

Modification of YOLOv11 Architecture with Convolutional Block Attention Module Integration for Batik Besurek Motif Detection

Reksi Hendra Pratama^{1*}, Ernawati², Widhia Oktoberza KZ³

^{1,2,3}Informatics Engineering, Faculty of Engineering, Universitas Bengkulu, Bengkulu, Indonesia
E-mail: ^{1*}reksihendrapratama637@gmail.com, ²ernawati@unib.ac.id, ³widhiakz@unib.ac.id

(Received: 22 Jan 2026, revised: 2 Mar 2026, accepted: 3 Mar 2026)

Abstract

Batik Besurek is a distinctive cultural heritage textile from Bengkulu, characterized by the combination of Arabic calligraphy, floral, and faunal motifs. However, public recognition of these motifs continues to decline, particularly among younger generations, creating an urgent need for digital preservation through modern computer vision techniques. This study aims to develop an enhanced object detection model for identifying Batik Besurek motifs by integrating the Convolutional Block Attention Module (CBAM) into the Neck (feature fusion layers) of the YOLOv11 architecture. A total of 3,876 images were collected from online repositories and field documentation in Bengkulu, followed by annotation and data augmentation using the Roboflow platform. Specifically, Mosaic augmentation combined with a close-mosaic strategy proved most significant in stabilizing the detection of dense and minority motifs. The research employs an experimental methodology by constructing two models, standard YOLOv11 as the baseline and YOLOv11+CBAM as the proposed model, trained under identical hyperparameter configurations. Model performance was evaluated using standard object detection metrics, namely precision, recall, mAP@50, and mAP@50–95. The experimental results demonstrate that the integration of CBAM significantly improves feature extraction capability, enabling the model to better capture dense, small, and visually complex batik motifs. The proposed model outperforms the baseline across all metrics, achieving a mAP@50 of 0.988, mAP@50-95 of 0.914, Recall of 0.971, and Precision of 0.959. These results indicate superior sensitivity and localization accuracy, particularly for minority motif classes, confirming that incorporating attention mechanisms into YOLOv11 provides a robust solution for the digital preservation of Besurek patterns.

Keywords: Batik Besurek, Object Detection, YOLOv11, CBAM, Deep Learning.

I. INTRODUCTION

Batik is not merely a patterned fabric but a visual manifestation of indigenous knowledge and cultural identity, recognized by UNESCO as Intangible Cultural Heritage [1]. Currently, efforts to preserve Indonesian textiles require a precise computational approach to ensure cultural sustainability and intellectual property rights are maintained amid the onslaught of the global creative industry [2]. Specifically, Batik Besurek from Bengkulu holds a unique position in Indonesian textiles due to its visual characteristics, which combine Arabic calligraphy (*huruf gundul*) with motifs of endemic flora and fauna, such as the Rafflesia Arnoldi flower. A recent dissertation study [3] highlights that Batik Besurek continues to evolve dynamically, from classic motifs dominated by intricate calligraphic curves to modern variations such as Gajah Seblat and Dol, which highlight local wisdom.

However, digitizing and automatically classifying these motifs poses considerable technical challenges. The main problem stems from their extreme visual complexity. As found in the Cantonese Embroidery detection case study, cultural heritage images often have highly variable morphology with dense fabric backgrounds, making it difficult to separate the main object from the supporting ornaments [4]. The same thing happens with Besurek Batik. Traditional motifs have a high degree of inter-class similarity; for example, the curves in the Calligraphy motif often visually overlap with the vines in the Moon motif. As a result, conventional feature extraction methods that rely on global shapes, such as the Histogram of Oriented Gradients (HOG), have proven inefficient in capturing subtle semantic details in these complex patterns [5].

This complexity is further exacerbated by the reality of imbalanced datasets in the field [6]. In digital repositories and conventional markets, image availability is often dominated by popular mass-produced motifs. In contrast, specific development motifs such as Gajah Seblat or Dol have very



minimal sample representation. This imbalance in data distribution is not a trivial matter. Research [7] confirms that these conditions cause Deep Learning models to become biased and “lazy,” predicting the majority class and ignoring unique features in the minority class. Therefore, simply adding data with standard augmentation is often not enough; it is necessary to modify the model architecture so that the artificial neural network can focus more on these rare features.

In the field of Computer Vision, the You Only Look Once (YOLO) algorithm has now become the de facto industry standard for real-time object detection. In the specific context of Bengkulu cultural preservation, the effectiveness of this architecture has been demonstrated by Sutresno [8], who utilized YOLOv8 to classify five main motifs of Batik Besurek (such as the Burung Kuau, Kaligrafi, and Rafflesia). The study recorded impressive performance with an average mAP@50 of 96% and Recall of 94%. However, this study was limited to a dataset with popular motif variations and did not specifically address the issue of extreme imbalance in rare motifs.

The latest version of YOLO, YOLOv11, offers significantly improved parameter efficiency and accuracy compared to its predecessors, YOLOv8 and YOLOv10. Research [9] notes that YOLOv11 achieves an average mAP@50-95 of 87.5% in multi-object classification. This capability has also proven robust in industrial sectors, such as in steel defect detection [10] and real-time textile inspection [11], thanks in particular to its advanced feature fusion mechanism that enables it to recognize small targets.

However, no matter how fast its performance is, the application of standard YOLO to complex textile patterns such as batik still leaves a research gap. Studies on surface crack detection [12] reveal that this model is often overwhelmed when dealing with irregular textures and occlusion interference without the help of additional mechanisms. The fundamental problem lies in the standard convolution operation, which treats all pixels with the same weight, thus failing to distinguish essential features from visual noise due to low contrast between the main motif and the background (background interference) [13]. In Batik Besurek, which is full of *isen-isen* (background filling dots), the model often misinterprets the background as part of the motif.

Preliminary data from previous implementations, such as the application of YOLOv8 on Batik Besurek [8], explicitly demonstrates the severity of this background interference. Based on the confusion matrix from that study, the standard model generated a significant number of false positives by misclassifying the background as primary motifs 48 times. Consistent with these findings, our preliminary observations using the standard YOLOv11 architecture prior to CBAM integration revealed identical vulnerabilities. When confronted with dense *isen-isen*, the baseline YOLOv11 severely struggled to differentiate textures, producing 34 misclassifications in the Kaligrafi Lawas class and 11 in the Kembang Melati class by incorrectly predicting background areas as motifs. This documented frequency of failure in both previous and current baseline models provides a strong, empirical justification for integrating an attention mechanism to suppress visual noise before further feature extraction.

Based on the analysis of these issues, this study proposes a solution by integrating the YOLOv11 architecture with the CBAM (Convolutional Block Attention Module). CBAM was chosen for its ability to adaptively refine feature maps through two sequential paths [14]. Channel Attention is tasked with selecting dominant features to reduce inter-class ambiguity [15]. Meanwhile, Spatial Attention specifically maps the locations of relevant features to separate motifs from the background (background suppression) and to minimize noise [16]. The effectiveness of this hybrid approach has been previously proven by [17] which significantly improved the accuracy of small object detection (mAP 76.68%), and by study [18] which noted an increase in segmentation precision for overlapping objects in dense environments.

This study aims to: (1) Implement and optimize the YOLOv11+CBAM architecture for the classification of Besurek Batik motifs in imbalanced datasets; and (2) Analyze in depth the effectiveness of this attention mechanism in boosting the Recall and mAP metrics, especially for minority motifs that have failed to be detected by the baseline model.

II. RESEARCH METHOD

A. Research Framework

This study employs a two-model approach, namely YOLOv11 Baseline and YOLOv11 + CBAM, by adapting the systematic framework from the latest batik classification study [6]. This approach was selected for its effectiveness in addressing the challenges of recognizing complex traditional textile patterns through structured stages, spanning data management to model performance validation.

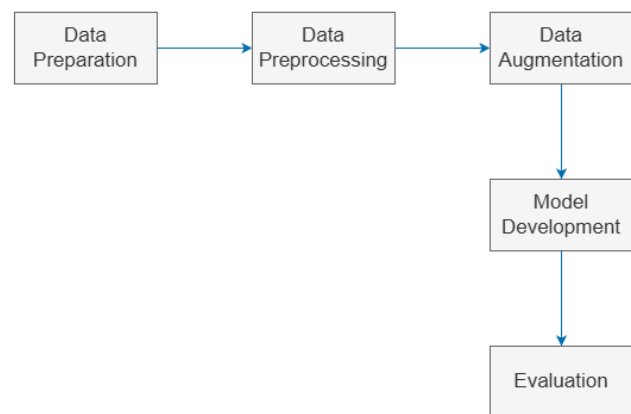


Figure 1. Research Process

The research process, as shown in Figure 1, was carried out through five main stages as follows:

1. Data Preparation

The initial stage focused on collecting Batik Besurek images from two sources to ensure visual variety. Secondary data were obtained from a public repository (Roboflow) to train generalization, while primary data was acquired independently using a smartphone to supplement the

representation of development motifs (Gajah Seblat, Dol, Kembang Cengkeh) that had minimal digital documentation. At this stage, data annotation was also carried out for primary data that had not yet been annotated [19].

2. Preprocessing

This stage aims to standardize raw data to match the model architecture input specifications. The process includes standardizing image dimensions to maintain visual feature consistency, as well as splitting the dataset into several separate subsets (training, validation, and testing) to ensure an objective evaluation process free from bias.

3. Data Augmentation

This stage is applied to training data to overcome class imbalance and prevent overfitting. Various geometric transformations and image manipulation techniques are applied to enrich visual pattern variation without changing object identity, and image fusion strategies are used to improve model robustness against complex backgrounds.

4. Model Development

At this stage, the Standard YOLOv11 model architecture and the YOLOv11 model integrated with the CBAM (Convolutional Block Attention Module) attention module are designed and developed. At this stage, the model is trained using optimal hyperparameter configurations to maximize feature extraction capabilities on complex and overlapping batik patterns.

5. Evaluation

The performance of the trained model is measured using standard object detection evaluation metrics, namely Precision, Recall, and Mean Average Precision (mAP). An in-depth analysis of the detection results is conducted to validate the effectiveness of the proposed method in overcoming data imbalance issues before the model is considered suitable for further development.

B. Dataset and Data Preparation

This study utilizes a dataset of Batik Besurek Bengkulu comprising 3,876 unique digital images. However, given that specific images contain multiple distinct motif classes, the cumulative class-wise count totals 4,161. Despite the finite number of source images, the dataset encompasses 10,059 annotated instances. This high density of instances relative to images is attributed to the inherent nature of batik textiles, which feature repeating patterns. Consequently, individual images frequently exhibit motif co-occurrence, containing multiple instances or diverse motif types simultaneously [20].

1. Data Sources

The data collection strategy employs a dual-source approach to ensure comprehensive visual variation across motifs:

- Secondary Data (Roboflow):** A total of 3,817 images were retrieved from the Roboflow public repository [21]. This subset comprises popular motifs set against diverse

backgrounds, facilitating the training of model generalization capabilities.

- Primary Data (Self-Acquisition):** Fifty nine images were manually acquired specifically for rare motifs unavailable in online repositories: *Gajah Seblat*, *Dol*, and *Kembang Cengkeh*. Image acquisition was conducted using a Xiaomi Redmi 12C smartphone camera without studio lighting. The deliberate use of a low-cost device aims to simulate "in-the-wild" scenarios, where system inputs frequently exhibit natural noise and suboptimal illumination [22]. Subsequently, the primary dataset was annotated via the Roboflow platform using a hybrid approach. Manual bounding boxes were applied to simple motifs, while the Segment Anything Model (SAM) was utilized for complex patterns to generate polygon segmentation that precisely traces the batik contours. The resulting annotations were converted to the YOLO format and validated before preprocessing and model development.

2. Class Distribution

Table 1. Distribution of the Batik Besurek Dataset

Motif Class	Image Counts	Annotated Instances	Data Source
Burung Punai	189	730	Secondary
Burung Kuau	532	1054	Secondary
Burung Kuau (Var. 1)	218	1013	Secondary
Dol	13	20	Primary
Gajah Seblat	31	61	Primary
Kaligrafi	221	977	Secondary
Kaligrafi Lawas	747	749	Secondary
Kembang Cempaka	829	829	Secondary
Kembang Cengkeh	15	41	Primary
Kembang Melati	244	1018	Secondary
Pohon Hayat	217	1149	Secondary
Rafflesia	380	1397	Secondary
Rembulan	525	1021	Secondary

The dataset exhibits a significant class imbalance, as shown in Table 1, where the majority classes contain hundreds of image samples, while the minority class (self-acquisition) is represented by only a few samples. Although the volume of primary data (59 images) appears limited compared to the secondary dataset, this accurately reflects the real-world constraints of Bengkulu's contemporary development motifs (Gajah Seblat, Dol, and Kembang Cengkeh), which currently have highly restricted physical production and digital availability. To ensure statistical validity and compensate for this scarcity, this study maximizes data quality and instance density. Due to the repeating nature of batik patterns, these 59 images yield 122 distinct annotated instances. Furthermore, utilizing the Segment Anything Model (SAM) for complex patterns generates highly precise polygon segmentations, maximizing the spatial feature quality extracted from each rare instance. Such an imbalance may lead to biased model predictions favoring the majority classes. Therefore, data



augmentation techniques and attention mechanisms are employed to enhance the model's focus on the minority class, proving that integrating CBAM is the exact solution to force the network to focus on minority features rather than relying solely on large data volumes.

C. Preprocessing and Data Augmentation Strategies

The pre-processing and data augmentation stages are designed to align the visual characteristics of the dataset with the input specifications of the YOLOv11 architecture and to enrich the training data variation to prevent overfitting on minority classes. In the initial data management stage in Roboflow, raw images only go through the Auto-Orient process to standardize pixel orientation without permanently changing the physical resolution. Image resizing (dimension adjustment) is performed automatically by the dataloader during training (runtime). The input resolution is set to 640×640 pixels, referring to the optimal configuration of the study [17], which has been proven effective in balancing GPU computational efficiency and detail feature preservation in solid object detection. The dataset is then partitioned using stratified random sampling with a ratio of 80:10:10. This scheme allocates 80% of the data for training, 10% for validation, and 10% for testing, to ensure that the proportion of each class remains evenly represented in each evaluation subset [23].



Figure 2. Original Samples vs. Augmented Variations

To improve the model's resilience to visual variations, a dynamic data augmentation scenario was applied, as shown in Figure 2. Geometric transformations in the form of random rotations of $\pm 15^\circ$ and horizontal flips were applied because batik motifs have rotational invariance, where the orientation of the motifs on the fabric is not always perpendicular. In addition, brightness adjustments of $\pm 15\%$ were made to simulate variations in ambient lighting, given that the primary data were acquired with a mobile phone camera under uncontrolled lighting [24] [25]. Apart from geometric manipulation, this study also considers the real conditions of data collection in the field, which are often less than ideal. Given that the primary images were acquired with a mobile phone camera, visual disturbances in the form of fine spots (noise) are common, especially in low-light conditions.

Therefore, data augmentation with noise was applied to the training data

The primary augmentation strategy implemented to overcome data imbalance is Mosaic Augmentation. This technique combines four different training images into a single input grid to enrich the background context and force the model to learn objects on a smaller scale. However, to ensure optimal detection precision in the final phase, the parameter `close_mosaic=10` was applied. This strategy turns off the Mosaic feature in the last 10 epochs of training, which aims to stabilize the model weights using the original data distribution without cropping (clean images). This method has been proven effective in stabilizing the training process in the YOLOv11 architecture [17].

D. Model Architecture

This study adopts the architecture development strategy proposed by Chen et al. [26] by modifying the base YOLOv11 model to address detection challenges associated with batik motifs that exhibit dense and overlapping patterns. Prior to integration, the model structure is enhanced with the Convolutional Block Attention Module (CBAM). CBAM is a lightweight attention module that operates sequentially through two submodules: the Channel Attention Module (CAM) and the Spatial Attention Module (SAM) [14]. In the context of batik motif detection, CAM exploits inter-channel relationships to determine 'what' objects are important, while SAM focuses on 'where' those objects are located [14]. This combination is crucial given the highly complex visual backgrounds of batik patterns [27], where attention mechanisms are required to suppress background noise [16] and enhance relevant motif features, as has been effectively demonstrated in textile defect detection tasks [13].

Unlike most previous approaches that integrate attention mechanisms into the backbone as part of the feature extraction process [17] [28], this study strategically inserts three CBAM units within the Neck component. Based on the model configuration built, the CBAM module is placed at layers 17, 21, and 25. This placement is done immediately after the feature aggregation process (C3k2 block) in the upsampling and concatenation paths. Integration at this point aims to purify the fused feature maps from background noise that often appears when combining high and low resolution features, before passing them on to the final detection layer.

The fundamental differences between the standard architecture and the proposed model are visualized in Figure 3 and Figure 4.

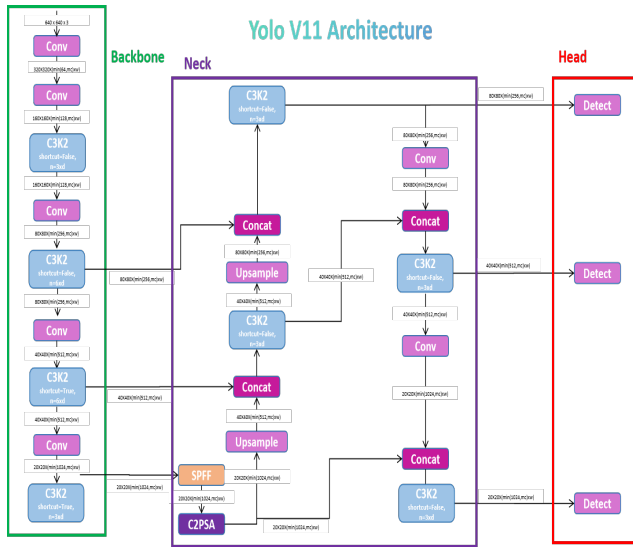


Figure 3. YOLOv11 Standard Architecture [29]

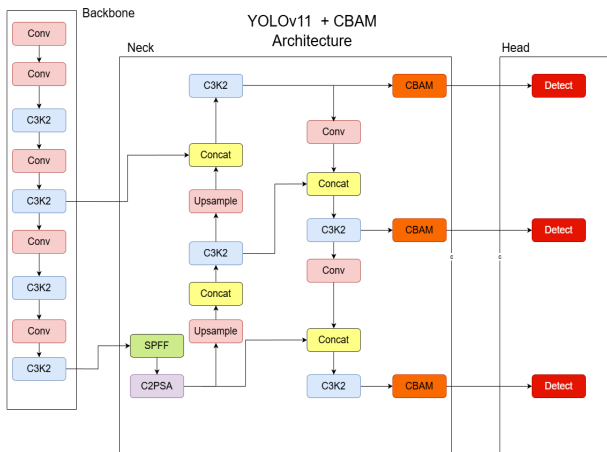


Figure 4. Proposed YOLOv11 with CBAM Integration

As illustrated in Figure 3, the standard YOLOv11 architecture forwards the aggregated features from the C3k2 blocks directly to the detection head without any filtering. In contrast, in the proposed architecture (Figure 4), this feature flow is intervened by the CBAM module, which produces three attention-enhanced output paths. This mechanism ensures that the model assigns greater weight to pixel regions containing batik *isen-isen* information.

E. Evaluation

1. Precision and Recall

Quantitative evaluation was conducted on the test dataset to measure the model's generalization performance on unseen data. The primary metrics used were Precision and Recall. Precision measures the accuracy of positive predictions to minimize false detections, while Recall assesses how completely the model identifies all existing batik motif objects. The mathematical relationship between these two metrics is presented in Equations (1) and (2) [30].

$$\text{Precision} = \frac{TP}{TP + FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (2)$$

Where:

- TP (True Positives): The number of batik motifs correctly detected.
- FP (False Positives): The number of incorrect detections, where non-motif objects are mistakenly identified as motifs.
- FN (False Negatives): The number of batik motif objects that are present in the ground truth but are not detected by the model.

2. Mean Average Precision (mAP)

For a more comprehensive performance assessment across all classes, this study uses the Mean Average Precision (mAP) metric. The mAP value is obtained by calculating the average of the Average Precision (AP) of each motif class, where AP represents the area under the Precision-Recall curve that describes the balance of model performance.

This evaluation includes mAP@50 (accuracy at IoU 0.50) and mAP@50–95. The use of mAP@50–95 is considered crucial for measuring the model's ability to accurately map the positions of dense and overlapping batik motif objects. The mAP formulation is shown in Equation (3) [30].

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (3)$$

Where :

- N = The total number of classes (batik motif categories).
- AP_i = the *Average Precision* of the i -th batik motif class.

III. RESULTS AND DISCUSSION

A. Training Environment and Parameters

In this study, the entire model training process was conducted using Google Colab as a cloud-based computing platform, which is widely used in research and education due to its ease of access and integrated GPU support. Training was run on an NVIDIA L4 GPU with 22.5 GB of memory, an accelerator designed for high-performance inference and deep learning computing. The Colab runtime provides 53 GB of system RAM for notebook execution in the Pro version, while the author's local device has 16 GB of RAM used for data preparation and experiment file management. On the software side, this research uses the Ultralytics YOLO framework version 8.3.252, Python 3.12.12, and PyTorch 2.9.0, a deep learning library designed for ease of programming and GPU-based execution efficiency. All hardware and software components are recorded in detail to ensure that this research can be consistently replicated by other researchers.

The training parameters used follow those in study [17]. The model was trained for 100 epochs with a batch size of 16 and an input resolution of 640×640 pixels. The optimizer used was AdamW with an initial learning rate of 0.001 and a warmup period of 10 epochs to aid learning stability in the early stages. The Mosaic augmentation technique was applied in the early training phase to increase sample diversity, but was disabled in the last 10 epochs through the setting `close_mosaic = 10`. This strategy allowed the model to benefit from strong augmentation in the early training, but still achieve stable convergence in the final phase. The model also saves checkpoints every 10 epochs and implements an early stopping mechanism with a patience of 30 epochs to avoid overfitting. All of these parameters are used consistently to ensure the replicability of the results and to ensure that the experiment can be revalidated in an equivalent computing environment.

B. Analysis of Training Dynamics

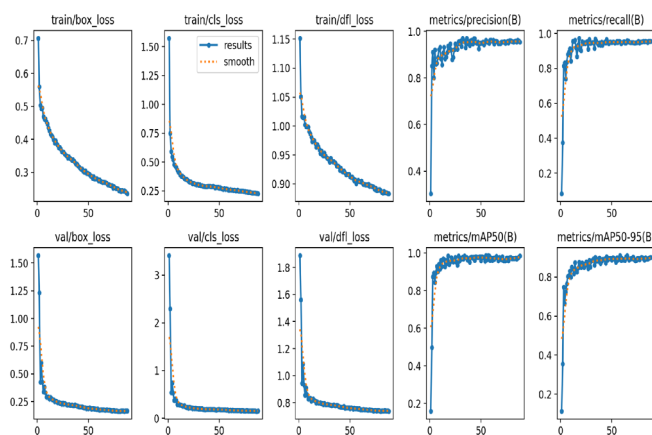


Figure 5. Training Curves of the Baseline Model

In the Baseline model, as shown in Figure 5, the training curve shows a very aggressive decline in loss in the early phase (0–20 epochs). This indicates that the standard model is able to quickly learn basic visual features (such as dominant shapes and colors). However, this ability soon reaches a saturation point. Based on the training data, the model reaches its peak performance (best fitness) relatively early, namely at the 58th epoch. After this point, the validation curve tends to plateau and does not show significant improvement even though training continues. This stagnation triggers the Early Stopping mechanism at the 88th epoch, which automatically stops training [31]. The phenomenon of “early death” in this training process implies that the Baseline model has limited capacity to extract the subtle features needed to distinguish complex batik motifs, causing the model to stop learning earlier.

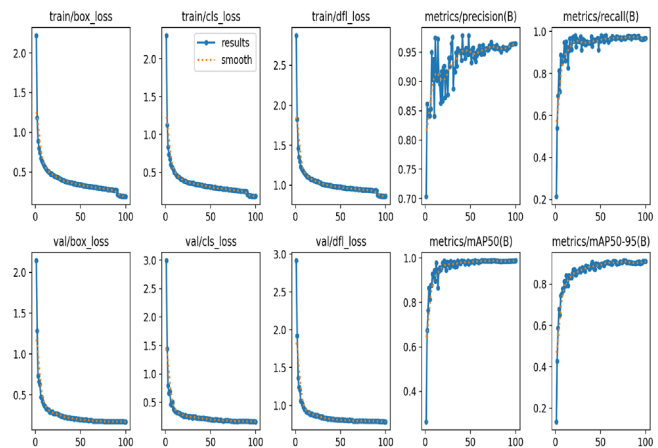


Figure 6. Training Curves of the YOLOv11+CBAM Model

In contrast, the YOLOv11 + CBAM model (Figure 6) shows a more robust and sustainable training profile. Although minor fluctuations (jitter) are seen in the precision metric in the early stages, which is common when the attention module is calibrating spatial weights and channels, this model has proven capable of maintaining longer learning momentum. Unlike the baseline, which stops at epoch 88, the CBAM model continues to learn until the whole 100 epochs. This allows the model to enter the crucial “Close Mosaic” phase in the last 10 epochs, where mosaic augmentation is disabled.

The impact of this phase is clearly visible in the training curve, where there is a sharp and significant decrease in loss immediately after epoch 90. This drastic decrease indicates that the model, which has been strengthened by the attention mechanism, is able to adapt very efficiently when switching from augmented images to unmixed images.

This extended effective training duration supports the hypothesis that the integration of the Convolutional Block Attention Module (CBAM) successfully enhances the model’s representational power. The Channel Attention and Spatial Attention mechanisms operate synergistically to suppress background noise and concentrate feature representations on relevant motif regions [14]. As a result, the model not only avoids premature saturation but is also able to achieve higher accuracy.

C. Quantitative Evaluation of Model Performance

1. Global Performance Analysis

A quantitative evaluation was conducted to compare the performance of the baseline YOLOv11 with the proposed YOLOv11 + CBAM model on the test dataset. Since both models were trained using identical configurations, the difference in performance reflects the architectural contribution of the CBAM module placed in the Neck section. A summary of the test results is presented in Table 2.

Table 2. Model Performance Comparison

Model	Precision	Recall	mAP@50	mAP@50-95
YOLOv11 Baseline	0.948	0.955	0.984	0.910
YOLOv11 + CBAM	0.959	0.971	0.988	0.914

The experimental results presented in Table 2 demonstrate that the integration of the CBAM module leads to a comprehensive performance improvement. By addressing the classical trade-off between precision and sensitivity, the proposed model is able to simultaneously increase Precision from 0.948 to 0.959 and Recall from 0.955 to 0.971. Improvements are also observed in mAP@50, which reaches 0.988, along with a substantial increase in mAP@50–95 to 0.914. The gain in this stricter metric reflects a crucial refinement in bounding box localization accuracy, particularly when handling challenging edge cases such as truncated or partially occluded objects.

This simultaneous improvement indicates that the CBAM module operates effectively according to its fundamental mechanism by emphasizing meaningful features while suppressing irrelevant ones [14]. This behavior enables a reduction in False Positives through the mitigation of complex background noise, which is consistent with recent findings showing that spatial attention is effective in lowering false detection rates in images with rich textures [32]. On the other hand, the improved model sensitivity helps minimize False Negatives by strengthening feature representations for motifs that were previously difficult to detect.

2. Confusion Matrix Analysis

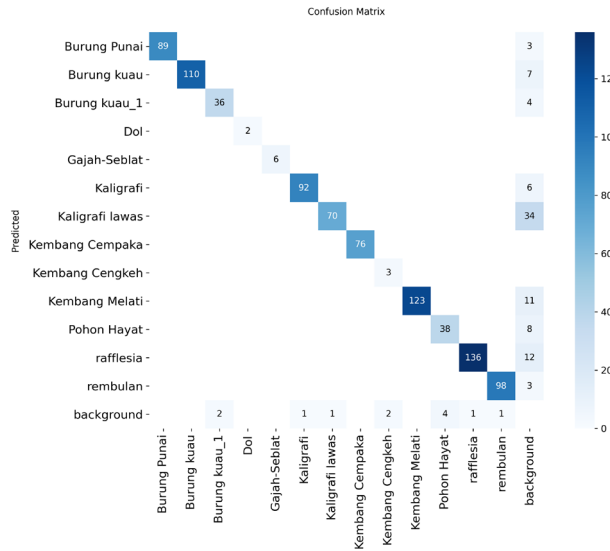


Figure 7. Confusion Matrix Baseline

The visualization in Figure 7 illustrates the prediction distribution of the baseline model, where the error patterns are predominantly driven by ambiguity between object regions and the background. This limitation is particularly evident in

classes with high visual complexity, such as “Kaligrafi Lawas,” where 34 misclassification instances occur due to background areas being incorrectly predicted as motifs. Similar errors are also prominent in the “Kembang Melati” class, with 11 false detection cases. The high number of False Positives indicates that the baseline model is not yet able to clearly distinguish subtle feature boundaries between batik motif textures and the underlying fabric fibers.

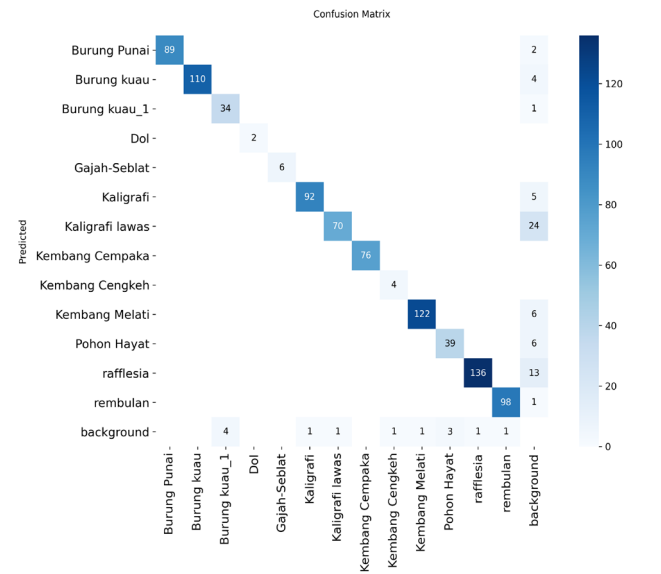


Figure 8. Confusion Matrix Proposed Model

Figure 8 illustrates how the integration of the CBAM module successfully addresses the limitations of the previous model by significantly reducing noise interference. The model demonstrates improved precision in recognizing object boundaries, with detection errors in the “Kaligrafi Lawas” class decreasing to 24 instances and those in the “Kembang Melati” class decreasing to 6 instances. This reduction confirms the critical role of Spatial Attention in filtering unnecessary visual information. This transformation serves as a valid indicator that the proposed model is better able to adapt to background variations compared to the baseline model.

3. Minority Class Performance Analysis

In addition to the global evaluation, performance analysis was also conducted at the class level to identify the specific characteristics of the improvements introduced by the attention module. Table 3 summarizes the comparison of performance metrics between the baseline model and YOLOv11 + CBAM across three minority classes, namely “Kembang Cengkeh”, “Dol”, and “Gajah Seblat”.

Table 3. Model Performance Comparison on Minority Classes

Motif Class	Model	Precision	Recall	mAP@50	mAP@50-95
Kembang Cengkeh	Baseline	1.000	0.659	0.962	0.519
	YOLOv11 + CBAM	1.000	0.884	0.995	0.504
Dol	Baseline	0.864	1.000	0.995	0.796
	YOLOv11 + CBAM	0.862	1.000	0.995	0.822
Gajah Seblat	Baseline	0.942	1.000	0.995	0.814
	YOLOv11 + CBAM	0.940	1.000	0.995	0.836

Table 3 presents class-specific performance dynamics following the integration of the CBAM module. The most pronounced change is observed in the “Kembang Cengkeh” class, where the proposed model achieves a substantial Recall improvement of 0.225, increasing from 0.659 to 0.884, while maintaining perfect Precision at 1.000. This sharp increase confirms that CBAM is effective in recovering objects that were previously missed, namely False Negatives, by the baseline model. In line with recent findings [32], this phenomenon indicates that feature enhancement through attention mechanisms is crucial for improving model sensitivity to objects with weak or ambiguous visual characteristics. The slight decrease in mAP@50–95 of 0.015 represents a reasonable trade-off, as the model is now able to recognize more challenging objects that were previously overlooked, albeit with bounding box accuracy that may be less optimal than for easily detectable objects.

In contrast, for the ‘Dol’ and ‘Gajah Seblat’ classes, Recall has already reached full saturation at 1.000 in both models. In this case, the role of CBAM shifts from improving detection to refining localization, reflected by consistent mAP@50-95 increases of 0.026 and 0.022, respectively. This demonstrates that when features are well-recognized, spatial attention becomes dominant in sharpening positional accuracy (the ‘where’ aspect), leading to more precise coordinate regression [14]. The very slight Precision decrease of 0.002 in both classes is considered negligible compared to these significant improvements in localization quality.

4. Qualitative Analysis of Model Failures

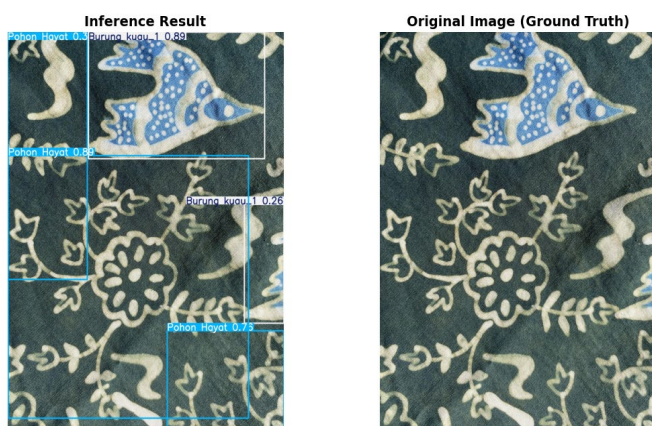


Figure 9. Qualitative Analysis of Model Failures Showing a False Positive Detection.

To provide a qualitative analysis of model failures for further study, Figure 9 illustrates a visual example of an image that failed to be detected accurately. In the top left corner of the inference result, a clear detection error occurs where a truncated edge and background texture are erroneously identified as the “Pohon Hayat” motif with a low confidence score of 0.3. This false positive explicitly demonstrates how the detection system can be easily misled by partial visual cues and background fabric interference without an attention mechanism to prioritize holistic object structures.

D. Comparison with Previous Studies

When compared with previous studies, the proposed YOLOv11 + CBAM model demonstrates consistent and meaningful performance improvements. The study by Sutresno employed the YOLOv8 architecture for Batik Besurek motif detection and classification using a dataset of 1,656 images distributed across eight motif classes, namely Burung Kuau (5 classes), Kaligrafi, Kembang Melati, Rafflesia, and Rembulan. That study reported an average Precision of approximately 0.96 and an average Recall of around 0.93 [8].

In this study, the proposed model achieves a Precision of 0.959 and a Recall of 0.971, indicating a stronger balance between detection accuracy and sensitivity. Moreover, while the prior work primarily focused on basic evaluation metrics, this study incorporates a stricter assessment using mAP@50–95 to evaluate bounding box localization accuracy. The achieved mAP@50–95 value of 0.914 reflects an improved ability to handle dense and overlapping batik motifs. These findings suggest that the integration of the CBAM module into the YOLOv11 architecture provides a tangible advantage in enhancing feature representation and robustness against complex background variations.

IV. CONCLUSION

This study successfully develops a modified YOLOv11 architecture by integrating the Convolutional Block Attention Module (CBAM) within the Neck pathway to address the visual complexity of Batik Besurek imagery. This mechanism has been proven effective in separating primary motifs from background noise, resulting in more accurate texture recognition.

Based on quantitative evaluation, the proposed model (YOLOv11 + CBAM) demonstrates substantial improvements over the baseline. The model achieves an mAP@50 of 0.988, increasing from 0.984, and a Recall of 0.971, improving from

0.955, which indicates significantly higher model sensitivity. In addition, Precision increases to 0.959, while localization performance, measured by mAP@50–95, improves to 0.914.

Overall, this study validates that a structural intervention at the Neck level, combined with adaptive data augmentation, represents an effective approach for the digital preservation of Indonesian traditional textiles. Future research may focus on expanding the primary dataset for underrepresented contemporary motifs (such as Gajah Seblat, Dol, and Kembang Cengkeh) and incorporating dataset variations under extreme lighting conditions. Furthermore, exploring advanced augmentation for handling truncated objects and applying model compression techniques will be crucial to enable optimal deployment on mobile devices for real-world applications.

REFERENCES

- [1] M. Igiriza, R. A. Valentino, G. Wiradharma, Z. Adamy, and A. Anjani, "Preservation of Balinese batik knowledge as an effort to preserve indigenous knowledge," *J. Kaji. Inf. Perpust.*, vol. 12, no. 2, pp. 205–220, 2024, [Online]. Available: <https://jurnal.unpad.ac.id/jkip/article/view/58663>
- [2] E. N. Intania, "Revitalisasi Batik sebagai Simbol Identitas Nasional dalam Era Globalisasi dan Industri Kreatif Global," *Maliki Interdiscip. J.*, vol. 3, pp. 1050–1055, 2025, [Online]. Available: <https://urj.uin-malang.ac.id/index.php/mij/article/view/14909>
- [3] M. Utami, "Model ResNet50 dengan Optimalisasi Activation Function dan Optimizer Berbasis Gamma - MSR Untuk Klasifikasi Citra Motif Batik," Universitas Sriwijaya, 2025.
- [4] Y. Rao *et al.*, "Feature enhancement and attention mechanism fusion method for Cantonese embroidery image detection based on YOLOv8," *npj Herit. Sci.*, vol. 13, no. 1, p. 521, 2025, doi: 10.1038/s40494-025-02079-x.
- [5] W. Azizah and S. Agustin, "Feature Extraction using Histogram of Oriented Gradients and Moments with Random Forest Classification for Batik Pattern Detection," *J. SISFOKOM (Sistem Inf. dan Komput.)*, vol. 14, pp. 8–14, 2025, doi: 10.32736/sisfokom.v14i1.2352.
- [6] E. Khoirunnisa, F. Alzami, R. A. Pramunendar, and R. A. Megantara, "Enhanced Semarang Batik Classification using MobileNetV2 and Data Augmentation," *Sinkron*, vol. 9, no. 1, pp. 43–54, 2025, doi: 10.33395/sinkron.v9i1.14308.
- [7] A. F. Rozi, A. Wibowo, and B. Warsito, "Resampling Technique for Imbalanced Class Handling on Educational Dataset," *JUITA*, vol. 11, no. 1, pp. 77–85, 2023, doi: 10.30595/juita.v11i1.15498.
- [8] S. A. Sutresno, "The Classification of Batik Besurek Fabric Motifs in Indonesia Utilizing YOLOv8 for Enhanced Cultural Preservation," *J. Comput. Syst. Informatics*, vol. 6, no. 1, pp. 86–95, 2024, doi: 10.47065/josyc.v6i1.6123.
- [9] M. Fajar, J. Permana, J. Caesar, R. Bakar, and N. A. Fauzan, "YOLOv11 Model as a Smart Solution for Waste Identification and Classification in Automated Waste Management System," *J. Ilmu Komput. dan Inf.*, vol. 2, pp. 207–216, 2025, doi: 10.21609/jiki.v18i2.1490.
- [10] Y. Wang *et al.*, "A lightweight YOLOv11-based framework for small steel defect detection with a newly enhanced feature fusion module," *Sci. Rep.*, vol. 15, no. 1, p. 34322, 2025, doi: 10.1038/s41598-025-16619-9.
- [11] M. Mao and M. Hong, "YOLO Object Detection for Real-Time Fabric Defect Inspection in the Textile Industry: A Review of YOLOv1 to YOLOv11," *Sensors*, vol. 25, no. 7, pp. 1–43, 2025, doi: 10.3390/s25072270.
- [12] Q. Li, T. Wu, T. Xu, J. Lei, and J. Liu, "A Novel YOLO Algorithm Integrating Attention Mechanisms and Fuzzy Information for Pavement Crack Detection," *Int. J. Comput. Intell. Syst.*, vol. 18, no. 1, p. 158, 2025, doi: 10.1007/s44196-025-00894-5.
- [13] L. Z. Id, B. Ma, Y. Dong, Z. Yin, and F. Lu, "DCFE-YOLO : A novel fabric defect detection method," pp. 1–21, 2025, doi: 10.1371/journal.pone.0314525.
- [14] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *Computer Vision -- ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Cham: Springer International Publishing, 2018, pp. 3–19. doi: 10.1007/978-3-030-01234-2_1.
- [15] H. Chen, Z. Chen, and H. Yu, "Enhanced YOLOv5 : An Efficient Road Object Detection Method," *Sensors*, vol. 23, no. 20, 2023, doi: 10.3390/s23208355.
- [16] S. Praveen and Y. Jung, "CBAM-STN-TPS-YOLO : Enhancing Agricultural Object Detection through Spatially Adaptive Attention Mechanisms," 2023, doi: 10.48550/arXiv.2506.07357.
- [17] N. Garshasebi, "Improving Small Object Detection in Remote Sensing with YOLO11- CBAM and Deep Learning," vol. 50, no. 1, pp. 217–227, 2025, doi: 10.5281/zenodo.15845503.
- [18] R. Sapkota and M. Karkee, "Integrating YOLO11 and Convolution Block Attention Module for Multi-Season Segmentation of Tree Trunks and Branches in Commercial," pp. 1–19, 2024, doi: 10.48550/arXiv.2412.05728.
- [19] N. Putu, D. Ariani, S. Dewi, J. Aji, D. Aryasa, and I. G. Hendrayana, "YOLOv8-Based Quality Detection of Bali MSMEs Staple Food," *TIERS*, vol. 6, no. 2, pp. 171–186, 2026, doi: 10.38043/tiers.v6i2.7144.
- [20] Y. Jin *et al.*, "A Real-Time Fabric Defect Detection Method Based on Improved YOLOv8," *Appl. Sci.*, vol. 15, no. 6, 2025, doi: 10.3390/app15063228.
- [21] M. B. BESUREK, "Object Detection Motif Asli Batik Besurek Dataset," Mar. 2024, *Roboflow*. [Online]. Available: <https://universe.roboflow.com/motif-batik-besurek/object-detection-motif-asli-batik-besurek>
- [22] S. Nepal *et al.*, "MoodCapture: Depression Detection Using In-the-Wild Smartphone Images.," *Proc. SIGCHI*



- Conf. Hum. factors Comput. Syst. CHI Conf.*, vol. 2024, May 2024, doi: 10.1145/3613904.3642680.
- [23] T. Kumar, R. O. B. Brennan, S. Member, A. Mileo, and M. Bendeche, "Image Data Augmentation Approaches: A Comprehensive Survey and Future Directions," 2023, doi: 10.48550/arXiv.2301.02830.
- [24] T. Islam, M. S. Hafiz, J. R. Jim, M. M. Kabir, and M. F. Mridha, "A systematic review of deep learning data augmentation in medical imaging: Recent advances and future research directions," *Healthc. Anal.*, vol. 5, p. 100340, 2024, doi: <https://doi.org/10.1016/j.health.2024.100340>.
- [25] M. H. Putri, U. Zaky, and B. A. Prabawa, "Optimizing Data Augmentation Parameters in YOLOv11 for Enhanced Rip Current Detection on Small Datasets from Depok-Parangtritis Coastline," *J. Tek. Inform.*, vol. 6, no. 5, pp. 3938–3957, 2025, doi: 10.52436/1.jutif.2025.6.5.5352.
- [26] S. Chen, Y. Liu, H. Zhang, and Y. Cai, "A human location and action recognition method based on improved Yolov11 model," *Discov. Artif. Intell.*, vol. 5, no. 1, p. 232, 2025, doi: 10.1007/s44163-025-00492-6.
- [27] R. R. Hamidi, M. K. Herliansyah, D. Sukma, and E. Atmaja, "Comparative Analysis of YOLOv5n and YOLOv8n Deep Learning Models for Precision Detection of Klowong Defects in Batik Fabric," *TIERS*, vol. 6, no. 1, pp. 74–86, 2025, doi: 10.38043/tiers.v6i1.6499.
- [28] J. Gao and Z. Wang, "YOLOv8-CBAM: a study of sheep head identification in Ujumqin sheep," no. February, pp. 1–10, 2025, doi: 10.3389/fvets.2025.1514212.
- [29] S. N. Rao, "YOLOv11 Architecture Explained: Next-Level Object Detection with Enhanced Speed and Accuracy," [www.medium.com](https://www.medium.com/@nikhil-rao-20/yolov11-explained-next-level-object-detection-with-enhanced-speed-and-accuracy-2dbe2d376f71). Accessed: Jan. 20, 2026. [Online]. Available: <https://medium.com/@nikhil-rao-20/yolov11-explained-next-level-object-detection-with-enhanced-speed-and-accuracy-2dbe2d376f71>
- [30] W. Cai, Y. Chen, X. Qiu, M. Niu, and J. Li, "LLD-YOLO: A Low-Light Object Detection Algorithm Based on Dynamic Weighted Fusion of Shallow and Deep Features," *IEEE Access*, vol. 13, no. January, pp. 69967–69979, 2025, doi: 10.1109/ACCESS.2025.3558574.
- [31] I. Goodfellow, *Deep Learning*. 2016.
- [32] X. Wang *et al.*, "Research on multi class pests identification and detection based on fusion attention mechanism with Mask-RCNN-CBAM," *Front. Agron.*, vol. Volume 7-2025, 2025, doi: 10.3389/fagro.2025.1578412.