https://doi.org/10.48047/AFJBS.6.Si4.2024.2501-2521

AFJBS African Journal of Biological Sciences
Journal homepage: http://www.afjbs.com
Research Paper Open Access

Attention Based Shuffle Bi-Directional Feature Pyramid Network Enclosed Yolov8 Model for Cotton Boll Detection and Counting

Uma N Dulhare,
Department of AI and Computer Science and Engineering, Muffakham Jah college of Engineering and Technology (MJCET). Hyderabad, Telangana, India.
Arathi Bairi,
Research Scholar, Department of Computer Science and Engineering, University College of Engineering, Osmania University Hyderabad, Telangana, India.
Email: artibairi@gmail.com
ORCID: https://orcid.org/0009-0003-8430-2021

Corresponding Author: Arathi Bairi

Article History
Volume 6, Issue Si4 2024
Received: 30 May 2024
Accepted : 30 June 2024
Doi:
10.48047/AFJBS.6.Si4.2024.2501-2521

Abstract
Cotton is one of the most widely grown and commercially significant crop globally, which is majorly utilized in textile production. The Cotton boll detection and counting technique is developed based on Machine learning (ML) and Deep Learning (DL) algorithms. There are limited techniques available for detection of cotton boll count, in which most of the techniques poses several difficulties, such as low performance, high time consumption, and computational complexity and so on.To overcome these kinds of issues, the proposed methodology is developed to provide efficient detection technique for cotton boll count. Initially, the data is gathered from dataset and augment the dataset using Geometric distortion and Photometric distortion approaches. Then, Rolling Joint Bilateral Filter (RJBF) is utilized to pre-process the image to enhance the quality of image. Finally, the Attention based Shuffle Bi-directional feature pyramid network enclosed YOLOv8 (ASB-FPNY)model is utilized to extract the feature from a pre-processed image for cotton boll detection and counting. The progressive attention function is utilized to enhance the performance of ASB-FPNY model for cotton boll detection.The accuracy of proposed technique is 98.18% which is higher than other related techniques. The precision achieved is 98.19%,recall is 97.18%,specificity is 98.2%, and F1-score is 97.68% with the proposed model.
Keywords: Cotton boll Detection, Geometric Distortion, Photometric Distortion, Bilateral Filter, ShuffleNet-V2, YOLOv8.

1. Introduction

Cotton is one of the essential fiber and oilseed crop, it provides nearly 35% of entire fiber over worldwide. One of the largest producers of cotton in the world is China, and it is a significant agricultural commodity for the country.[1]. During 2022-23, the second largest cotton producer is India which produces 5.84 Million Metric Tons, nearly 23.83% of world cotton is produced in India [2].Cotton is a crop which has a significant growing season, an extensive range of pests and illnesses, and a very high pesticide usage rate during the entire growing season. Therefore, for the purpose to get a better spray effect, it is crucial to study how cotton targets are interpreted [3].It is one of the essential textile crops and a substantial source of foodsuch as cottonseed oil and hulls. It is also a high-value crop that fulfills the requirements for food, feed, and fiber.[4]. Cotton harvesting is one of the difficult task which requires more skilled labor, possess maximum return on investments and so on. But, sometimes the production of cotton become lows due to low investment and lack of suitable conditions for harvesting [5].

The density of cotton plants initially influences yield potential, but can also be impacted by severe weather, disease, and pest pressure.The primary cause of production variations in cotton plants can be the preservation of bolls and blooms, which can impact the efficiency of harvesting and lead to losses due to decreased flow of harvested material [6].The cotton boll counting technique is utilized to count the production rate of cotton using Machine Learning (ML) and Deep Learning (DL) techniques. There are very limited techniques were developed for Cotton boll detection and counting [7].Cotton boll counting in the field is essential for predicting fiber yield and gaining deeper insights into the genetic and physiological processes driving the crop's growth and development.

Counting frequently may evaluate the rate of growth in the blooming and blossom maturity stages and facilitate the selection of genotypes that use their energy more efficiently for growing cotton goods [8]. The Cotton boll counting techniques are developed based on various factor such as stand count, blossom and blooming count [9]. A normalized difference vegetation index (NDVI) utilized to segment and count the cotton plants. Here, the K-nearest Neighbor (KNN) algorithm is used for feature detection and matching.Then, combine the obtained features from feature detection and matching to develop an efficient cotton counting technique [10]. The Cotton boll counting in two-dimensional (2D) imaging technologies is the widely utilized techniques which performed based on phenotyping of cotton. The presence of Cotton boll are detected by various neural network-based data process pipeline in color images.

Phenotypic traits of various plants and crops are detected using image-based methods, which are also utilized to detect the count of cotton bolls and measure the size of cotton. [11]. The UAV-based RGB imager and DL technology utilized to provide efficient imagery data processing and framework for cotton emergence detection. Some of geo-reference crop row in every individual image were segmented the pre-processing pipeline. Then, the count of cotton stand and canopy size are evaluated by DL mode for automatic geo-referenced emergence maps in cotton [12]. The weeds specific to cotton is identified by the deep transfer learning (DTL) to analyze the supervised modelling using large volume of annotated data [13]. The cultivation of cotton wereanalyzed based on various factors such as fiber maturity, color, trash traits and fiber length. The production of Cotton bolls is evaluated based on humidity, rainfall, temperature as regional climatic characteristics [14, 15]. The CNN technique utilized to extract the features effectively for the image processing and DenseNet-121 pre-trained Model utilized for classify or detect the plant. This transfer learning technique utilizes ImageNet weights for cotton plant disease detection.[16]. Assorted pre-trained models of RegNet utilized to extract the feature of leaf.

Kernel-PCA model utilized to reduce the dimensionality of extracted feature and XGBoost utilized for classification [17]. An Unmanned Aerial Vehicles (UAV's) and AI technique utilized to classify the image for soil fertility analysis which enhance the production. Utilizing the early growth stage of crops during harvesting reduces the usage of fertilizers and pesticides.[18].

1.1 Motivation

Accurate cotton boll counting is essential for yield estimation, crop monitoring, resource allocation, and research in agriculture. It enables farmers to make informed decisions about harvesting schedules, resource management, and variety selection, ultimately leading to improved productivity and profitability in cotton production. Developing an automated system for cotton boll counting in images addresses the need for efficient and accurate crop monitoring and management. The challenge lies in designing a computer vision or machine learning-based solution capable of accurately detecting and counting cotton bolls amidst varying backgrounds, occlusions, and lighting conditions, while also being robust to factors such as plant growth stage and image quality. Existing model provide a drawbacks like overfitting, consumption of time for processing, inaccurate performance in the count of cotton bolls. These challenges motivate the proposal of an efficient attention based deep learning model for achieving accurate cotton boll detection and counting.Leveraging attention mechanisms, the model dynamically focuses on relevant image regions, improving robustness to varying conditions. By reducing processing time and enhancing accuracy, this solution aims to empower farmers with reliable information for better decision-making in cotton production.The major contribution of proposed methodologyare described as follows.

- To present an efficient attention based deep learning model for attaining accurate cotton boll detection along with its counts.
- To introduce Rolling joint bilateral filter (RJBF)approachfor denoising the images to provide appropriate features
- To implement Attention based Shuffle Bi-directional feature pyramid network enclosed YOLOv8 (ASB-FPNY) model for detection of cotton boll count.
- To compare the performance of proposed approach with the recently developed approaches for cotton boll detection.

The remaining content of paper is organized in the following manner: Section 2 includes the survey of related technique with its performance and limitation. In Section 3, the details and work flow of proposed technique is described with Figures. The results are analyzed and discussed in Section 4 and overall conclusion of this paper is given in Section 5.

2. Related works

Survey of several related techniques were analyzed onCotton boll counting and are described as follows.

Ch. Gangadhar et al. [19] developed a novel deep adversarial network (DAN) named CropCycleNet which combines the features of Convolutional Neural Network (CNN) and Generative Adversarial Networks (GAN). The stage of cotton plant was identified by DAN and Histogram base Gradients Feature Orientation transform method is utilized for feature descriptors and detection performed by feature level fusion. This technique obtains 93.27% accuracy in the harvesting stage in cotton fields.

Zhang Guohui et al. [20] developed a binocular matching algorithm named Shape feature extraction matching method (SFEMM). This technique utilized a matching equation to match the objects and it extracts the binocular matching images based on shape and space position of cotton. The matching precision was obtained to be 87.52% which also provide a feasible algorithm for benchmark picking of cotton-picking robots.

Ioannis E. Livieris et al. [21] developed a multi-input neural network model to predict the production of cotton. This technique utilize three kinds of data as input such as cultivation management data, yield management data and soil data. This technique reduced the overfitting, provide more flexibility and adapt low computational cost. Limited sample dataset is utilized for the prediction of cotton production.

Le Wand et al. [22] developed an automated crop stand counting using CNN based on color images which extracted from the video clips. The corn seedlings were counted by Kalman filter which achieved 98% accuracy. This technique combines YoloV3-tiny with Kalman filter and a high-speed one-stage detection network when overlapping occurred the accuracy of the model is reduced.

KhadijehAlibabaei et al. [23] developed a two DL models such as bidirectional LSTM (Bi-LSTM) and Bidirectional Gated Recurrent unit (BiGRU) to predict the end-of season yields which analyze the time-series data. This technique is performed based on some historical data such as irrigation scheduling, soil water content and climate data. This technique obtained 0.17 to 0.038 MSE to predict the yields. The survey of various related techniques, along with their performance and limitations, is described in Table 1.

Table 1: Survey of existing techniques with its performance and limitations

Author name and reference	**Techniques used**	**Limitations**	**Performance**
Ch. Gangadhar et al. [19]	CropCycleNet	Computationally expensive and time-consuming.	93.27% accuracy
Zhang Guohui et al. [20]	SFEMM	Difficult to collect data from farmland and generate the cotton dataset.	87.52% matching precision
Ioannis E. Livieris et al. [21]	Multi-input neural network model	Limited samples	38.07 - MAE
Le Wand et al. [22]	CNN	Occurrence of overfitting reduce the accuracy of this model	98% accuracy
KhadijehAlibabaei et al. [23]	Bi-LSTM and BiGRU	High time consumption and difficult for training	0.17 to 0.038 MSE

The challenge is in creating a computer vision or machine learning-based system that reliably recognize and count cotton bolls in a variety of lighting, occlusion, and background situations while remaining resilient to elements like image quality and plant growth stage. The existing model has several problems including overfitting, processing time consumption, and unreliable cotton boll counting.Some techniques were difficult to use for collecting data from farmland and generating the cotton dataset.Fewer techniques achieve high time consumption and are difficult for training.

3. Proposed methodology

Monitoring the number of bolls on cotton plants throughout the growing season provides valuable insights into the health and development of the crop. Changes in boll count over time can indicate factors such as nutrient deficiencies, pest infestations, water stress, or disease outbreaks. Moreover, Cotton boll count detection is pivotal in agriculture, aiding farmers in estimating crop yield, monitoring plant health, and optimizing resource allocation. By accurately assessing boll numbers, farmers can make informed decisions about harvest timing and resource management, ensuring optimal crop quality and yield. Furthermore, this data is invaluable for researchers and breeders in developing resilient cotton varieties and enhancing agricultural practices. Additionally, it supports crop insurance assessments and risk management strategies, contributing to the sustainability and profitability of cotton production.

This detection serves as a fundamental tool in modern farming, driving efficiency, resilience, and innovation in the cotton industry. Existing frame work for cotton boll detection possess some disadvantage like overfitting, consumption of time for processing, inaccurate performance in the count of cotton bolls. This research work providea deep learning based model to detect the cotton boll count in an efficient manner. Initially, the images collected from the dataset are augmented to balance the dataset and also reduce the overfitting effect by using geometric distortion and photometric distortion approaches. After the augmentations, the imagesget pre-processed to reduce the noise by using RJBF approach. From the pre-processed image the features are extracted and cotton bolls are detected with counts using ASB-FPNY. In the proposed architecture, progressive attention function is used for adjusting the model to focus over iterations dynamically. The components of YOLOv8 network are input, backbone, neck, and head. The Backbone is primarily used in the position of converting raw pixels into high-level semantic information by extracting features from the input image data. Here, the YOLOv8 architecture performs the fusion while ShuffleNet serves as the backbone. Therefore, the multi-scale features recovered by the Backbone were fused together by the Neck to achieve more accurate cotton boll detection which performed as the intermediary layer of the network structure. Head is the network's output layer which generates the final bounding boxes and cotton boll counts by using semantic and spatial information from the input. The workflow of proposed technique is represented in Figure 1.

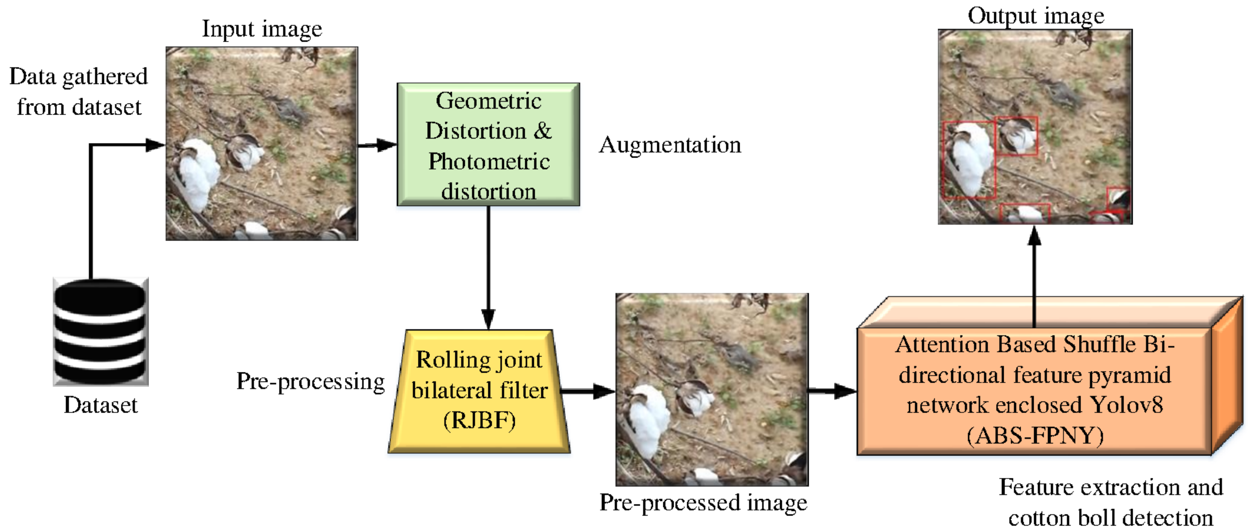


Figure 1: Workflow of proposed technique

3.1 Augmentation using Geometric distortions and photometric distortion approaches

Augmentation is a technique used to expand the dataset artificially by generating new transformed versions of an image from the input image to enhance its diversity. In this process, the Geometric distortion and photometric distortion approaches are utilized to augment the input image.

3.1.1 Geometric distortion

Geometric distortion is a technique utilized to address the geometric distortions in images [24]. The distortion are occurred due to various factors such as object movement, lens imperfections and camera perspective. In this technique, the geometric distortion flipping vertical, geometric distortions flipping horizontal, geometric distortions shearing, geometric distortions Translation and geometric distortions Rescaling are utilized for correcting the geometric distortions.

- ***Geometric distortions flipping vertical:*** Every row of pixels in the input image are reversed to perform the vertical flip. In this process, the top pixel row becomes last pixel row, second row become last second row, etc. until the last pixel row become the top pixel row.
- ***Geometric distortions flipping horizontal:*** Every row of pixels in the input image are reversed horizontally to perform the horizontal flip. In this process, the first column becomes last pixel column, second column become last second column, etc. until the last column becomes the first pixel column.
- ***Geometric distortions shearing:*** Shearing is a geometric transformation which distorts the shape of an object by shifting every point along a parallel line.
- ***Geometric distortions translation:***Translation is utilized to shift the pixel of image along a particular direction which is also known as the process of geometric transformation in image processing.
- ***Geometric distortions rescaling:***Rescaling isone of the geometric transformation utilized to change the size of images which is also known as scaling and resizing. To change the size of an image, the coordinates of each pixel are multiplied or divided by a scaling factor.

3.1.2 Photometric Distortion

The photometric distortion is the variation in brightness over the image which is caused by imperfection in uneven illumination or optical system [25]. In this technique, some process are performed to reduce the photometric distortion. The photometric Distortion is corrected by processing brightness, contrast enhancement, saturation, hue & saturation, noise and blur reduction. The brightness, saturation, hue & saturation, noise and blur are reduced, then the contrast of image are enhanced to correct the photometric distortion.

3.2 Pre-processing

The pre-processing stage is performed to improve the quality of image for efficient feature extraction. In this stage the RJBF filter is utilized to reduce the noise from image.

3.2.1 Rolling Joint bilateral filter (RJBF) for noise reduction

The Rolling Joint bilateral filter (RJBF) is used for noise reduction which is a successor of Rolling Guidance Filter (RGF), in which the bilateral filter is utilized instead of Guidance filter for effective outcome [26]. The RJBF is majorly utilized to remove the texture effectively and preservation of structure. It have various advantages such as scalable, adaptable, simple and easy to implement. The mathematical expression for the RJBF is described below.

$$RJBF(n) = \frac{\sum_{m \in F(n)} G_{n,m} B(m)}{\sum_{m \in F(n)} G_{n,m}} \tag{1}$$

Here, the mathematical expression for the bilateral kernel is described below.

$$G_{n,m} = e^{-\frac{(n-m)^2}{2\sigma_a^2}} e^{-\frac{(SP(n)-SP(m))^2}{2\sigma_b^2}}$$

(2)

Here, the input image and the sample image are represented as B and SP. The parameters to control the weights are represented as σ_a and σ_b, the window centered at pixel n is denoted as $F(n)$ and if $SP = B$, the RJBF is the bilateral filter (BF). It includes two steps, i.e., small structure removal and edge restoration. Initially, Gaussian filter is utilized to filter the input image which obtain the variance σ_a^2 and a filtered image R_{σ_a}. The output image is completely eliminated when size of input structure is less than σ_a and reduce other image structure. Then, other steps in RGF is utilized to perform iterative edge restoration. In this step, the sample image C^n is iteratively updated, C^0 is determined as R_{σ_a} for initialization. The output of $(n+1)^{th}$ iteration is denoted as C^{n+1} which is generated by RJBF, whose input image is original image B and reference image is previous iteration C^n. The RJBF is utilized to extract the main structure of input image B and the similarity between pixels are acquired by utilizing C^n which is also similar to the output image structures.

3.3 Feature Extraction and Cotton Boll Detection using ASB-FPNY model

The ASB-FPNY model is proposed for feature extraction and cotton boll detection which are a combination of Yolov8, ShuffleNetv2 module, Ghost module,Squeeze and Excitation (SE)attentionmodule and progressive attention module.

3.3.1 ShuffleNet2 Module

The shuffleNetv2 module is constructed by considering the cause of various factors such as degree of parallelism, memory access cost on speed [27]. Then, it was developed based on the four guidelines for classification. In the basic unit, $s=1$ where s represented as stride andthe input feature channel divided as two categories using input feature channels which is placed at head of the unit. To improve network parallelism and reduce fragmented operations, one part was carried out on the left branch without any operations. Thenother part flowed into the right branch and underwent 1×1 regular convolution, 3×3 depth convolution and 1×1 regular convolution to avoid pointwise convolution and reduce time consumption.

After convolution, the two branches combine while retaining the same number of input and output channels in order to lower the cost of memory access. Through channel shuffling, shuffle the feature map at random along the channel dimension. Also it facilitates channel fusion and information interchange between the two branches.Then, $s=2$ is used and the channel splitting were no longer performed. An equal amount of feature maps are received by the two branches. The right branch undergone 1×1 regular convolution, 3×3 depth convolution and 1×1 regular convolution while the left branch undergone 3×3 depth convolution and 1×1 regular convolution. The depth convolution in both left and right branches were added together after the dimension reduction. Similarly, the information exchange between various groups are facilitated by applying the channel shuffling. In Yolov8, the backbone portion is replaced with down sampling and ShuffleNetv2 basic unit. It is utilized for network complexity reduction, number of parameters and facilitate the network design.The architecture of ShuffleNet2 Module for $s=1$ and $s=2$ are represented in Figure 2.

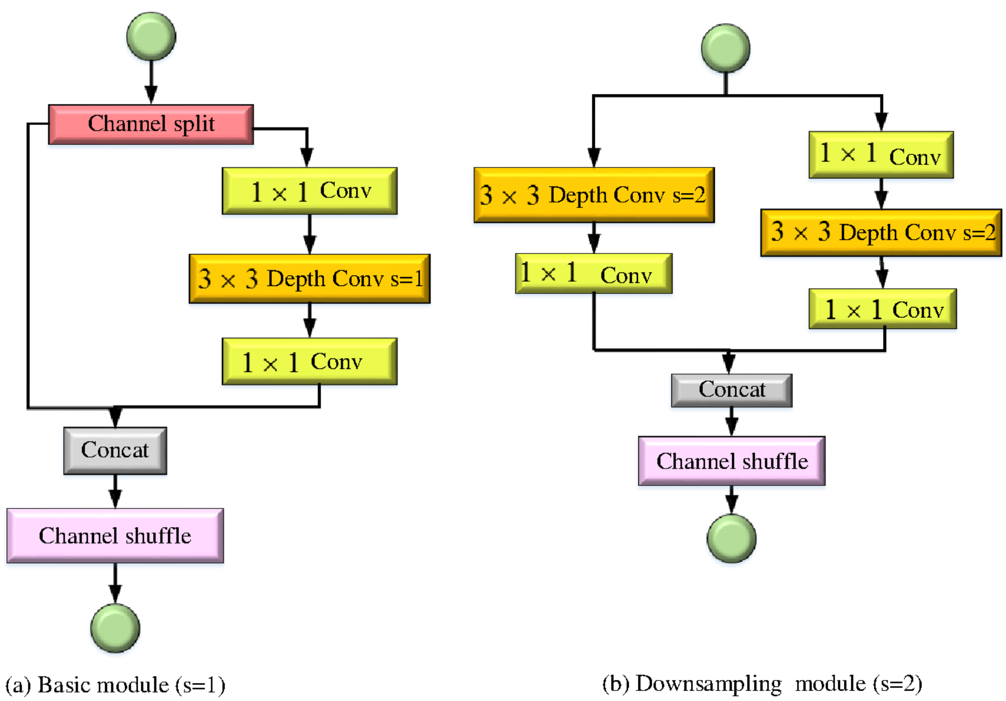


Figure 2: Architecture of ShuffleNet2 Module for (a) s=1 and (b) s=2

3.3.2 Ghost module

The Ghost module is an efficient and compact model with fewer parameters and it is utilized to find the feature redundancy issue by stationing large number of convolutional layer in deep convolutional neural network (DCNN).Ittechnically process regular non-linear convolutional layer when ghost module adopts the phased convolutional computation approach. It performs a low-cost linear convolutional operation on the feature maps generated by a nonlinear convolution employing half of the convolutional kernels in order to acquire extra features. Through combining the feature maps generated by the linear and nonlinear convolutions, the original feature map and channel sizes remained while the network's parameter and computational requirements were greatly reduced.

The GhostBottleneck is utilized as a suitable network structure of small-scale CNN model to develop the Ghost module. It includes two branches named major branch and residual branch. If $s=1$ where stride denoted as s, then major branch combines two ghost modules. Although the second Ghost module connected to the residual branch and decreases the number of channels in a feature map, the first Ghost module mainly enhances the number of input feature channels. If $s=2$, before connecting with the residual branch, the main branch combines the spatial dimension of the feature map using two Ghost modules and depthwise separable convolution. This model, which uses the Ghost module in Yolov8's neck, replaces the conventional convolution. The C2fGhost module is generated by combining the C2f module with GhostBottleneck. Then, it used to achieve model compression and lower the computational requirements.

3.3.3 SE attention module

SE attention mechanism extracts significant features which have major effects on the network and improve neural networks' representational capacity. The SE attention module includes two parts such as Squeeze and Excitation. When combining the feature maps of size, the global average pooling method is employed with the channel dimension $A\times B\times C$ into $1\times 1\times C$ in the squeeze stage. The whole feature map is replaced by combining every channel into a feature vector along with a single value. It is utilized to reduce the computational cost and overfitting issues. Then, two fully connected layersare utilized in the excitation stage. The vector output from convolutional layer scaled by the first fully connected layer to emphasize efficient feature. Then, the weighted excitation is applied to the outcome of the initial layer by the second fully connected layer and assign weight coefficients to every channel. In order to enhance the network's perception, the network explicitly models the interdependencies across channels, recalibrates feature maps according to the value of different channels, and emphasizes certain features while reducing irrelevant ones.

The backbone of Yolov8 is embedded by SE attention mechanism at output position of 80×80, 40×40 and 20×20 feature maps. Then, the output position in neck of Yolov8 embedded with C2fGhost module. This technique improves the expressiveness of feature maps, overall accuracy and reduce the computational time by suppressing the irrelevant features.

3.3.4 Progressive attention module

Let the output feature map of layer $s\in\{0,1,...,S\}$ in Yolov8 is $p^s\in\Re^{A_s\times B_s\times C_s}$ with width B_s, height A_s and number of channels C_s. Then, $p^s_{i,j}\in\Re^{C_s}$ is assumed to be feature at $(i,j)^{th}$ feature map p^s. The attentive process applied in this model to multiple layers of Yolosv8 to obtain the attended feature map $\hat{p}^s=\left[\hat{p}^s_{i,j}\right]$ which is described in the below equation.

$$\hat{p}^s_{i,j}=\alpha^s_{i,j}p^s_{i,j} \tag{3}$$

The attention probability $\alpha_{i,j}^{s}$ for a feature $p_{i,j}^{s}$ and a query u are described in the below equation.

$$e_{i,j}^{s} = h_{att}^{s}\left(p_{i,j}^{s}, u; \theta_{att}^{s}\right) \tag{4}$$

$$\alpha_{i,j}^{s} = \begin{cases} soft\max_{i,j}\left(e^{s}\right) & if \;\; s = S \\ \sigma\left(e_{i,j}^{s}\right) & else \end{cases} \tag{5}$$

Here, the attention function with a set of parameter θ_{att}^{s} is denoted as $h_{att}^{s}(\cdot)$ and the attention score at (i, j) is denoted as $e_{i,j}^{s}$ and sigmoid function is represented as $\sigma(\cdot)$. The same feature map is used to estimate the attention probability at every point individually which is limited between 0 and 1 using a sigmoid function. The last layer of attention is utilized by applying a SoftMax function in the whole spatial region for subsequent feature aggregation.

The attended feature map $\hat{P}^{s}$ in the intermediate attention layers is not added together to generate one vector representation of all attended regions, compared to the soft attention model. Actually, the attended feature map is transmitted to the following layer, where it is utilized as an input to generate the subsequent feature map which is described in the below equation.

$$p^{s+1} = h_{Conv}^{s+1}\left(\hat{p}^{s}; \theta_{Conv}^{s+1}\right) \tag{6}$$

Here, convolution operation at $s+1$ layer in convolution parameterized by θ_{Conv}^{s+1} is denoted as $h_{Conv}^{s+1}(\cdot)$. The feedforward procedure performed with attentive process from input layer to obtain $\hat{p}^{S}$, where $p^{0} = X$. Finally, add every features in final attended feature map $\hat{p}^{S}$ to retrieve the attended feature as in soft attention is described in the below equation.

$$p^{att} = \sum_{i}^{A}\sum_{j}^{B}\hat{p}_{i,j}^{S} = \sum_{i}^{A}\sum_{j}^{B}\alpha_{i,j}^{S}p_{i,j}^{S}$$

(7)

This process provides attended feature p^{att} as output which is used as input to the visual attribute classifier. This attention layer is placed to the output of C2fGhost module instead of Ghost module. It is utilized to enhance the overall accuracy of cotton boll detecting and counting.

3.3.5 YOLOv8 model

The YOLO model is an effective and successive computer vision technique by adding more modules.YOLOv1 is utilized for objected detection but, it is difficult to detect the small object in the image. YOLOv2 have a difficult in object detection which were overlapping and closer together. YOLOv3 is difficult in objection detection which were heavily occluded and truncated. YOLOv4 have a complex structure while extracting the features. YOLOv5 has lower accuracy compared to other object detection model. YOLOv6 still have difficulties in small object detection and YOLOv7 has more complex structure than YOLOv5 and YOLOv6 which made YOLOv7 model very hard to train. To overcomes these kinds of issues, YOLOv8 is utilized with ShuffleNetv2 and Ghost module, SE attention module and progressive attention model to provide efficient feature extraction and cotton boll detection.

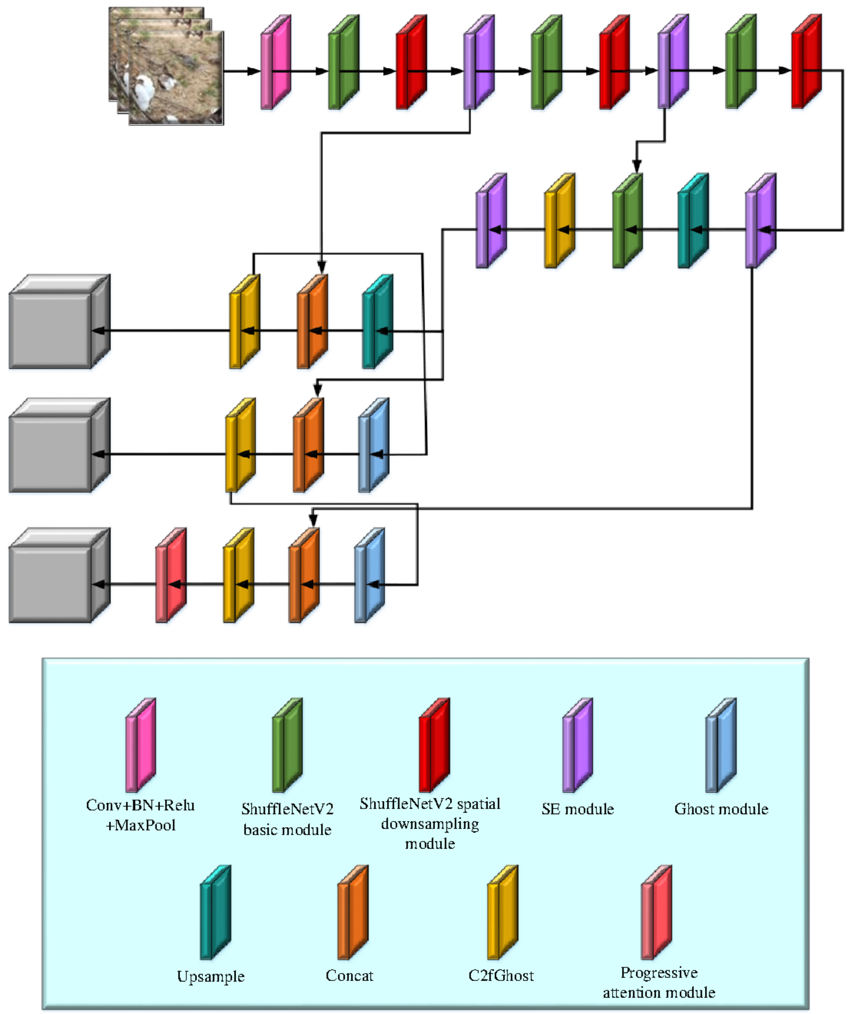
YOLOv8 enable to provide accurate and efficient object detection in real-time scenarios with its advanced architecture and cutting edge algorithm. YOLOv8 model utilizes three scale-detection layer to identify the object with various scales [28]. The bounding box regression loss are reduced using WIoU v3 which incorporates a dynamic non-monotonic mechanism. It is utilized to reduce the occurrence of harmful and large gradients from sample images.It is mainly utilized to enhance the generalization ability and quality of images.

3.3.6 ASB-FPNY model

The ASB-FPNY model is developed using YOLOv8, ShuffleNetv2 module, Ghost module, SE attention module and progressive attention module. The structure of YOLOv8 network includes the four parts such as input, Backbone, Neck and Head. The Backbone part is utilized to extract the features from the pre-processed image and convert the raw pixel into high-level semantic features. Then, in order to ensure effective object detection, the neck part functions as an intermediary layer by fusing multi-scale features that the backbone has extracted. Finally, the output layer as Head provides the final detection boxes and layers of classification by extracting the spatial and syntactic data from the input.

Three modules such as CBS, C2f, and Spatial Pyramid Pooling-Fast (SPPF) are utilized to develop the YOLOv8 model's basis. The C2f module uses cross-layer connections to incorporate more branches in order to produce gradient flow data that is more detailed. In order to reduce computational costs while generating a lightweight model that has higher accuracy, and also modifies the structural characteristics. The SPPF module combined local and global data by connecting three max-pooling layers. The combination of Path Aggregation Network (PAN) and Feature Pyramid Network (FPN) (PAN-FPN) are adapted by the Neck. It utilizes the top-down upsampling and bottom-up downsampling to fuse the multi-scale features. The disk storage burden is increased due to the large number of convolutional layers in Backbone and Neck.

The efficiency of objection model is developed by ensuring more accuracy while the size of model is compressed by reducing the parameters and computational costs. The YOLOv8 model includes Input, Backbone, Neck and Head. The Conv_BN_Relu_Maxpool module is adapted by backbone of YOLOv8 to reduce the parameters and computational cost of input to next layers. It is performed by staked arrangement of three ShuffleNetV2 module and downsampling modules. The dimensionality of input features were reduced by the combination of ShuffleNetV2 module and downsampling modules. The output feature maps of size 80×80, 40×40 and 20×20 were obtained by the three combination modules. The SE attention mechanism modules were connected to output feature maps in order to mitigate network fragmentation, reduce model size and enhance network perception. By this method, essential attributes could be effectively extracted and irrelevant characteristics are suppressed. Decreasing network complexity and parameters even further, the C2fGhost module in the Neck is created by merging the original C2f module with the GhostBottleneck module.The architecture of ASB-FPNY model is represented in Figure 3.



V

Figure 3: Architecture of ASB-FPNY model

4. Results and discussion

The overall performance of proposed technique is analyzed based on various performance metrics such as Accuracy, Precision, Recall and so on. Then, comparison is conducted with various related technique to determine the efficient performance of proposed technique.

4.1 Dataset description

The dataset is known as Annotated dataset for weakly supervised Cotton boll detection and counting. It covered nearly 290 RGB image of cotton plants from indoor, output and in-field through the consumer-grade cameras. Then, it divided into 4350 image tiles for model training and testing. The link of dataset utilized in proposed technique is givenhttps://figshare.com/articles/dataset/Annotated_dataset_for_weakly_supervised_Cotton_boll_detection_and_counting/19665096.

4.2 Performance metric and its formulation

The performance analysis of proposed technique is evaluated based on various performance metrics which are described as follows.

4.2.1 Accuracy

Accuracy is a performance statistic used to assess the efficacy of the suggested approach by comparing the number of corrected predictions to the total number of predictions. The mathematical expression of proposed technique is described as below.

$$acc = \frac{(T_{pos} + T_{neg})}{(T_{pos} + F_{pos} + T_{neg} + F_{neg})}$$

(8)

Here, T_{pos} and T_{neg} are represented as true positive and true negative rate. The false positive rate and false negative rate are denoted as F_{pos} and F_{neg}.

4.2.2 Precision

Precision is a performance parameter that is measured by the percentage of detection specified in the equation below.

$$P = \frac{T_{pos}}{T_{pos} + F_{pos}} \tag{9}$$

4.2.3 Recall

Recall is the percentage of relevant data items properly discovered by the model. The mathematical expression for proposed method is described in the below equation.

$$R = \frac{T_{pos}}{T_{pos} + F_{neg}} \tag{10}$$

4.2.4 F1 score

The F1-score is also a performance metric analyzed by the perfect precision and recall score, which is described in the equation below.

$$F1 = \frac{2 \times P \times R}{P + R} \tag{11}$$

4.2.5 Specificity

The percentage of every positivity analysis is defined as specificity rate which is evaluatedusing thebelow described equation.

$$S = \frac{T_{neg}}{T_{meg} + F_{pos}} \tag{12}$$

4.2.6 MAE

The average absolute difference between the predicted and actual value of model is utilized to measure the effectiveness is known as MAE which is described in the equation below.

$$MAE = \frac{1}{x}\sum_{s=0}^{x} |a_s - \hat{a}_s| \tag{13}$$

Here, the predicted and actual value of model is denoted as a_s and $\hat{a}_s$.

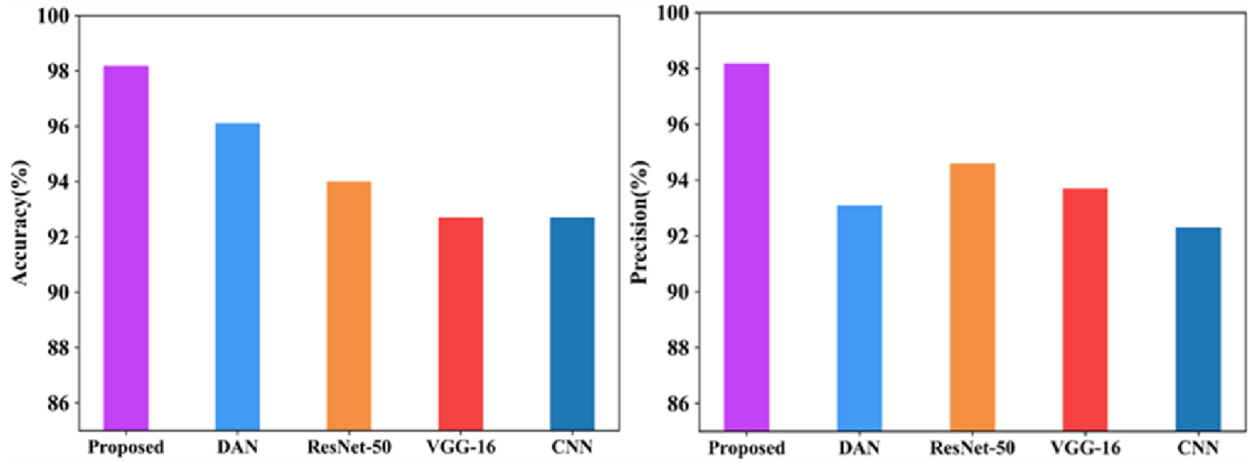
4.2.7 RMSE

The squared difference between the predicted and actual value are divided by total number of predictions and then square root the output is known as RMSE which is described in the below equation.

$$RMSE = \sqrt{\frac{\sum_{s=0}^{x} (a_s - \hat{a}_s)^2}{x}} \tag{14}$$

4.3 Performance analysis

The performance of proposed technique analyzed with various techniques such as CNN, VGG-16), Residual Network version 50 (ResNet-50) and Dense Adversarial Network (DAN). The comparison of Accuracy and Precision are represented in Figure 4.

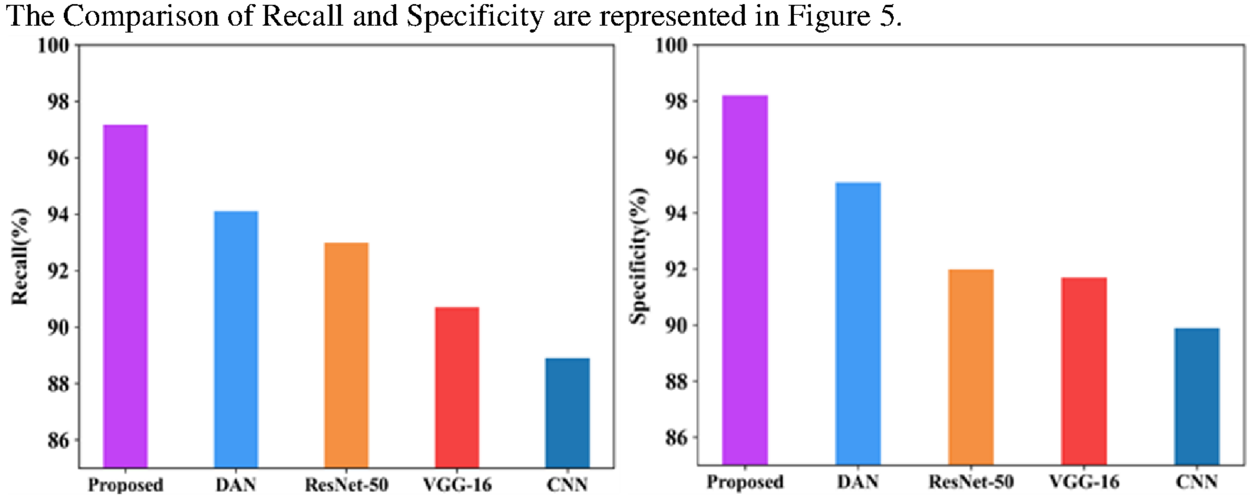


(a) (b)

Figure 4: Comparison of (a) Accuracy and (b) Precision

The accuracy of proposed technique is obtained to be 98.18% which is higher than other related techniques such as CNN (90.9%), VGG-16 (92.7%), ResNet-50 (94%) and DAN (96.1%). It clearly determines the efficient performance of proposed technique compared to other related technique. Similarly, the proposed technique achieves 98.19% precision which is higher than the related techniques.

The Comparison of Recall and Specificity are represented in Figure 5.



(a) (b)

Figure 5: Comparison of (a) Recall and (b) Specificity

The Recall and Specificity of proposed technique is obtained to be 97.18% and 97.68% for cotton boll counting. Then, comparisons were conducted to determine the efficiency of the proposed techniques. The recall rate of various existing techniques were CNN (88.9%), VGG-16

(90.7%), ResNet-50 (92.9%) and DAN (94.1%). The specificity of the existing techniques were as follows: CNN (89.9%), VGG-16 (91.7%), ResNet-50 (91.9%) and DAN (95.1%). The analysis of recall and specificity clearly determines the efficient performance of the proposed technique. The F1-score of both proposed and related techniques is represented in Figure 6.

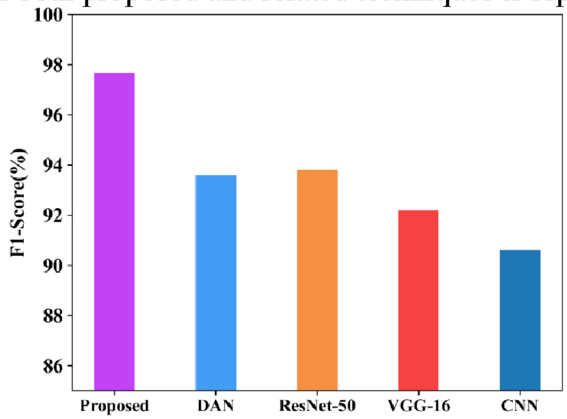
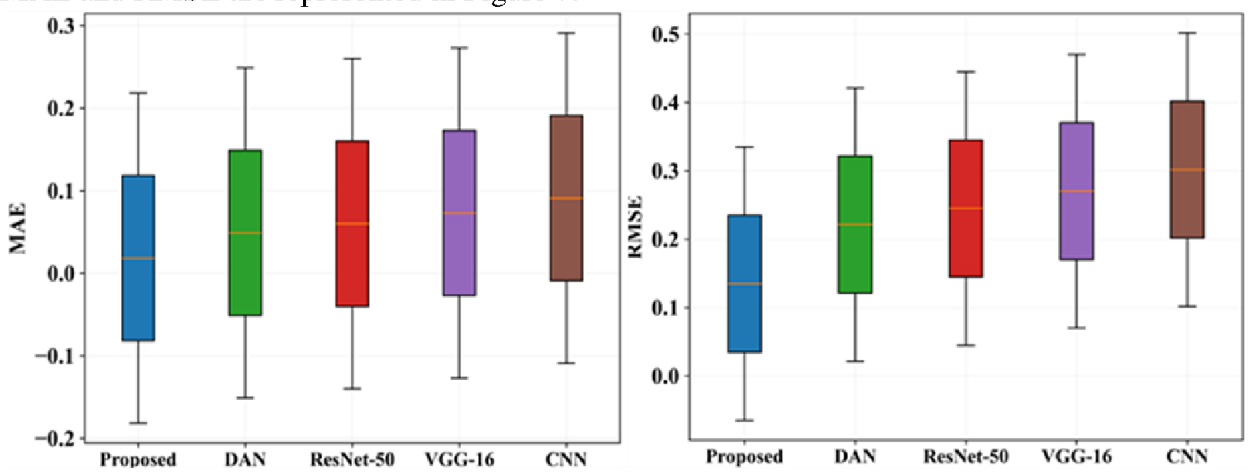


Figure 6: Comparison of F1-score

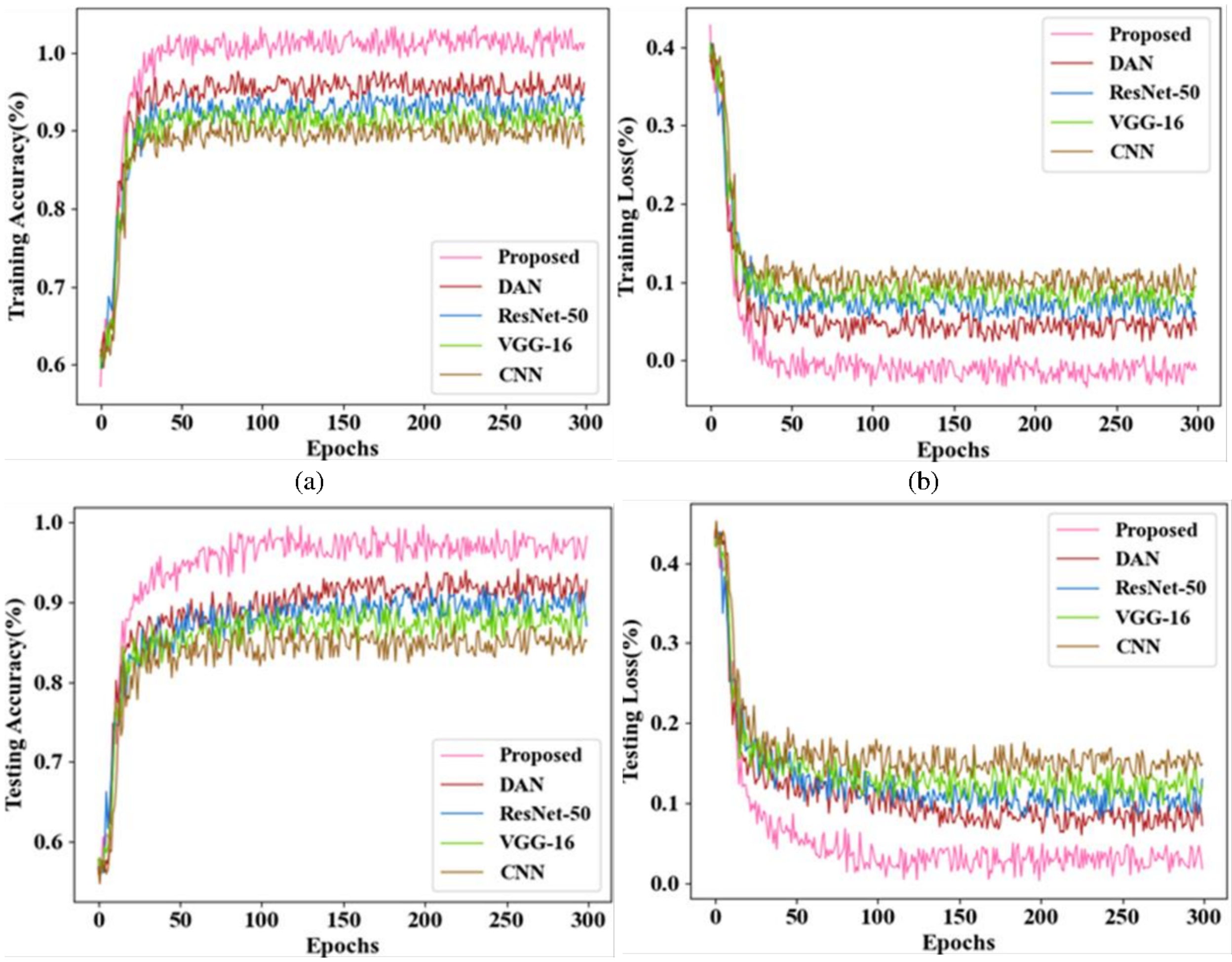
The F1-score of proposed technique is obtained to be 97.68% which is higher than other related techniques such as CNN (90.6%), VGG-16 (92.2%), ResNet-10 (93.8%) and DAN (93.6%). These performance analysis for F1-score determines that the proposed technique achieves efficient performance compared to other related techniques. The Error metrics such as MAE and RMSE are represented in Figure 7.



(a) (b)

Figure 7: Error metric of (a) MAE and (b) RMSE

The MAE rate of proposed technique is obtained to be 0.0182 which is much lower than other related techniques such as CNN (0.091), VGG-16 (0.073), ResNet-50 (0.06) and DAN (0.049). Then, the RMSE rate of the proposed technique were analyzed and it achievedto be0.1349 which is lower than CNN (0.3017), VGG-16 (0.2702), ResNet-50 (0.2449) and DAN (0.2214). These analysis clearly determines the efficient performance of proposed technique. The comparison analysis of accuracy and loss for both training and testing are represented in Figure 8.



(c) (d)

Figure 8: Comparison of (a) Training Accuracy, (b) Training Loss, (c) Testing Accuracy and (d) Testing Loss

The accuracy and loss of both training and testing are analyzed with various related techniques to determine the efficiency of the proposed technique. The dataset were divided into 80:20, where 80% of data is utilized to train the model and 20% of data is utilized for testing. Figure 8clearly shows that the accuracy of training and testing is slightly enhanced compared to other related techniques. Similarly, the loss of training and testing are reduced compared to other existing techniques which determines the efficient performance of the proposed technique. The comparison of ROC curve is represented in Figure 9.

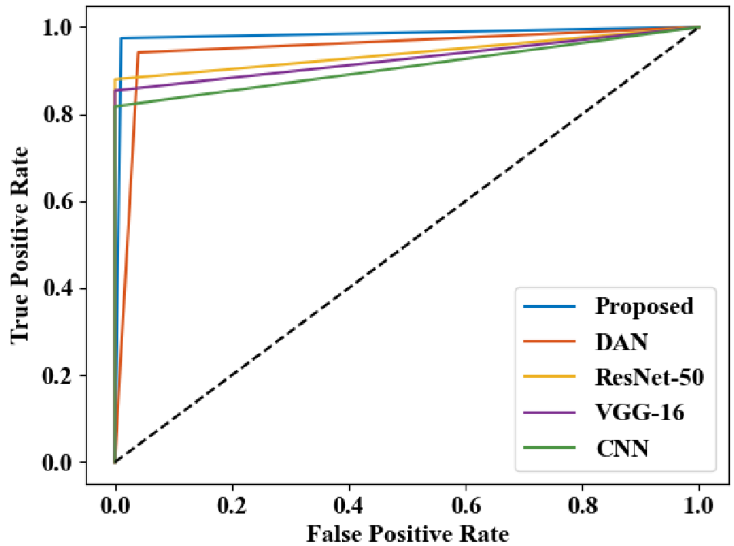


Figure 9: Comparison of ROC curve

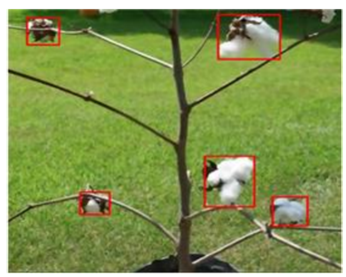
The ROC curve of the proposed technique is analyzed with various other techniques to determine the effectiveness of the proposed technique for cotton boll detection and counting. It is analyzed based on the AUC values between the True Positive Rate and False Positive Rate.This ROC is also analyzed with various related techniques such as CNN, VGG-16, ResNet-50 and DAN. It clearly determines the efficient performance of the proposed technique compared to other related techniques. The performance analysis of the proposed technique with various related techniques is determined in Table 2.

Table 2: Performance analysis of the proposed technique

Technique used	**Performance analysis**						
	Accuracy (%)	**Precision (%)**	**Recall (%)**	**Specificity (%)**	**F1-score (%)**	**MAE**	**RMSE**
CNN	90.9	92.3	88.9	89.9	90.6	0.091	0.3017
VGG-16	92.7	93.7	90.7	91.7	92.2	0.073	0.2702
ResNet-50	94	94.6	92.9	91.9	93.8	0.06	0.2449
DAN	96.1	93.09	94.1	95.1	93.6	0.049	0.2214
Proposed	**98.18**	**98.19**	**97.18**	**98.2**	**97.68**	**0.0182**	**0.1349**

The performance analysis of the proposed technique in Table 2 clearly determines its efficiency compared to other related techniques such as CNN, VGG-16, ResNet-50, and DAN.Comparing to other related techniques, The proposed technique achieves 98.18% accuracy, 98.19% precision and 97.18% recall. The outcomes of each technique utilized in the proposed method are described in Table 3.

Table 3: Outcomes of each technique in proposed methodology

Technique used	Output
Input image	
Rolling joint bilateral filter (RJBF)	
Attention Based Shuffle Bi-directional feature Pyramid Network enclosed YOLOv8 (ABS-FPNY) model	

The output of each technique utilized in the proposed technique such as RJBF and ASB-FPNY model is given in Table 3. It determines the quality and efficiency of each technique utilized in the proposed method. Here, the input image, output of RJBF and output of ASB-FPNY model are given. The final output of the proposed technique is represented in Table 3, where the cotton bolls are detected and displayed with bounding boxes.

4.4 Discussion

The performance of proposed technique on cotton boll detection and counting is analyzed and the efficiency is compared with various related techniques. The comparative analysis of the proposed technique with various existing approaches based on their performance is represented in Table 4.

Table 4: Comparative analysis

Author name and reference	**Techniques used**	**Performance**
Ch. Gangadhar et al. [19]	CropCycleNet	93.27% accuracy
Zhang Guohui et al. [20]	SFEMM	87.52% matching precision
Ioannis E. Livieris et al. [21]	Multi-input neural network model	38.07 - MAE
Le Wand et al. [22]	CNN	98% accuracy
KhadijehAlibabaei et al. [23]	Bi-LSTM and BiGRU	0.17 to 0.038 MSE

Proposed	**ASB-FPNY model**	**98.18% accuracy**

Ch. Gangadhar et al. [19] developed CropCycleNet to achieve 93.27% accuracy and Zhang Guohui et al. [20] developed SFEMM to obtain 87.52% matching precision. The MAE of Multi-input neural network model is 38.07 which developed by Ioannis E. Livieris et al. [21]. Similarly, the other related techniques were analyzed and compared with proposed technique to determine its efficient performance. The proposed model ASB-FPNY achieved 98.18%.

5. Conclusion

The proposed approach has been developed to provide an effective cotton boll detection method. First, the data was collected from datasets, which were improved utilizing geometric and photometric distortions techniques. Then the image is pre-processed using RJBF to improve the image quality. Finally, the feature extraction from the pre-processed image and cotton boll detection are done using the ASB-FPNY model. The performance of the ASB-FPNY model for cotton boll detection is improved by the progressive attention feature. The effectiveness of the proposed technique is determined through comparing its performance with several related techniques. The proposed technique achieves an accuracy of 98.18% which is higher than that of othertechniques.Alsothe proposed method achieved 98.19% precision, 97.18% recall, 98.2% specificity, and 97.68% F1-score.The MAE and RMSE rates of the proposed technique are obtained as 0.0182 and 0.1349 respectively, which are lower than those of other related techniques. In future, the performance of the proposed technique will be enhanced by exploring an advanced DL-based detection techniques. The efficiency of the proposed technique will be enhanced by extracting efficient features using a hybrid technique.

References

[1] Ni, Chao, Zhenye Li, Xiong Zhang, Xinyan Sun, Yuping Huang, Ling Zhao, Tingting Zhu, and Dongyi Wang. "Online sorting of the film on cotton based on deep learning and hyperspectral imaging." *Ieee Access* 8 (2020): 93028-93038.

[2] Syazwani, R. Wan Nurazwin, H. Muhammad Asraf, MA MegatSyahirul Amin, and KA Nur Dalila. "Automated image identification, detection and fruit counting of top-view pineapple crown using machine learning." *Alexandria Engineering Journal* 61, no. 2 (2022): 1265-1276.

[3] Wang, Shanping, Yang Li, Jin Yuan, Laiqi Song, Xinghua Liu, and Xuemei Liu. "Recognition of cotton growth period for precise spraying based on convolution neural network." *Information processing in agriculture* 8, no. 2 (2021): 219-231.

[4] Ramacharan, S. "A 3-stage method for disease detection of cotton plant leaf using deep learning CNN algorithm." *Int. J. Res. Appl. Sci. Eng. Technol.* 9, no. VII (2021): 2503-2510.

[5] Jiang, Yu, Changying Li, Rui Xu, Shangpeng Sun, Jon S. Robertson, and Andrew H. Paterson. "DeepFlower: a deep learning-based approach to characterize flowering patterns of cotton plants in the field." *Plant methods* 16 (2020): 1-17.

[6] Oh, Sungchan, Anjin Chang, Akash Ashapure, Jinha Jung, Nothabo Dube, Murilo Maeda, Daniel Gonzalez, and Juan Landivar. "Plant counting of cotton from UAS imagery using deep learning-based object detection framework." *Remote Sensing* 12, no. 18 (2020): 2981.

[7] Verma, Pooja, Gagandeep Kaur, Rajendra Machavaram, and Mahua Bhattacharya. "Intelligent Cotton boll maturity prediction model for smart agriculture." In *AIP Conference Proceedings*, vol. 2919, no. 1. AIP Publishing, 2024.

[8] Dewangan, Umashankar, R. H. Talwekar, and SwagotaBera. "Systematic Literature Review on Crop Yield Prediction using Machine & Deep Learning Algorithm." In *2022 5th International Conference on Advances in Science and Technology (ICAST)*, pp. 654-661. IEEE, 2022.

[9] Lin, Zhe, and Wenxuan Guo. "Cotton stand counting from unmanned aerial system imagery using mobilenet and centernet deep learning models." *Remote Sensing* 13, no. 14 (2021): 2822.
[10] Liu, Qianhui, Yan Zhang, and Gongping Yang. "Small unopened cotton boll counting by detection with MRF-YOLO in the wild." *Computers and Electronics in Agriculture* 204 (2023): 107576.
[11] Sun, Shangpeng, Changying Li, Peng W. Chee, Andrew H. Paterson, Yu Jiang, Rui Xu, Jon S. Robertson, Jeevan Adhikari, and Tariq Shehzad. "Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering." *ISPRS Journal of Photogrammetry and Remote Sensing* 160 (2020): 195-207.
[12] Feng, Aijing, Jianfeng Zhou, Earl Vories, and Kenneth A. Sudduth. "Evaluation of cotton emergence using UAV-based imagery and deep learning." *Computers and Electronics in Agriculture* 177 (2020): 105711.
[13] Chen, Dong, Yuzhen Lu, Zhaojian Li, and Sierra Young. "Performance evaluation of deep transfer learning on multi-class identification of common weed species in cotton production systems." *Computers and Electronics in Agriculture* 198 (2022): 107091.
[14] Darawsheh, Mohammed K., DimitriosBeslemes, Varvara Kouneli, EvangeliaTigka, DimitriosBilalis, IoannisRoussis, Stella Karydogianni et al. "Environmental and regional effects on fiber quality of cotton cultivated in Greece." *Agronomy* 12, no. 4 (2022): 943.
[15] Kumar, Sandeep, Arpit Jain, Anand Prakash Shukla, Satyendr Singh, Rohit Raja, Shilpa Rani, G. Harshitha, Mohammed A. AlZain, and Mehedi Masud. "A comparative analysis of machine learning algorithms for detection of organic and nonorganic cotton diseases." *Mathematical Problems in Engineering* 2021 (2021): 1-18.
[16] Arathi, B., and Uma N. Dulhare. "Classification of cotton leaf diseases using transfer learning-DenseNet-121." In *Proceedings of third international conference on advances in computer engineering and communication systems: ICACECS 2022*, pp. 393-405. Singapore: Springer Nature Singapore, 2023.
[17] Gouse, Sheikh, and Uma N. Dulhare. "Automation of Rice Leaf Diseases Prediction Using Deep Learning Hybrid Model VVIR." In *International Conference on Advancements in Smart Computing and Information Security*, pp. 133-143. Cham: Springer Nature Switzerland, 2022.
[18] Dulhare, Uma N., and Sheikh Gouse. "Automation of rice cultivation from ploughing–harvesting with diseases, pests and weeds to increase the yield using AI." In ICCCE 2021: Proceedings of the 4th International Conference on Communications and Cyber Physical Engineering, pp. 505-513. Singapore: Springer Nature Singapore, 2022.
[19] Gangadhar, Ch, R. Reji, MusmadeBhausahebBhanudas, Swati Vijay Shinde, Devendra Kumar, and D. Kavitha. "Prediction and detection of harvesting stage in cotton fields using deep adversarial networks." *Soft Computing* 28, no. 2 (2024): 1819-1831.
[20] Guohui, Zhang, Gulbahar Tohti, Chen Ping, MamtiminGeni, and Fan Yixuan. "SFEMM: A cotton binocular matching method based on YOLOv7x." *Mathematical Biosciences and Engineering* 21, no. 3 (2024): 3618-3630.
[21] Livieris, Ioannis E., Spiros D. Dafnis, George K. Papadopoulos, and Dionissios P. Kalivas. "A multiple-input neural network model for predicting cotton production quantity: A case study." *Algorithms* 13, no. 11 (2020): 273.
[22] Wang, Le, Lirong Xiang, Lie Tang, and Huanyu Jiang. "A convolutional neural network-based method for corn stand counting in the field." *Sensors* 21, no. 2 (2021): 507.

[23] Alibabaei, Khadijeh, Pedro D. Gaspar, and Tania M. Lima. "Crop yield estimation using deep learning based on climate big data and irrigation scheduling." *Energies* 14, no. 11 (2021): 3004.
[24] Yamamoto, Tetsuya, Masaki Fukunaga, Sho K. Sugawara, Yuki H. Hamano, and NorihiroSadato. "Quantitative evaluations of geometrical distortion corrections in cortical surface-based analysis of high-resolution functional MRI data at 7T." *Journal of Magnetic Resonance Imaging* 53, no. 4 (2021): 1220-1234.
[25] García-Vega, Axel, Ricardo Espinosa, Gilberto Ochoa-Ruiz, Thomas Bazin, Luis Falcón-Morales, Dominique Lamarque, and Christian Daul. "A novel hybrid endoscopic dataset for evaluating machine learning-based photometric image enhancement models." In *Mexican International Conference on Artificial Intelligence*, pp. 267-281. Cham: Springer Nature Switzerland, 2022.
[26] Suryanarayana, Gunnam, Karthik Chandran, Osamah Ibrahim Khalaf, Youseef Alotaibi, AbdulmajeedAlsufyani, and Saleh Ahmed Alghamdi. "Accurate magnetic resonance image super-resolution using deep networks and Gaussian filtering in the stationary wavelet domain." IEEE Access 9 (2021): 71406-71417.
[27] Chen, Zhichao, Jie Yang, Lifang Chen, and Haining Jiao. "Garbage classification system based on improved ShuffleNet v2." *Resources, Conservation and Recycling* 178 (2022): 106090.
[28] Xiao, Bingjie, Minh Nguyen, and Wei Qi Yan. "Fruit ripeness identification using YOLOv8 model." *Multimedia Tools and Applications* 83, no. 9 (2024): 28039-28056.