggested a method that accurately identifies the location and category of defects in color fabric image data sets, thus improving defect detection. Compared to the previously discussed detection algorithms, Squeeze-and-Excitation Networks (SENet) gained prominence for their effectiveness, securing first place in the ImageNet 2017 classification competition. It's the SE module, which is straightforward to implement, primarily learns channel correlations and filters out channel-specific attention. This model slightly increases computational effort with its superior results. SENet is not a standalone network but a substructure that is incorporated into other classification or detection models. The core concept of SENet is to learn the model by adjusting feature weights based on the loss function, giving more weight to effective feature maps and less weight to ineffective or minor feature maps, thereby achieving better performance. Many reasons arise for preferring deep networks over other traditional ones. First, fabric defects have distinct features compared to common defects. Some defects, such as WEFTS, WARPS, STAINS, FLOATS, and CRACK WEFTS, cover a large portion of the image, while others, like NEPS, HOLES, SNAGS, and KNOTS, are very small, often only a few pixels. Detecting large defects is relatively straightforward, but identifying small defects is challenging, especially when they only occupy a few pixels. Although in a two-stage network [16], a combination of image processing with deep learning is proposed. Proposing a method where image enhancement is performed before using different convolutional networks, resulting in improved accuracy. Some proposed networks, like the region proposal network (RPN) in Faster R-CNN [17], combined with a loss, reduce the weight

of large sample losses and increase the weight of small sample losses.

The precision of a two-stage network is generally higher than that of its one-stage counterpart. However, despite Faster R-CNN being a typical two-stage network, its speed is generally lower compared to other methods. Among the first-order algorithms, YOLO has undergone substantial improvements across multiple generations. YOLOv1 [18] faced various accuracy limitations, which were addressed in YOLOv2 [19] and further enhanced in YOLOv3 [20]. YOLOv4 [21] offers even better performance, incorporating techniques such as mosaic data augmentation, MISH [22] activation function, FPN-net [23], PAN-net [24], SPP-net [25], and CSPdarknet53 [26] as backbones. Additionally, YOLOv4 is accessible for training with a graphics processing unit (GPU), making it advantageous for many researchers and convenient for industry applications. Unlike two-stage algorithms, one-order algorithms like YOLO are capable of meeting real-time detection requirements.

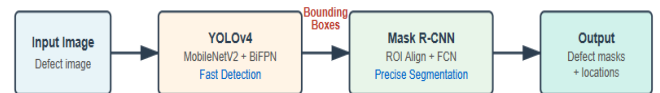
The review examines various machine vision techniques for fabric defect detection alongside deep learning approaches, emphasizing their strengths and limitations. Traditional image processing methods, such as Gabor and Wiener filters, are effective for defect classification but fall short in precise localization. Neural network-based methods, including CNNs and advanced architectures like Faster R-CNN and SENet, significantly enhance detection accuracy but often struggle with speed and computational efficiency. The evolution of YOLO models, particularly YOLOv4, demonstrates promising results in real-time defect detection by integrating advanced feature extraction and optimization techniques. While two-stage networks offer higher precision, they are generally slower compared to one-stage alternatives like YOLO. The fusion of deep learning with advanced attention mechanisms presents a robust solution for detecting both large and small textile defects, focusing mainly on optimizing computational efficiency, improving dataset diversity, and further enhancing detection accuracy and real-world applicability.

### 3. METHODOLOGY

This paper introduces a cost-effective automated system designed to identify defects in handloom fabrics, emphasizing its practical benefits. The development of a fully automated inspection system requires a highly efficient and reliable approach to fabric flaw detection, ensuring superior accuracy and consistency compared to manual inspection methods. Our proposed system follows a well-structured workflow, as in Figure 2, that begins with preprocessing, where high-resolution fabric images undergo enhancement and augmentation to improve diversity. The training setup involves feeding these preprocessed images into the hybrid framework, utilizing backbone networks for efficient feature extraction. The multi-scale feature fusion ensures accurate detection across different defect sizes. Validation strategies include splitting the dataset into training, validation, and testing, and optimizing hyperparameters that reduce overfitting. Finally, model evaluation is conducted using the metrics and inference speed to assess both classification and segmentation accuracy.

The learning networks were trained using a comprehensive dataset of handloom fabrics, carefully standardized to

encompass a wide range of defects. High-quality datasets play a vital role in training and evaluating deep learning models, ensuring accurate and well-balanced outcomes. However, scraping web and e-commerce sites yielded low-quality, low-resolution images, making dedicated data collection necessary. In several defect detection studies, datasets such as the MVTec dataset (MV) [27], TILDA Fabric database (TL) [28], and Fabric Stain dataset (ST) [29] are utilized. These datasets encompass over 540,000 images, featuring a wide range of defect classes across different types of fabrics. Firstly, we utilized a pre-trained network for initial training, but to achieve efficient outputs, original handloom defect samples were essential. Collecting many defect samples with sufficient variation is challenging, and capturing high-resolution images can be resource-intensive. Accurate annotation of each image is also a labor-intensive and time-consuming process. To overcome these limitations, we visited the silk city of Assam, “Sualkuchi,” [30] where we collaborated with weavers to capture around 240 samples of four different flaws and took 60 normal samples, focusing on five specific classes: Float, Slub, Stain, Selvedge, and non-defective. Each defect type was manually annotated by domain experts to maintain high labeling accuracy. High-resolution images were gathered using quality mobile equipment to maintain clarity and detail. To improve the effectiveness of our research, we applied adaptive histogram equalization (AHE) [31] to enhance subtle variations and detect defects that might otherwise be concealed due to inconsistent lighting conditions. This method is especially beneficial for inspecting handloom fabrics, as it minimizes the effects of uneven illumination while maintaining the quality and integrity of the textiles. To maintain dataset balance and prevent bias, an equal distribution of defect types was ensured by augmenting underrepresented categories. We augmented the dataset by processing images through cropping, rotating, and adjusting brightness and contrast to simulate various real-world conditions. Cropping helped simulate partial defects, rotation introduced variations in defect orientation, and brightness adjustments accounted for different lighting conditions. To expand the dataset and improve model performance, data augmentation techniques such as flipping and scaling were applied. These methods help create additional training samples, enhancing the model’s ability to recognize fabric defects under varying conditions. Each image was meticulously annotated with precise labels indicating the defect type, facilitating the training and evaluation of deep learning models. This comprehensive dataset aims to enhance the detection accuracy of proposed automated systems, providing a valuable resource for advancing quality control in the handloom industry. The generated YMask dataset consisted of 4500 samples with 900 in each class. For our proposed approach, we considered 70% and 20% data for training, validation, respectively, and the remaining 10% data for testing. The process of our dataset creation is shown in Figure 3.

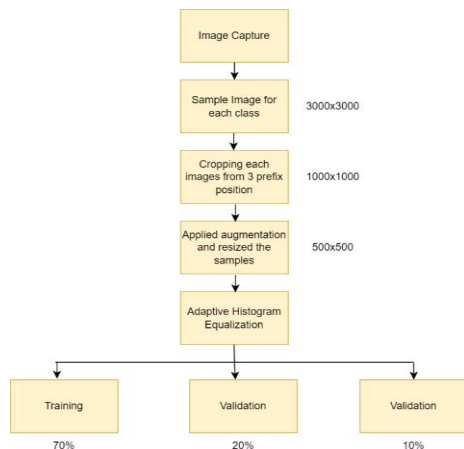


**Figure 2.** Workflow of the proposed network

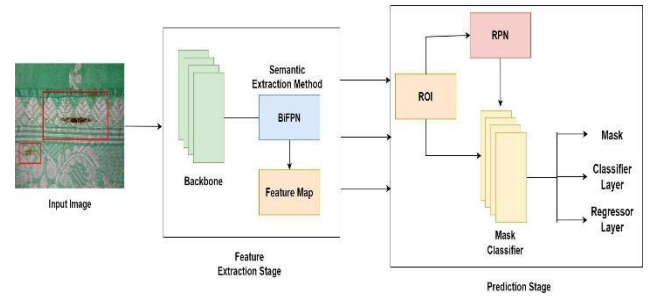
The YOLOv4 network [32] is composed of three key components: the backbone, the neck, and the head, as illustrated in Figure 4. While Darknet-53 [33] serves as the



backbone of YOLOv4, it presents several challenges. One major limitation is its high computational demand, requiring substantial processing power and memory. This makes it less ideal for deployment on resource-limited devices like mobile phones or embedded systems. Additionally, the increased complexity and depth of Darknet-53 result in slower inference times, which can be a significant drawback for real-time applications, where high processing speed is essential for efficient performance and timely defect detection. The high computational requirements also lead to increased energy consumption, posing challenges for battery-powered devices and reducing their operational efficiency. Moreover, implementing and optimizing a model with Darknet-53 for different hardware platforms can be complex, often necessitating specialized hardware or GPU acceleration to achieve acceptable performance. Finally, the extensive depth of Darknet-53 can lead to longer training times, increasing the resources and time required to develop and fine-tune models. To overcome these challenges, replacing the Darknet-53 backbone in YOLOv4 with the lightweight MobileNetV2 [34] significantly reduces the number of network parameters, making the model more efficient. MobileNetV2 utilizes Depthwise Separable Convolution instead of traditional convolution in both the feature extraction and detection head networks. This approach consists of two key operations: depthwise convolution (DWConv) and pointwise convolution (PWConv). The depthwise convolution applies a single convolutional filter to each input channel separately, while the pointwise convolution then combines the extracted features from all channels using a standard convolution operation. This structure enhances computational efficiency while maintaining strong feature extraction capabilities. MobileNetV2 was selected over other lightweight models, such as EfficientNet or ResNet, which are preferred for their ability to maintain a balance between accuracy, computational efficiency, and real-time processing. While other networks provide accuracy through compound scaling, higher computational power, which may not be ideal for resource-limited environments. Some effective deep feature extractors have a larger model size and higher memory consumption. This makes it particularly effective for identifying subtle fabric defects, ensuring fast and accurate detection even on mobile or embedded systems. This approach significantly reduces the number of model calculations without sacrificing accuracy. The complete architecture of our YMaskNet is shown in Figure 5.

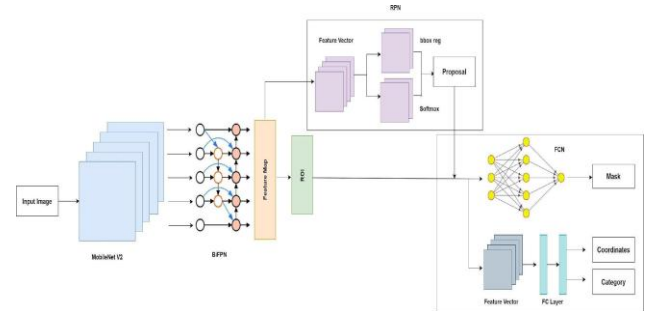


**Figure 3.** Dataset preparation flowchart



**Figure 4.** The framework of YMaskNet

MobileNetV2 introduces an inverted residual module with linear bottleneck blocks, significantly improving accuracy in image classification and detection tasks, especially on mobile devices. Unlike MobileNetV1, which applies rectified linear unit (ReLU) activation functions across all layers, MobileNetV2 strategically avoids using activation functions in the low-dimensional convolutional layers of the bottleneck block. This approach minimizes the loss of crucial feature information by employing a linear bottleneck layer, ensuring that essential details are preserved. The inverted residual module enhances the efficient transfer of multi-layer feature information, strengthening the network's ability to extract meaningful features. It first expands the number of channels using a  $1 \times 1$  convolution, followed by a  $3 \times 3$  depthwise convolution to extract features. Finally, a  $1 \times 1$  pointwise convolution in the linear bottleneck layer reduces the channel dimensions. This design not only optimizes memory usage but also improves computational efficiency while maintaining high accuracy. For defect detection, the pre-trained MobileNetV2, which incorporates a 16-block architecture, is an excellent choice for feature extraction. It provides a well-balanced trade-off between model size, computational efficiency, and detection accuracy, making it particularly suitable for resource-constrained environments.



**Figure 5.** YMaskNet architecture

BiFPN [35] serves as the effective neck in YOLO for object detection by enhancing feature fusion across different scales. The BiFPN architecture improves upon traditional feature pyramid networks (FPNs) [36] by enabling efficient multi-scale feature integration through a bi-directional approach. It seamlessly combines top-down and bottom-up pathways, allowing high-level semantic features to merge with low-level features. This integration enhances the model's ability to capture both abstract patterns and fine-grained details. This network accepts three outputs from the backbone, compared to the five in EfficientDet's BiFPN [37], aligning with the three scale inputs expected by YOLO's neck. Following this, three scale predictions are also used as the output from the neck of YOLO, which are then fused into a single output by

the final fully connected layer. This represents the original implementation of BiFPN, where the scaled predictions are fused within the neck itself. Additionally, it utilizes learnable weighted connections to prioritize the most relevant features during the fusion process, optimizing the network's performance. By effectively combining features at various resolutions, BiFPN enhances the detection accuracy of small objects while maintaining robust performance on larger objects. Its design also emphasizes computational efficiency, making it a suitable choice for real-time applications where speed and accuracy are crucial. This improved feature fusion capability enables YOLO models with BiFPN as the neck to achieve superior detection results, particularly in complex and varied environments. The refined regions of interest (ROIs) are then processed using the RoI Align module, a technique designed to enhance accuracy in object detection and segmentation. Unlike traditional RoI Pooling, which applies quantization, RoI Align eliminates this step by using bilinear interpolation. This approach ensures that pixel values are computed with floating-point precision, making the feature aggregation process continuous rather than discrete. The process begins with RoI Align extracting candidate regions from the feature maps and resizing them into fixed dimensions, typically  $7 \times 7$  and  $14 \times 14$  feature maps, depending on the specific requirements of classification, localization, or mask generation tasks. Unlike conventional methods, RoI Align preserves the original floating-point coordinates of region proposals without rounding them off. These proposals are then divided into  $k \times k$  cells, with each cell further subdivided to determine four key points. The feature values at these points are estimated using bilinear interpolation, followed by a max pooling operation. By converting feature aggregation from a discrete to a continuous process, RoI Align effectively eliminates errors caused by mismatched quantization in traditional RoI Pooling. This results in more accurate and reliable feature maps, ultimately improving the performance of object detection and segmentation models.

Mask R-CNN [38], an advanced version of Faster R-CNN [39], utilizes a two-stage object detection approach. The first stage involves the RPN [40], which generates candidate regions that may contain objects. In the second stage, features are extracted from these proposed regions, leading to classification, localization, and mask generation for each detected object. The feature extraction process in Mask R-CNN follows a dual-path structure: bottom-up and top-down. The bottom-up pathway consists of modules with residual structures of varying sizes, responsible for extracting image features. As the feature maps progress through the network, their resolution is halved at each stage due to the residual structures and stride-2 convolutions, ensuring a hierarchical representation of features. In contrast, the top-down pathway refines these features by incorporating high-level semantic information with finer details. This is achieved by upsampling the high-level feature maps by a factor of two to align them with the resolutions of the bottom-up feature maps. The integration of these features produces new feature maps (P2, P3, P4, and P5), which undergo pixel-wise addition followed by  $3 \times 3$  convolutions. This process enhances the representation of high-level semantic information, improves detection accuracy. Once feature extraction is complete, the generated feature maps are sent to the RPN for further processing. The RPN performs initial classification and bounding box regression to predict the presence of objects and their locations. It generates preliminary regions of interest

(ROIs) by placing anchors on the feature map. A sliding window mechanism scans the feature map, generating  $k$  anchors at each point. To accommodate objects of different sizes, the anchors are defined using three different scales and three aspect ratios, resulting in a total of nine anchors per location. The sliding window then convolves across the feature map, and a fully connected network classifies the anchors and refines their bounding box positions. During output generation, convolutional neural networks (CNNs) define the shape and extent of objects within the image. The RPN assigns objectness scores to each anchor, determining whether it contains an object or belongs to the background. Simultaneously, the regression layer fine-tunes the bounding box coordinates for better localization, leading to more precise object detection and segmentation. Non-Max Suppression (NMS) [41] then removes redundant and overlapping proposals, retaining only the most relevant ones. These refined proposals pass through the RoI Align module and then to the mask classifier. The mask classifier, using CNNs, generates segmentation masks for each region, predicting the objects' precise shapes and extents within the proposals. Each component in the proposed architecture plays a distinct role in enhancing overall model performance. MobileNetV2 acts as a lightweight and efficient backbone, enabling fast and effective feature extraction with reduced computational cost. YOLOv4 contributes to real-time object detection by rapidly identifying potential defect regions in the fabric. BiFPN strengthens the model by allowing bidirectional feature flow and multi-scale feature fusion, improving the detection of small and varied defects. Finally, Mask R-CNN adds precise pixel-level segmentation, allowing accurate localization and shape identification of defects. Together, these components create a balanced and high-performing system for complex handloom fabric defect detection. This multi-step process allows the network to accurately classify objects, precisely locate them, and generate detailed segmentation masks. The final output includes class labels, bounding box coordinates, and segmentation masks for each detected object, providing comprehensive outputs for object detection and segmentation tasks.

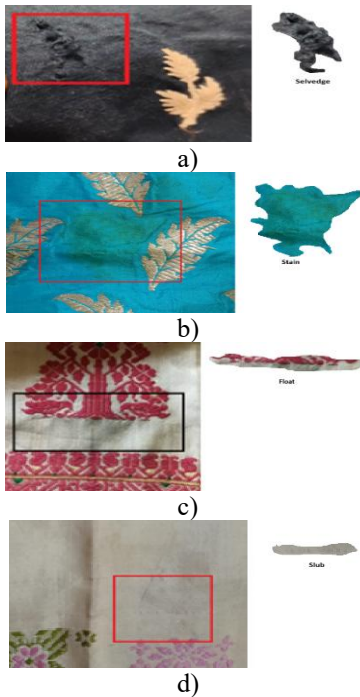
## 4. EXPERIMENTAL RESULTS

In our implementation, we used the NVIDIA GeForce GTX 1650 Max-Q GPU, which offers robust hardware acceleration essential for executing deep learning models like YMaskNet. This setup ensures real-time or near-real-time image processing capabilities. Both the input and output data are 500x500 pixels in size with three color channels in the YMask dataset. The network learns to detect and segment defects within these images, aiding in the visualization and analysis of detection results. This visualization step enables users to easily identify and confirm detected defects. Our system is designed with intuitive software for data annotation, model fine-tuning, and performance evaluation. This well-rounded experimental setup takes advantage of advanced computational power to ensure precise and efficient defect detection in handloom fibers. By improving quality control, it also contributes to preserving the rich heritage and craftsmanship of these valuable textiles.

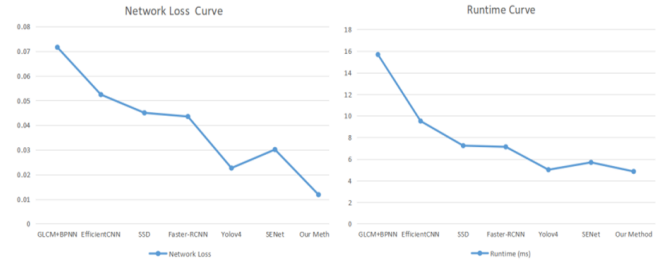
Our defect detection approach in handloom fabrics integrates several advanced networks, such as YOLOv4 with a MobileNetV2 backbone, BiFPN, and Mask R-CNN. This

combination leverages the strengths of each component to achieve superior results. The MobileNetV2 backbone provides efficient feature extraction with a smaller computational footprint, making the model lightweight and fast. BiFPN enhances the feature pyramid by enabling a bidirectional flow of information, thereby improving the model's capability to detect defects across various scales. Mask R-CNN adds a robust segmentation capability, enabling precise localization and identification of defects within the fabric. The model was trained with a learning rate of 0.001, a batch size of 32, and 50 epochs, optimized with the Adam optimizer to ensure stable convergence. Cross-entropy loss was employed for classification, while L2 regularization and dropout techniques were applied to mitigate overfitting.

The configuration for our integrated network was optimized to achieve the highest efficiency and effectiveness. Some of the visual outcomes of various fabric types are shown in Figure 6, and graphs show loss and runtime comparisons for various methods against our proposed approach, as in Figure 7. Our integrated model contains around 26 million parameters and performs at 15 GFLOPs. It achieves real-time inference with a latency of 12.5 ms per image, enabling real-time processing at nearly 80 FPS and uses approximately 850 MB of GPU memory. The network loss function and runtime curves illustrate over 1000 iterations, with the total loss stabilizing at approximately 0.0118 in 4.85 (ms). The model's energy consumption during inference ranged between 17-22 watts, and the end-to-end latency for a batch of eight images was approximately 102 milliseconds. These evaluations show that our model effectively learns from the data. The proposed model achieves an optimal balance between accuracy and inference speed, providing consistently superior overall performance among lightweight models such as YOLOv5s and NanoDet. Additionally, the runtime analysis highlights the superior efficiency of our method compared to others, highlighting its capability to provide fast and precise defect detection.



**Figure 6.** Visual outcomes of our method under varying fabric types



**Figure 7.** Graphical representation of network loss and runtime curve comparison between our approach and other methods

Additionally, comparing our accuracies and losses, we employed the following metrics to evaluate our classification task with the 10% testing dataset: precision, recall, F1-score, and mean average precision (mAP), as in Tables 1-4. Precision measures the model's effectiveness in accurately identifying true positive labels. Recall assesses the model's ability to detect all actual positive instances. The F1-score combines precision and recall into a single metric, providing a balanced measure of the performance. Lastly, the mAP offers a comprehensive evaluation by considering precision and recall across different thresholds. These matrices are evaluated in the case of our proposed approach, considering our generated dataset along with the other 3 available ones. Each of these metrics was meticulously analyzed for all models to ensure a robust solution for our classification problem.

**Table 1.** Comparisons (precision) between our approach and other methods

Method	TL	Precision		
		MV	ST	YMask
GLCM+BPNN [7]	0.5905	0.6466	0.7382	0.7854
EfficientCNN [14]	0.6023	0.6550	0.7842	0.7905
SSD [13]	0.6243	0.5870	0.7522	0.8135
Faster-RCNN [17]	0.6311	0.6820	0.7900	0.8235
YOLOv4 [21]	0.6790	0.6903	0.7842	0.8355
SENet [15]	0.6500	0.6899	0.8090	0.8286
Our Method	0.6886	0.7453	0.8340	0.9125

**Table 2.** Comparisons (recall) between our approach and other methods

Method	TL	Recall		
		MV	ST	YMask
GLCM+BPNN [7]	0.4105	0.4857	0.6786	0.7980
EfficientCNN [14]	0.4456	0.4932	0.7100	0.8032
SSD [13]	0.5475	0.5357	0.7367	0.8133
Faster-RCNN [17]	0.5510	0.5889	0.7786	0.8280
YOLOv4 [21]	0.5475	0.5857	0.6786	0.8980
SENet [15]	0.6090	0.5807	0.6654	0.8560
Our Method	0.6232	0.6721	0.8506	0.9365

Whether the defect area is large or small, our model significantly increases accuracy compared to earlier methods. This holds significant importance for quality assurance in handloom fabrics, where detecting even the smallest defect is crucial. Our method's outstanding performance, compared to other techniques, highlights its potential for real-world applications in maintaining the high quality of handloom textiles. Traditional methods, such as the Gabor filter technique, Wiener filter method, and thermal-based classification methods, struggled with precise defect

localization and often suffered from slow recognition speeds and low precision. Our integrated approach leverages the lightweight and efficient feature extraction capabilities combined with the powerful multi-scale feature of the network.

**Table 3.** Comparisons (F1-score) between our approach and other methods

Method	F1-Score			
	TL	MV	ST	YMask
GLCM+BPNN [7]	0.4857	0.5547	0.7075	0.7917
EfficientCNN [14]	0.5156	0.5635	0.7463	0.7968
SSD [13]	0.6243	0.5870	0.7522	0.8135
Faster-RCNN [17]	0.5900	0.6304	0.7842	0.8257
YOLO [21]	0.6095	0.6345	0.7276	0.8656
SENet [15]	0.6285	0.6312	0.7320	0.8413
Our Method	0.6546	0.7073	0.8392	0.9242

**Table 4.** Comparisons (mAP) between our approach and other methods

Method	mAP			
	TL	MV	ST	YMask
GLCM+BPNN [7]	0.5005	0.5661	0.7084	0.7917
EfficientCNN [14]	0.5239	0.5741	0.7471	0.7969
SSD [13]	0.5859	0.5614	0.7445	0.8134
Faster-RCNN [17]	0.5910	0.6354	0.7843	0.8258
YOLO [21]	0.6133	0.6380	0.7314	0.8668
SENet [15]	0.6295	0.6353	0.7372	0.8382
Our Method	0.6559	0.7087	0.8423	0.9245

Previous neural network-based methods, such as those employing Faster R-CNN and SSD, offered improvements that were still limited by either speed or accuracy in comparison to YOLOv4. Although advancements like the paired-potential activation layer in CNNs and the use of deep residual networks in enhanced Faster R-CNN methods have improved defect detection, our approach surpasses these by integrating the strengths of all four networks. This fusion augments the capability by providing precise defect localization and segmentation, addressing the issues of efficiently identifying small defects that the state-of-the-art methods struggled with. The network also produces some false predictions in handloom fabric defect detection due to small defect regions, color similarities between flaws and fabric, or complex weave patterns. False positives result from textures resembling defects, while false negatives occur with minor flaws that are overlooked. Real-time calibration for varying lighting conditions in fabric textures also affects the network. Some advanced adaptive learning strategies and self-supervised techniques in the near future can help the model better distinguish subtle defects. The proposed method is also an efficient, lightweight network for production with sufficient computing power. However, implementation on low-power or embedded targets might have a negative impact on inference speed and latency. To solve this problem, various computing optimization techniques such as quantization, pruning, and hardware acceleration can be used to ensure the computational load is minimized without sacrificing accuracy. Also, the modular and scalable nature allows YMaskNet to be easily deployed on a centralized as well as edge-based industrial system.

The ablation study demonstrated in Figure 8 shows the impact of each component on the performance of the proposed defect detection network. The full integrated model, combining YOLOv4, MobileNetV2, BiFPN, and Mask R-

CNN, achieves the highest performance with 26 million parameters, 15 GFLOPs, a precision of 0.9125, recall of 0.9365, F1-score of 0.9242, and mAP of 0.9245, while maintaining efficient computation. Replacing MobileNetV2 with a standard CNN backbone increases the parameter count to 32 million and GFLOPs to 20, resulting.

In a reduction in precision (0.8355), recall (0.8980), F1-score (0.8656), and mAP (0.8668), indicating decreased efficiency and slightly lower accuracy. Removing BiFPN reduces the network to 24 million parameters and 13 GFLOPs, lowering recall to 0.8560 and mAP to 0.382 due to diminished multi-scale detection capability. Excluding Mask R-CNN leaves the model with 22 million parameters and 12 GFLOPs, with a precision of 0.944, recall of 0.935, F1-score of 0.8257, and mAP of 0.8258, demonstrating loss of precise segmentation, especially for small defects. Using only YOLOv4 with MobileNetV2, without BiFPN or Mask R-CNN, further reduces performance to 20 million parameters and 10 GFLOPs, with precision of 0.8355, recall of 0.8280, F1-score of 0.8257, and mAP of 0.8258, highlighting the necessity of integrating all components for accurate and efficient defect detection.

Model Configuration	Parameters	GFLOPs	Precision	Recall	F1-Score	mAP
Proposed Model	26	15	0.9125	0.9365	0.9242	0.9245
Replace MobileNetV2 with Standard CNN	32	20	0.8355	0.8980	0.8656	0.8668
Remove BiFPN (FPN only)	24	13	0.8286	0.8560	0.8413	0.8382
Remove Mask R-CNN (Bounding Box Only)	22	12	0.8355	0.8280	0.8257	0.8258
YOLOv4 + MobileNetV2 Only	20	10	0.8355	0.8980	0.8656	0.8668

**Figure 8.** Ablation study of the proposed network

## 5. DISCUSSION

The comparative evaluation, as in tables, of our proposed model against several established methods such as GLCM+BPNN, EfficientCNN, SSD, Faster R-CNN, YOLOv4, and SENet, demonstrates clear and consistent performance improvements across all four fabric types: TL, MV, ST, and YMask. In particular, our method achieves the highest scores in all key metrics, including a precision of 0.9125, recall of 0.9365, F1-score of 0.9242, and mAP of 0.9245 on the YMask dataset, which includes the most visually complex samples. This significant margin of improvement reflects the effectiveness of integrating YOLOv4 with MobileNetV2, BiFPN, and Mask R-CNN. While traditional methods such as GLCM+BPNN and EfficientCNN perform reasonably well on simpler patterns, their inability to generalize across varying textures and thread densities is evident in their lower precision and F1-scores. Even more advanced architectures like SSD and Faster R-CNN, though better in recall and mAP, fall short in maintaining consistent accuracy across different fabric types.

Our architecture's success lies in the synergy between its components: MobileNetV2 provides efficient feature extraction with minimal computational cost, BiFPN enables

enhanced multi-scale feature fusion, and Mask R-CNN ensures accurate pixel-level segmentation. This combination leads to significantly improved detection of small, irregular defects, which are common in handloom fabrics. Competing models, including SENet and YOLOv4, show strength in isolated metrics but lack the balance of precision, recall, and segmentation accuracy that our model delivers. To establish the reliability of our model's performance gains, we calculated 95% confidence intervals for core evaluation metrics across all fabric categories. The results demonstrate that our model consistently outperforms existing methods within statistically significant margins. For instance, the best mAP achieved by our integrated model was 0.9245 on the YMask dataset, with a 95% confidence interval of [0.9181, 0.9302]. Similarly, the highest F1-score observed was 0.9242, with a confidence interval of [0.9185, 0.9298], indicating high robustness in segmentation accuracy. A paired t-test was performed on the per-fold mean of mAP results of YMaskNet against other methods, for five different folds. The results demonstrate that the t-statistic is equal to 26.08 with  $p < 0.001$ , implying that the improvement of the performance of YMaskNet is statistically significant. This validates the robustness and reliability of the proposed model on all of the folds. Quantitatively, YMaskNet obtained better precision and recall than the described U-Net and YOLOv7/YOLOv8 across all tasks, especially in multiple-minute and low contrast defect detection, where U-Net and YOLOv7/8 failed. The integration of BiFPN allowed the improvement of multi-scale feature fusion, which improved the detection of different fabric texture defects, and the Mask R-CNN component offered the accurate segmentation boundaries, which were superior area under the curve (AUC) of defect positions than the DETR. Furthermore, the inclusion of the MobileNetV2 backbone has resulted in greatly reduced computational complexity, so that YMaskNet was able to achieve real-time inference speeds of about 80 frames per second (FPS) with a minute amount of energy consumption. These narrow intervals reinforce the stability and generalizability of our approach, confirming that the performance improvements are both meaningful and reproducible. Furthermore, the high scores across all datasets indicate that the model is suitable for practical deployment in quality control systems within the textile industry. Future improvements could optimize computational efficiency through pruning or quantization to support real-time applications on edge devices and expand the model's industrial applicability.

### 5.1 Limitations

While the proposed method shows promising results in detecting defects in handloom fabrics, there are certain limitations that need to be addressed. One major challenge lies in the computational overhead of the integrated model, particularly due to the inclusion of Mask-RCNN, which may result in slower inference compared to simpler models, especially on resource-constrained devices. Additionally, the model's performance might vary depending on the characteristics of various fabric types, such as varying textures, patterns, or thread densities, which could affect the generalizability of the approach. The dataset used for training may not capture the full diversity of handloom fabrics, which may limit the model's robustness on unseen or highly complex designs. Finally, while the fusion of BiFPN and MobileNetV2 enhances precision, fine-tuning the model to balance accuracy

with speed for real-time applications could require further optimization.

## 6. CONCLUSION

This paper shows that the fusion of YOLOv4 with MobileNetV2 as the backbone, along with BiFPN and Mask RCNN, offers a robust and efficient method for detecting defects in handloom fabrics. MobileNet's lightweight, high-performance framework ensures rapid processing, while BiFPN improves feature extraction at multiple scales, greatly enhancing detection precision. Incorporating Mask RCNN further sharpens the model's effectiveness by allowing for detailed defect segmentation, crucial for identifying and categorizing various textile imperfections. This integrated model effectively tackles the specific challenges associated with detecting flaws in handloom textiles, which often feature complex patterns and textures. The experimental findings demonstrate that this method offers superior accuracy and dependability, but also demonstrate significant potential for practical applications, leading to better quality control and increased efficiency in the textile industry. In the future, this model can focus on improving its real-time performance and scalability for deployment in production environments. Exploring techniques such as model pruning or quantization could help reduce the computational load, making the system more efficient for use on edge devices or low-resource platforms. Additionally, expanding the dataset to include a wider variety of handloom fabrics with more intricate patterns would help improve the model's generalization and robustness. Future research can explore the integration of advanced segmentation techniques or attention mechanisms to enhance complex fabric defect detection and classification, improving both efficiency and industrial applicability.

## REFERENCES

- [1] Roy Maulik, S. (2021). Handloom—The challenges and opportunities. *Handloom Sustainability and Culture: Product Development, Design and Environmental Aspects*, 97-117. [https://doi.org/10.1007/978-981-16-5665-1\\_5](https://doi.org/10.1007/978-981-16-5665-1_5)
- [2] Pande, S. (2022). Problems and prospects of handloom industries: A regional study. *International Journal of Management, Accounting and Economics*, 9(11): 734-748. <https://doi.org/10.5281/zenodo.7433077>
- [3] Kumar, A., Pang, G.K. (2002). Defect detection in textured materials using Gabor filters. *IEEE Transactions on Industry Applications*, 38(2): 425-440. <https://doi.org/10.1109/28.993164>
- [4] Tolba, A.S., Raafat, H.M. (2015). Multiscale image quality measures for defect detection in thin films. *The International Journal of Advanced Manufacturing Technology*, 79(1-4): 113-122. <https://doi.org/10.1007/s00170-014-6758-7>
- [5] Yıldız, K., Buldu, A., Demetgul, M. (2016). A thermal-based defect classification method in textile fabrics with K-nearest neighbor algorithm. *Journal of Industrial Textiles*, 45(5): 780-795. <https://doi.org/10.1177/1528083714555777>
- [6] Shrifan, N.H., Jawad, G.N., Isa, N.A.M., Akbar, M.F. (2020). Microwave nondestructive testing for defect



- detection in composites based on K-means clustering algorithm. *IEEE Access*, 9: 4820-4828. <https://doi.org/10.1109/ACCESS.2020.3048147>
- [7] Song, W., Lang, D., Zhang, J., Zheng, M., Li, X. (2025). Textile defect detection algorithm based on the improved YOLOv8. *IEEE Access*, 13: 11217-11231. <https://doi.org/10.1109/ACCESS.2025.3528771>
- [8] Ouyang, W., Xu, B., Hou, J., Yuan, X. (2019). Fabric defect detection using activation layer embedded convolutional neural network. *IEEE Access*, 7: 70130-70140. <https://doi.org/10.1109/ACCESS.2019.2913620>
- [9] Li, C.J., Qu, Z., Wang, S.Y., Bao, K.H., Wang, S.Y. (2021). A method of defect detection for focal hard samples PCB based on an extended FPN model. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 12(2): 217-227. <https://doi.org/10.1109/TCPMT.2021.3136823>
- [10] Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P. (2017). Focal loss for dense object detection. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2980-2988.
- [11] Mukti, I.Z., Biswas, D. (2019). Transfer learning-based plant diseases detection using ResNet50. In 2019 4th International Conference on Electrical Information and Communication Technology (EICT), Khulna, Bangladesh, pp. 1-6. <https://doi.org/10.1109/EICT48899.2019.9068805>
- [12] Wang, D., Chen, X., Yi, H., Zhao, F. (2019). Improvement of non-maximum suppression in RGB-D object detection. *IEEE Access*, 7: 144134-144143. <https://doi.org/10.1109/ACCESS.2019.2945834>
- [13] Liu, S., Huang, L., Zhao, Y., Wu, X. (2024). Lightweight single shot multi-box detector: A fabric defect detection algorithm incorporating parallel dilated convolution and dual channel attention. *Textile Research Journal*, 94(1-2): 209-224. <https://doi.org/10.1177/00405175231202817>
- [14] Wang, T., Chen, Y., Qiao, M., Snoussi, H. (2018). A fast and robust convolutional neural network-based defect detection model in product quality control. *The International Journal of Advanced Manufacturing Technology*, 94(9): 3465-3471. <https://doi.org/10.1007/s00170-017-0882-0>
- [15] Zhao, H., Zhang, T. (2022). Fabric surface defect detection using SE-SSDNet. *Symmetry*, 14(11): 2373. <https://doi.org/10.3390/sym14112373>
- [16] Wei, X., Yang, Z., Liu, Y., Wei, D., Jia, L., Li, Y. (2019). Railway track fastener defect detection based on image processing and deep learning techniques: A comparative study. *Engineering Applications of Artificial Intelligence*, 80: 66-81. <https://doi.org/10.1016/j.engappai.2019.01.008>
- [17] Ren, S., He, K., Girshick, R., Sun, J. (2016). Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6): 1137-1149. <https://doi.org/10.1109/TPAMI.2016.2577031>
- [18] Ahmad, T., Ma, Y., Yahya, M., Ahmad, B., Nazir, S., Haq, A.U. (2020). Object detection through modified YOLO neural network. *Scientific Programming*, 2020(1): 8403262. <https://doi.org/10.1155/2020/8403262>
- [19] Dong, E., Zhu, Y., Ji, Y., Du, S. (2018). An improved convolution neural network for object detection using YOLOv2. In 2018 IEEE International Conference on Mechatronics and Automation (ICMA), Changchun, China, pp. 1184-1188. <https://doi.org/10.1109/ICMA.2018.8484733>
- [20] Li, C., Wang, R., Li, J., Fei, L. (2019). Face detection based on YOLOv3. *Recent Trends in Intelligent Computing, Communication and Devices: Proceedings of ICCD 2018*, pp. 277-284. [https://doi.org/10.1007/978-981-13-9406-5\\_34](https://doi.org/10.1007/978-981-13-9406-5_34)
- [21] Alhassan, M.A.M., Yilmaz, E. (2025). Evaluating YOLOv4 and YOLOv5 for enhanced object detection in UAV-based surveillance. *Processes*, 13(1): 254. <https://doi.org/10.3390/pr13010254>
- [22] Ren, P., Pan, T.Y., Yang, G., Guo, Y., Wei, W., Pan, Z. (2025). ReLU-like non-monotonic smooth activation functions based on regularized heaviside functions and extensions. *Mathematical Foundations of Computing*, 9: 60-80. <https://doi.org/10.3934/mfc.2025009>
- [23] Gong, Y., Yu, X., Ding, Y., Peng, X., Zhao, J., Han, Z. (2021). Effective fusion factor in FPN for tiny object detection. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1160-1168.
- [24] Yang, J., Fu, X., Hu, Y., Huang, Y., Ding, X., Paisley, J. (2017). PanNet: A deep network architecture for pan-sharpening. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 5449-5457.
- [25] Hua, W., Chen, Q. (2025). A survey of small object detection based on deep learning in aerial images. *Artificial Intelligence Review*, 58(6): 1-67. <https://doi.org/10.1007/s10462-025-11150-9>
- [26] Gouider, C., Seddik, H. (2022). YOLOv4 enhancement with efficient channel recalibration approach in CSPdarknet53. In 2022 IEEE Information Technologies and Smart Industrial Systems (ITSIS), Paris, France, pp. 1-6. <https://doi.org/10.1109/ITSIS56166.2022.10118431>
- [27] Bergmann, P., Fauser, M., Sattlegger, D., Steger, C. (2019). MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9592-9600.
- [28] Wu, J., Le, J., Xiao, Z., Zhang, F., Geng, L., Liu, Y., Wang, W. (2021). Automatic fabric defect detection using a wide-and-light network. *Applied Intelligence*, 51(7): 4945-4961. <https://doi.org/10.1007/s10489-020-02084-6>
- [29] Chandran, L., Bhat, P.N., Kanade, B. (2019). Identification of stain on white fabric and data set generation. In 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT), Kannur, India, pp. 496-500. <https://doi.org/10.1109/ICICT46008.2019.8993243>
- [30] Saikia, J.N. (2011). A study of the Muga silk reeler in the world's biggest Muga weaving cluster-Sualkuchi. *Asian Journal of Research in Business Economics and Management*, 1(3): 257-266.
- [31] Härtinger, P., Steger, C. (2024). Adaptive histogram equalization in constant time. *Journal of Real-Time Image Processing*, 21(3): 93. <https://doi.org/10.1007/s11554-024-01465-1>
- [32] Liu, J., Cui, G., Xiao, C. (2023). A real-time and efficient surface defect detection method based on YOLOv4. *Journal of Real-Time Image Processing*, 20(4): 77. <https://doi.org/10.1007/s11554-023-01333-4>
- [33] JV, N.L., Malwa, K. (2021). Fabric defect detection and classification using YOLOv4. *i-Manager's Journal on*

- Software Engineering, 16(1): 1-14.  
<https://doi.org/10.26634/jse.16.1.18465>
- [34] Dong, K., Zhou, C., Ruan, Y., Li, Y. (2020). MobileNetV2 model for image classification. In 2020 2nd International Conference on Information Technology and Computer Application (ITCA), Guangzhou, China, pp. 476-480.  
<https://doi.org/10.1109/ITCA52113.2020.00106>
- [35] Quang, T.N., Lee, S., Song, B.C. (2021). Object detection using improved bi-directional feature pyramid network. Electronics, 10(6): 746.  
<https://doi.org/10.3390/electronics10060746>
- [36] Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S. (2017). Feature pyramid networks for object detection. Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, pp. 2117-2125.
- [37] Tan, M., Pang, R., Le, Q.V. (2020). Efficientdet: Scalable and efficient object detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10781-10790.
- [38] He, K., Gkioxari, G., Dollár, P., Girshick, R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision, pp. 2961-2969.
- [39] Jiang, H., Learned-Miller, E. (2017). Face detection with the faster R-CNN. In 2017 12th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2017), Washington, USA, pp. 650-657.  
<https://doi.org/10.1109/FG.2017.82>
- [40] Chen, Y.P., Li, Y., Wang, G., Xu, Q. (2018). A multi-strategy region proposal network. Expert Systems with Applications, 113: 1-17.  
<https://doi.org/10.1016/j.eswa.2018.06.043>
- [41] Tyhsen-Smith, L., Petersson, L. (2018). Improving object localization with fitness NMS and bounded IOU loss. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6877-6885.